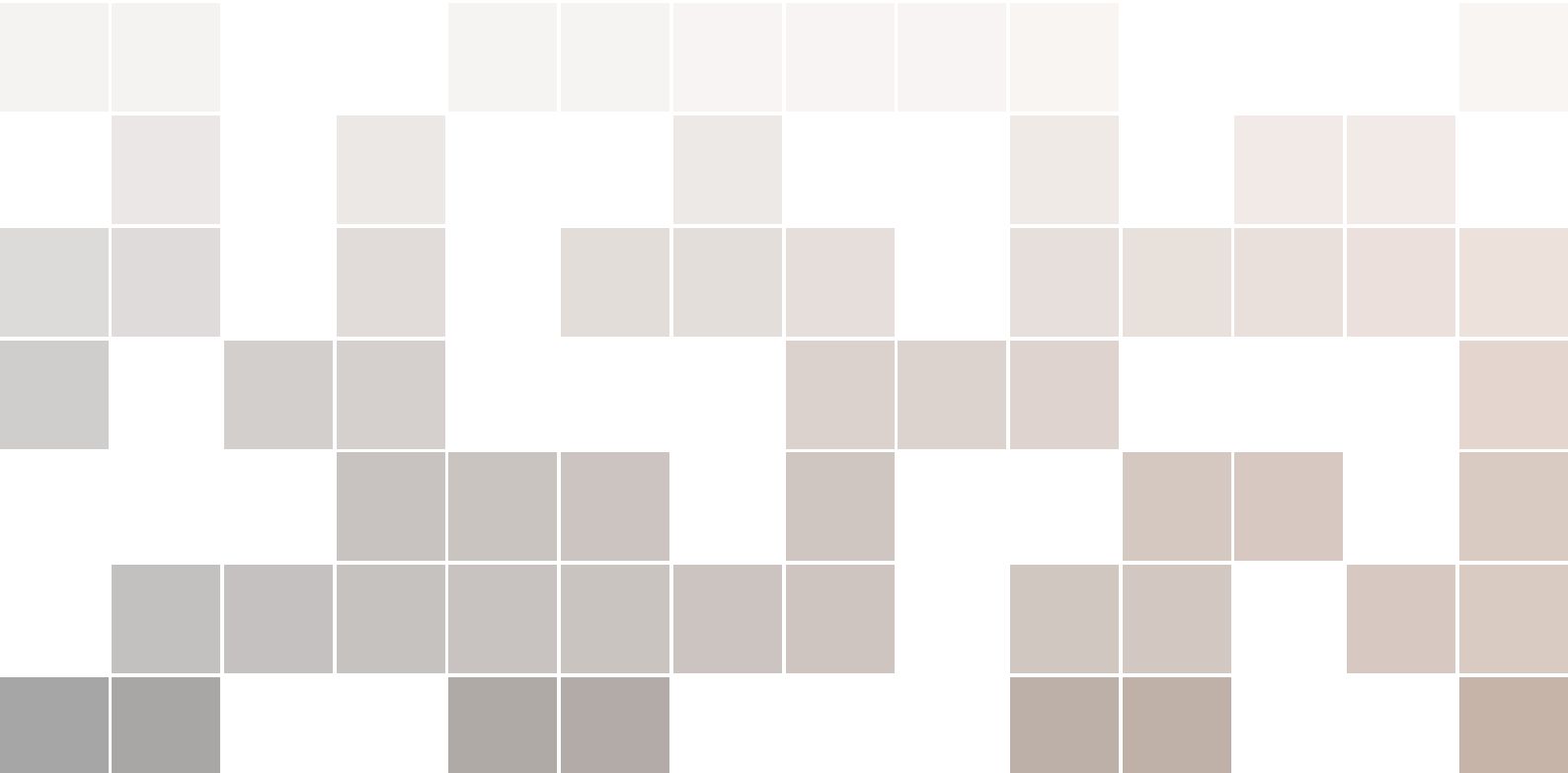


经典力学新讲

现代导引

陈童



Copyright © 2021 陈童

PUBLISHED BY PUBLISHER

下载网址: <http://newquanta.com/>

作者本人 (即陈童) 对本讲义保有所有版权, 目前读者可在上面的网站上自由下载, 也可以私自传播, 但仅限用于学习或教学等非商业用途, 禁止私自以本讲义的部分或全部来获利, 一经发现, 作者将保留诉诸法律的权利。如果你有任何反馈, 或者发现书中的任何错误, 请给作者发邮件, tongchen@ecut.edu.cn


2021 年 6 月

Contents

1	哈密顿力学初步	9
1.1	势能与力	10
1.1.1	能量与力	10
1.1.2	哈密顿正则方程	11
1.1.3	势能曲线	12
1.2	无磨损摩擦的一个物理模型	16
1.3	微分形式	18
1.3.1	微分形式	19
1.3.2	外微分与斯托克斯公式	21
1.3.3	保守力的特性	24
2	最小作用量原理	27
2.1	物理学的局部视角与全局视角	28
2.2	从费马原理到变分法	29
2.2.1	费马原理	29
2.2.2	变分和泛函导数	30
2.2.3	欧拉-拉格朗日方程	32
2.3	相空间的最小作用量原理	34
2.4	坐标空间的最小作用量原理	36
2.4.1	最小作用量原理与拉格朗日方程	36

2.4.2	发现洛伦兹力	39
2.4.3	最小作用量原理能够导出任何方程吗?	40
2.4.4	真实路径的作用量是否取极小值?	41
2.5	广义坐标和广义动量	42
3	对称性	47
3.1	对称性与守恒定律	47
3.1.1	对称性就是作用量的不变性	47
3.1.2	诺特定理	48
3.1.3	动量守恒、角动量守恒、能量守恒	50
3.1.4	宇宙的总能量是零	54
3.1.5	* 伽利略不变性	55
3.2	对称性决定作用量	56
3.2.1	相对论不变性	57
3.2.2	相对论粒子的作用量	59
3.2.3	与电磁场的耦合	61
3.2.4	对称性决定一切?	63
4	两体问题和受限三体问题	65
4.1	两体问题	65
4.2	中心力场	67
4.2.1	中心力场的一般讨论	67
4.2.2	开普勒问题	69
4.2.3	散射问题	72
4.3	受限三体问题和拉格朗日点	74
5	再谈哈密顿力学	79
5.1	泊松括号	80
5.1.1	辛结构	80
5.1.2	泊松括号	81
5.1.3	涡旋的运动	84
5.2	正则变换	87
5.2.1	正则变换作为相空间坐标变换	87
5.2.2	正则变换作为相空间的微分同胚映射	89
5.2.3	无穷小正则变换与诺特定理	90
5.2.4	刘维尔定理和庞加莱回归定理	93
5.2.5	生成函数与哈密顿正则方程	96

5.3	哈密顿-雅可比方程	99
5.3.1	哈密顿-雅可比方程	99
5.3.2	可积系统	101
5.3.3	举例：哈密顿雅可比方程求解中心力场问题	102
6	哈密顿系统中的规则与混沌	105
6.1	最规则的哈密顿系统-可积系统	105
6.1.1	不变环面	106
6.1.2	另一个刘维尔定理	107
6.1.3	作用-角变量	109
6.1.4	开普勒问题和两维谐振子问题	110
6.1.5	绝热不变量	113
6.2	近可积系统与混沌系统	115
7	振动	117
7.1	小振动	117
7.1.1	小振动的拉格朗日量	117
7.1.2	简正坐标的存在性	118
7.1.3	如何寻找简正坐标	119
7.1.4	举例	121
7.2	参数共振	122
7.2.1	现象与问题	122
7.2.2	参数共振的一般性分析	123
7.2.3	单自由度系统的参数共振	127
8	刚体的运动	131
8.1	运动学	132
8.1.1	刚体转动的位形空间	132
8.1.2	角速度矢量	133
8.2	惯量张量	134
8.2.1	惯量张量	134
8.2.2	平行轴定理	137
8.2.3	同时考虑平动和转动	137
8.3	欧拉陀螺	138
8.4	欧拉角	141
8.4.1	欧拉角	141
8.4.2	角速度的欧拉角表示法	143
8.5	拉格朗日陀螺	144



前言

《经典力学新讲》的核心目标是用现代的观点重新整理和澄清经典力学的理论结构，凸显理论之美。这个讲义中的有些内容是我多年的教学积累，但促使我开始写这个新式讲义的主要动机，却是物理学大师 Polyakov 关于现代经典力学课程的在线视频。读者也许可以在我讲义的多处找到 Polyakov 的痕迹。

不过，本讲义并不是对 Polyakov 课程的跟随，讲义的理论逻辑和 Polyakov 的课程有些根本性的不同，也和通常的理论力学教材（包括朗道的《力学》）有根本性的不同。比方说，本讲义是将能量和哈密顿量作为第一概念，而不是将拉格朗日量作为第一概念，我们对力这个概念的处理也是这样，是直接用能量随空间位置的变化来定义力，而不是将力作为基本概念引入，总之，我试图将整个理论力学的大厦根基于能量和哈密顿量的概念之上。当然，我并没有抛弃最小作用量原理（本讲义的第二章就是讲最小作用量原理），但由于以哈密顿量为第一概念，我对最小作用量原理的处理也是由相空间的最小作用量原理出发，导出通常的最小作用量原理（朗道的《力学》就是以这一最小作用量原理为出发点的）。在这种处理中，勒让德变换将是一个自然而然的东西。

值得说明的是，在我看来，这么多年来没有人这样处理理论力学是一件很奇怪的事情。原因有二，第一，哈密顿量由于直接对应能量，当然比拉格朗日量更物理也更好接受，这样看的话通常先引入拉格朗日量，进而将哈密顿量定义为拉格朗日量的勒让德变换就是一件不可思议的事情。从我的教学经验来看，学生往往既不理解为什么有这么一个奇怪的拉格朗日量，也无从理解为什么突然要来这么一个勒让德变换，然后得出一个奇怪的哈密顿量。更有甚者，很多学生学完理论力学之后依然觉得哈密顿量只是碰巧通常等于能量，而不理解哈密顿量就是能量的物理描述，不理解如果你的哈密顿量不能完全描述能量，那要么你就得在哈密顿量中加减一些项，要么就是你对能量的定义不合适，要修改的是你对能量的定义。在我看来，正确的处理就是，先作为对能量的描述引入哈密顿量，再通过最小

作用量原理自然地勒让德变换到拉格朗日量。第二，从量子力学来看，量子系统会随时间么正演化，么正演化的生成元自然定义了哈密顿算符，哈密顿算符就是能量算符，其本征值就是能量（可见，哈密顿量不是碰巧等于能量，它描述的就是能量！）。无论是从实际的量子力学课程来看，还是从量子力学么正性来看，哈密顿算符无疑都应该是第一概念。当然，量子力学有路径积分表述，但为了保证么正性，人们也是需要么正演化算符和哈密顿算符出发推导路径积分表述。值得强调的是，在这样的推导中，首先出现的也是相空间作用量，然后你把动量先积掉，才导出通常的作用量，导出费曼的路径积分。这么看来，量子力学的这个推理逻辑天然对应本讲义的逻辑，而不是通常理论力学教材那样建立理论力学的逻辑。

本书采用诺特技巧（Noether's trick）来处理诺特定理，当然这主要是因为诺特技巧本身非常优雅。同时，由此不仅能给出诺特定理恐怕是最简单的证明，而且产生了一个从作用量中得到守恒量的系统办法。

本书的另一个特色是介绍了外微分形式，并在第五章和第六章中将之应用到哈密顿力学的详细讨论中。正因为如此我们才可以通过辛形式（或者说辛结构）串起哈密顿力学的各个知识点，体现了哈密顿力学就是相空间的辛几何这一现代观点。

除此之外，我们还证明了哈密顿力学中一系列漂亮而重要的定理，包括诺特定理的重新回顾，刘维尔定理，庞加莱回归定理，还有可积系统的刘维尔-阿诺德定理。尤其是对于后面两个定理，许多经典力学教材是不作讨论或者不给出证明的。

相比于传统经典力学教材，在理论的应用性内容中，我们额外讨论了限制性三体问题，讨论了什么是拉格朗日点。此外我们还跟着阿诺德的书给出了关于多自由度参数共振的一个精妙讨论。

由于是草稿版，讲义中有一些插图是手绘的，还有一些插图直接取自网络或者是从其它资料里面截图出来的，文中并没有全部一一说明，后续版本会逐渐改过来。



1. 哈密顿力学初步

本章初步介绍哈密顿正则方程，并详细讨论力的概念如何定义，势能与力的关系，以及保守力有何特性。同时本章也详细讨论了势能曲线，并利用它讨论了一个重要的关于无磨损摩擦的物理模型，独立振子模型。最后，本章也为读者讲述了微分形式这一重要数学方法的初步知识。特别是阐述了 3 维矢量分析中的斯托克斯公式和高斯定理如何统一成一个漂亮的公式！

1.1 势能与力

1.1.1 能量与力

这本书要讲述的是力学，那么什么是力？学过牛顿运动定律的读者会回答力是物体运动状态改变的原因。不过，在现代物理中，我们并不这样定义力，这时候，力不再是一个最基本的概念，能量才是，力的概念是通过粒子间相互作用的能量来定义的，大体来说，力是相互作用能量随着粒子空间位置变化的剧烈程度。

从这个定义可以看出，对于粒子我们才能谈它受到的力，如果是场，比方说电磁场，由于它并不具有一个位置坐标，而是空间分布，所以我们就不能定义受力这一概念，但是场的相互作用能量始终是有定义的。

那么，什么是能量呢？通常的回答是，能量是做功的本领，但是为了定义功，我们又需要先定义力，这不符合我们这里采用的逻辑，我们这里是把能量而不是力作为第一概念的。能量是什么呢？这在现代物理中有直接定义，不过我们想将相关的讨论放在后面的章节中，暂时让我们不加定义地接受能量的概念。当然，我们还得接受能量的转化和守恒定律，根据这条定律，我们暂时可以含糊地说，能量是一种具有多种表现形式的东西，能量的这些不同形式可以相互转化，但是，对于一个孤立系统，能量的总量保持不变。

因此，力是相互作用能量随粒子空间位置的变化，不过，为了把这个概念进一步精确化，很多时候我们需要势能的概念。势能就是一种相互作用能量，通常它和粒子的运动速度无关，只和粒子的空间位置有关。假设一个粒子受某个外力场作用，则它和这个外力场之间就有相互作用势能，可以用函数 $V(\mathbf{x})$ 来表示。比方说，地面附近一个高度为 z 的物体处于地球的重力场中，它和地球相互作用的势能为 $V(\mathbf{x}) = mgz$ 。再比方说，两个质点之间的万有引力势能为 $V(\mathbf{x}) = -G\frac{m_1 m_2}{r}$ ，两个点电荷之间的库伦相互作用势能为 $V(\mathbf{x}) = \frac{e_1 e_2}{4\pi\epsilon_0 r}$ 。这些势能的表达式是如何得到的呢？回答是，最初都是从实验中归纳出来的，不过在现代物理中，我们往往可以通过研究场的规律来推导出它们，通过电磁场的麦克斯威方程可以得出点电荷之间的库伦相互作用势能，通过研究引力场的爱因斯坦方程可以得出万有引力势能。

现在，假设我们知道了一个粒子与外力场的相互作用能量，也就是势能 $V(\mathbf{x})$ ，则这个粒子受到的力 \mathbf{F} 将由下式定义

$$\mathbf{F} \cdot d\mathbf{x} = -dV(\mathbf{x}), \quad (1.1)$$

或者也可以写作

$$\mathbf{F} = -\frac{\partial V}{\partial \mathbf{x}} = -\nabla V = -\left(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}, \frac{\partial V}{\partial z}\right). \quad (1.2)$$

这些式子中额外的那个负号是为了使得牛顿运动定律的形式或者机械能的表达式与通常的约定一致而引入的。

从上面两个式子可以看出，即使给势能函数 $V(\mathbf{x})$ 加上一个任意的常数，力的定义也将保持不变。实际上，在非相对论物理中，能量的定义本来就可以相差一个常数，因为在非相对论物理中，我们只能测量能量的差值，而无法测量能量的绝对值。能量定义的这个

不定常数反映在势能函数中，就是势能函数可以相差一个常数，为了确定这个常数，我们需要人为地选定一个空间位置，称之为势能参考点，令粒子在这个点处的势能为零。

1.1.2 哈密顿正则方程

利用上面关于力的定义，我们就可以将牛顿运动定律重写为

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} = -\frac{\partial V}{\partial \mathbf{x}}. \quad (1.3)$$

其中 \mathbf{p} 为粒子的动量，其定义为

$$\frac{d\mathbf{x}}{dt} = \frac{\mathbf{p}}{m}, \quad (1.4)$$

这里 m 为粒子的质量。

假设我们定义物理量 $H(\mathbf{x}, \mathbf{p}) = \frac{\mathbf{p}^2}{2m} + V(\mathbf{x})$ ，并将 \mathbf{x} 和 \mathbf{p} 看成是两个相互独立的变量，则很容易验证上面关于牛顿运动定律的两个方程 (1.3)、(1.4) 可以重写成一种非常对称、非常漂亮的形式

$$\frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{x}}, \quad \frac{d\mathbf{x}}{dt} = \frac{\partial H}{\partial \mathbf{p}}. \quad (1.5)$$

由这两个方程，我们将有

$$\frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{x}} \cdot \frac{d\mathbf{x}}{dt} + \frac{\partial H}{\partial \mathbf{p}} \cdot \frac{d\mathbf{p}}{dt} = \frac{\partial H}{\partial \mathbf{x}} \cdot \frac{\partial H}{\partial \mathbf{p}} - \frac{\partial H}{\partial \mathbf{p}} \cdot \frac{\partial H}{\partial \mathbf{x}} = 0. \quad (1.6)$$

也即是说， H 是一个守恒量。实际上， H 就是粒子的总能量， $\frac{\mathbf{p}^2}{2m}$ 是动能， $V(\mathbf{x})$ 是势能， H 是动能势能的总和。因此，除了重写了牛顿定律之外，我们刚才只不过是证明了能量守恒定律。

(1.5) 式也称作哈密顿正则方程，很显然，它和牛顿运动定律相互等价。物理量 $H(\mathbf{x}, \mathbf{p})$ 也称作哈密顿量。值得强调的是，作为一个物理量，哈密顿量代表的是粒子的总能量，但作为一个数学表达式， $H(\mathbf{x}, \mathbf{p})$ 得是将坐标 \mathbf{x} 和动量 \mathbf{p} 作为相互独立的变量表达出来的能量，因此作为一个表达式，哈密顿量 H 与 $\frac{1}{2}m\left(\frac{d\mathbf{x}}{dt}\right)^2 + V(\mathbf{x})$ 是不同的。

哈密顿正则方程很容易推广到多粒子情形，这时候系统的相互作用势能一般来说依赖于每一个粒子的位置坐标，表示为函数 $V(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ 。比方说，两粒子系统的万有引力势能为 $V(\mathbf{x}_1, \mathbf{x}_2) = -G \frac{m_1 m_2}{|\mathbf{x}_1 - \mathbf{x}_2|}$ 。多粒子系统的哈密顿量同样代表总能量，是每个粒子的动能与系统的相互作用势能之和，写成

$$H = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N). \quad (1.7)$$

可见，这时候哈密顿量依然以粒子坐标和动量为独立变量，不过是以每一个粒子的坐标和动量为独立变量。相应的，多粒子体系的哈密顿正则方程为

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial H}{\partial \mathbf{x}_i}, \quad \frac{d\mathbf{x}_i}{dt} = \frac{\partial H}{\partial \mathbf{p}_i}. \quad (1.8)$$

为了验证这个方程等价于多粒子体系的牛顿运动定律，我们需要注意到现在第 i 个粒子的受力 \mathbf{F}_i 由下式定义

$$\sum_i \mathbf{F}_i \cdot d\mathbf{x}_i = -dV \Leftrightarrow \mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{x}_i}. \quad (1.9)$$

完全类似于单粒子情形，读者也可以证明多粒子体系的能量守恒定律。

以上关于多粒子哈密顿正则方程的讨论演示了现代物理研究一个物理系统的基本方法：即首先我们找到系统的基本力学变量，这在多粒子系统中就是所有的 $\mathbf{x}_i, \mathbf{p}_i$ ，在场论系统中就是场位形和场的正则动量（读者可以先跳过关于场论系统的这句话）；其次，利用这些基本力学变量写出系统的哈密顿量，也就是写出系统总能量的表达式；最后，写出形如 (1.8) 式的哈密顿正则方程。

所以，在现代物理中，力不是最基本的，哈密顿量才是，相应的，牛顿运动定律也不是最基本的，哈密顿正则方程才是。当然，对于单粒子和多粒子系统，正如我们已经看到的，这时候哈密顿正则方程和牛顿运动定律的 $F = ma$ 完全等价。哈密顿正则方程的好处是给力学设定了一个一般性的框架，使得我们的力学体系不仅可以研究粒子，还可以研究自旋系统，研究场等等。

哈密顿正则方程当然也有其局限，那就是它不能包括所有的力学系统。能被哈密顿正则方程描述的力学系统称之为哈密顿系统，它包括几乎所有的基本物理系统。但是，有一些力不属于自然界中的基本力，这时候相应的力学系统可能就不能用哈密顿正则方程来描述，比方说，摩擦力就不是基本力，如果我们关心摩擦力的表现，那相应的方程就不能是哈密顿正则方程。然而，即使对于摩擦力，从现代物理学的眼光来看，我们更关心的也是其本质为何，即如何从基本物理的角度理解摩擦力的起源，而这时候哈密顿正则方程又有用武之地了，后文我们会进一步考察摩擦力的物理模型。

1.1.3 势能曲线

作为例子，下面我们来分析一下单粒子的一维运动。这时候粒子的坐标为 x ，相应的动量为 p 。粒子的势能可以表示成一个单变量函数 $V(x)$ ，我们可以以 x 为横坐标，以 V 为纵坐标，画出势能函数 $V(x)$ 的曲线，如图 (1.1) 所示，这一曲线就是所谓的势能曲线。

给定能量 E ，由于能量守恒，粒子运动到任何 x 处其总能量将都是 E ，我们可以画一条水平线表示它，如图 (1.1) 中虚线所示。根据能量守恒 $E = \frac{p^2}{2m} + V(x)$ ，动能 $p^2/(2m)$ 总大于等于零，由此可见，粒子的可能运动区域必定满足

$$E \geq V(x), \quad (1.10)$$

如图 (1.1) 中的 $(-\infty, x_a] \cup [x_b, x_c] \cup [x_d, +\infty)$ 。很显然，由于中间间隔的能量小山包（称之为势垒）是运动禁区，所以 $(-\infty, x_a]$ 、 $[x_b, x_c]$ 、 $[x_d, +\infty)$ 这三个可能运动区域是相互隔绝的。

我们来分析一下粒子在 $[x_b, x_c]$ 区间上的运动。假定粒子从 x_b 点出发，在这一点上，粒子的全部能量都是势能，动能为零，因此初速度等于零。由于 x_b 点左侧的势垒将粒子挡住了，所以粒子只能向右边运动，这时候它的势能逐渐减少，动能逐渐增加，因此粒子会运动得越来越快，当粒子到达 $[x_b, x_c]$ 区间上势能曲线的最低点时，其势能达到最小，动能

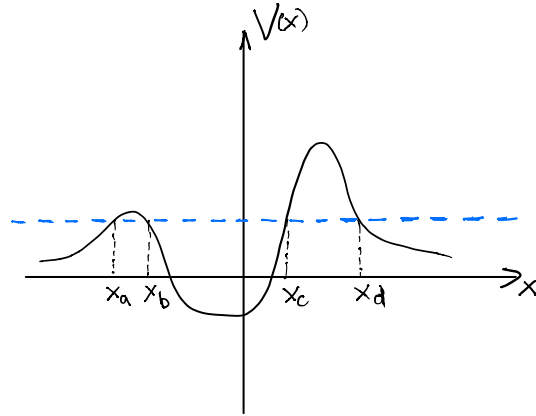


Figure 1.1: 势能曲线

达到最大，粒子的运动速度达到最大。此后，由于惯性，粒子继续向右边运动，但这时候势能在逐渐增加，因此动能会逐渐减小，粒子会逐渐减速。当达到 x_c 点时，所有的动能都转化为了势能，粒子的速度将会减为零。由于 x_c 的右边又是势垒，所以在 x_c 点的粒子只能掉头往回跑，经过加速阶段和减速阶段，粒子最终将回到 x_b 点。此后，粒子将重复上面的运动过程。所以，粒子实际上是在 $[x_b, x_c]$ 区间上作周期性来回运动。

但是，对于 $(-\infty, x_a]$ 运动区间，假设粒子从 $-\infty$ 处出发往右边运动，则它会一直运动到 x_a 处，然后被 x_a 右边的势垒挡住，从而掉头，然后一直往左边跑，一直跑到无穷远。但是， $-\infty$ 处不再有势垒使得粒子再次掉头了！所以，粒子在 $(-\infty, x_a]$ 区间上的运动并没有周期性来回运动。同样，由于右边没有足够高的势垒挡住粒子， $[x_d, +\infty)$ 区间上的运动也是类似的。我们称 $(-\infty, x_a]$ 区间和 $[x_d, +\infty)$ 区间上的这种运动为**散射**，它描述无穷远处入射粒子被势垒散射回去的过程。相反，我们称 $[x_b, x_c]$ 区间上的运动为**束缚运动**，相应的运动轨道为**束缚轨道**，因为这时粒子被束缚在一个势能的小山谷（称为势阱）中，无法跑到无穷远处去。很显然，散射不是一种束缚运动。

根据哈密顿正则方程的精神，我们将 x, p 看成两个独立变量，以 x 为横坐标， p 为纵坐标所构成的两维空间，就称为粒子的**相空间**，粒子在相空间中按照哈密顿正则方程演化。粒子在相空间中的运动轨迹就称为相空间轨道。在我们这个例子中，相空间轨道显然由能量守恒方程给出，即满足

$$E = \frac{p^2}{2m} + V(x). \quad (1.11)$$

我们可以示意性地画出 $(-\infty, x_a]$ 、 $[x_b, x_c]$ 、 $[x_d, +\infty)$ 这三个运动区间的相空间轨道，如图 (1.2) 所示。很明显，这里的束缚轨道正好是闭合轨道¹，而散射运动则对应非闭合轨道。

在现在的例子中，相空间轨道主要由其守恒能量 E 刻画，不同的能量对应不同的相空间轨道，我们可以示意性地画出不同 E 值的相空间轨道，如图 (1.3) 所示。

利用相空间轨道，我们很容易得出 $[x_b, x_c]$ 区间周期运动的周期公式，记周期为 T ，显然它依赖于守恒能量 E ，可以写作 $T(E)$ ，利用哈密顿正则方程 $\frac{dx}{dt} = \frac{\partial H}{\partial p} = \frac{p}{m}$ ，我们很容易积

¹这个结论不能推广到一般情形，在非一维运动中，束缚轨道不一定是闭合轨道。

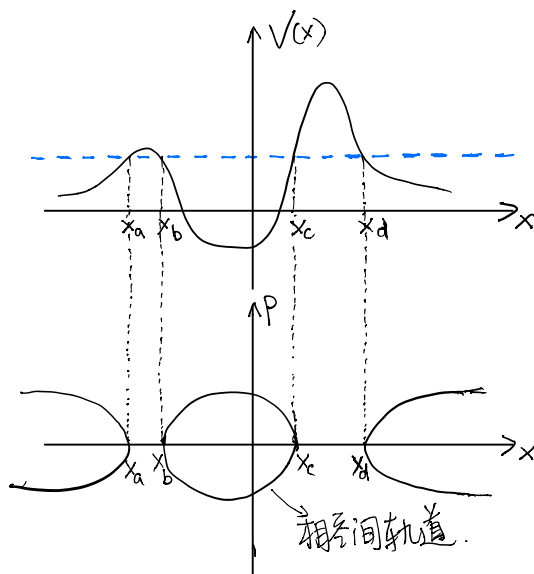
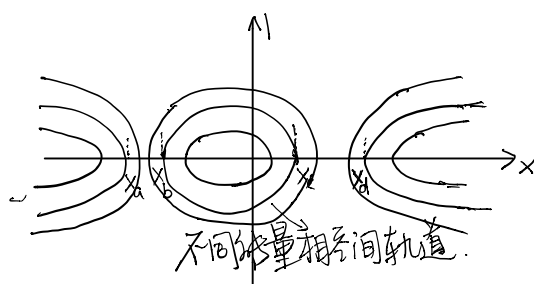
Figure 1.2: 给定能量 E 的相空间轨道。

Figure 1.3: 不同能量的相空间轨道。

分得出

$$T(E) = \oint_E dt = \oint_E \frac{m}{p} dx, \quad (1.12)$$

式中的积分沿着能量为 E 的闭合相空间轨道进行。或者我们也可以代入相空间轨道的方程 (1.11), 即代入 $p = \sqrt{2m(E - V(x))}$, 进而将这个周期公式写成

$$T(E) = \oint_E \frac{m}{\sqrt{2m(E - V(x))}} dx. \quad (1.13)$$

另一方面, 我们很容易计算闭合相空间轨道所包围的相空间面积, 通常记作 $2\pi I(E)$, 很显然

$$2\pi I(E) = \oint_E p dx = \oint_E \sqrt{2m(E - V(x))} dx. \quad (1.14)$$

由于粒子总是沿着相空间轨道在“演化”, 每一条相空间轨道在时间演化下都是不变的, 因此 $I(E)$ 作为沿着闭合轨道积分的结果当然也是不变的, 即 $I(E)$ 是一个守恒量, 实际上,

上面的公式告诉我们，它只依赖于守恒能量 E 。从 (1.14) 这个式子出发，再利用上面的周期公式 (1.13)，我们很容易得到

$$\frac{\partial I(E)}{\partial E} = \frac{1}{2\pi} \oint_E \frac{\partial}{\partial E} \sqrt{2m(E - V(x))} dx = \frac{T(E)}{2\pi}. \quad (1.15)$$

或者也可以反过来，将系统能量 E 看成是相轨道包围“面积” I 的函数，然后利用周期运动的角频率 $\omega = 2\pi/T$ ，进而将上面这个公式改写成

$$\frac{\partial E}{\partial I} = \omega(I). \quad (1.16)$$

这里只是对一维运动推导出这个结果，后面我们会把它推广到更一般的所谓可积系统的情形。

上面的结果 (1.16) 在量子物理的建立过程中起过重要的作用。根据玻尔的定态假设，一个作束缚运动的粒子，其能量是量子化的，能量的每一个量子化值称为能级，我们不妨记第 n 能级的能量为 E_n 。根据定态跃迁假设，粒子从第 n 能级往下跃迁到第 $n-1$ 能级，会放出一个角频率为 ω 的光子，根据爱因斯坦的光量子学说，这个光子的能量将为 $\hbar\omega$ ，由能量守恒，我们将有 $E_n - E_{n-1} = \hbar\omega$ ，当量子数 n 很大时，我们可以将这个方程近似写成

$$\frac{\partial E_n}{\partial n} = \hbar\omega. \quad (1.17)$$

将上式与前面的 (1.16) 式比较，我们可以知道，当 n 很大时，应该有

$$I = n\hbar. \quad (1.18)$$

这就是玻尔-索末菲量子化条件。

特别的，我们可以考察所谓的一维线性谐振子，也就是一个弹簧，它的弹性势能当然是 $V(x) = \frac{1}{2}kx^2$ ， x 为弹簧的伸长量， k 为弹簧的劲度系数。通常人们会引入一个参数 $\omega = \sqrt{k/m}$ (ω 的含义我们马上会看到)，进而将弹性势能写成

$$V(x) = \frac{1}{2}m\omega^2 x^2. \quad (1.19)$$

从而相空间轨道是一个椭圆，满足方程

$$\frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2 = E \Leftrightarrow \frac{x^2}{(\frac{2E}{m\omega^2})} + \frac{p^2}{2mE} = 1. \quad (1.20)$$

利用椭圆的面积公式，我们容易得到相空间轨道包围的面积为 $2\pi E/\omega$ ，进而有

$$2\pi I = 2\pi E/\omega \Leftrightarrow E = I\omega. \quad (1.21)$$

与前面的 (1.16) 式比较就可以知道， ω 正是这个弹簧振子振动的角频率。

如果我们将玻尔-索末菲量子化条件代入 (1.21) 式，就可以得到一维线性谐振子的量子化能级

$$E = n\hbar\omega. \quad (1.22)$$

谐振子的这一量子化能级最早是普朗克在推导其黑体辐射的普朗克公式时假设的，就是著名的能量量子化假设，正是这个假设引入了量子的概念。谐振子能级的第一个完全量子力学的正确推导由海森堡给出，当时是作为他的矩阵力学的第一个应用。

从 (1.22) 式容易知道，谐振子能级有一个重要的特点，即假设 $E_1 = n_1 \hbar \omega$ 是系统的一个能级， $E_2 = n_2 \hbar \omega$ 也是系统的能级，则 $E_1 + E_2 = (n_1 + n_2) \hbar \omega$ 也必定是系统的一个能级。即是说，谐振子能谱具有某种可加性。这一性质正好和粒子性吻合，假设系统的某个能态对应一个粒子，其能量为 E_1 ，另有一个能态对应粒子 E_2 ，则这两个粒子同时存在时的能态就具有能量 $E_1 + E_2$ ，也即是说，粒子系统的能谱也具有某种可加性。这两者的吻合并非某种巧合，因为在量子场论中，粒子是场的激发态，而自由场的行为正好如同一些谐振子。实际上，两者的吻合反过来从物理上（而不是从数学推导上）解释了为什么自由量子场可以看成是多自由度的谐振子。

1.2 无磨损摩擦的一个物理模型

前面我们提到，摩擦力起作用的力学系统不是一个哈密顿系统。然而，我们也说了，摩擦力不是一种基本力，它只是一种宏观表现，是有其微观机制的。对于理解摩擦力的微观机制而言，哈密顿系统的势能和势能曲线概念可能又会很重要。当然，实际中的摩擦力很复杂，机制也很复杂，这里我们主要讨论无磨损摩擦的一个简化物理模型，通常叫作独立振子模型²。

在讨论独立振子模型之前，我们简单地讨论一下稳定平衡和不稳定平衡的概念。首先，粒子在平衡位置受力为零，而力又是势能对空间位置的偏导，受力为零就意味着这个偏导等于零，这对应于势能函数的极值位置，所以**平衡位置就是势能函数的极值位置**。但是，极值有极大值和极小值之分，稳定平衡位置就是势能函数的极小值位置，这时候你将粒子从平衡位置偏离一点，它的势能就会升高，所以达成这种偏离需要一定量的能量输入，无限小的扰动不能使粒子偏离平衡位置，所以是稳定平衡。不稳定平衡则与之相反，是势能函数的极大值位置，因此任意小的偏离都能使粒子的势能降低，使粒子偏离平衡位置不需要能量输入，即是说，任意扰动都能使得粒子偏离得越来越远，因此是不稳定平衡。从势能曲线来说，图 (1.4) 中的左图粒子就处在稳定平衡位置，而右图的粒子就处在不稳定平衡位置。

下面回到关于摩擦力的独立振子模型。摩擦力是一种耗散力，它会将机械能耗散为热能，要解释摩擦力的机制关键就是要解释能量是怎么耗散掉的。不妨考察 A, B 两块固体，它们沿着一个平整的接触面相对滑动，我们称这个接触面为界面。为了清楚地论述能耗散的机制，我们不妨假设系统处在一个零温环境中，因此任何无规则热运动的能量都能迅速地被环境吸收掉。

²G. M. McClelland and J. N. Glosli, in Ref "Fundamentals of Friction", edited by I. L. Singer and H. M. Pollock (Kluwer, Dordrecht, 1992), p. 405.

知乎上贾明子老师的一篇有趣回答将我引导到了这个重要的物理模型，在此向贾明子老师表示感谢。贾明子回答的链接是

<https://www.zhihu.com/question/384020347/answer/1585610578>

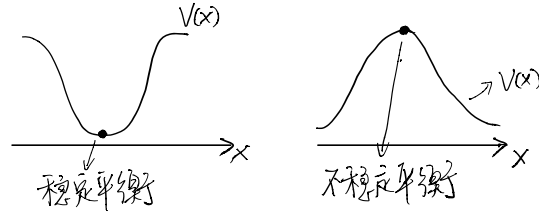


Figure 1.4: 稳定平衡和不稳定平衡。

所谓的独立振子模型，其关键是一个关于界面两边的固体原子结构的模型，如图 (1.5) 所示。我们假定 B 固体在界面附近是一些独立的原子，它们通过相互独立的弹簧振子 (可

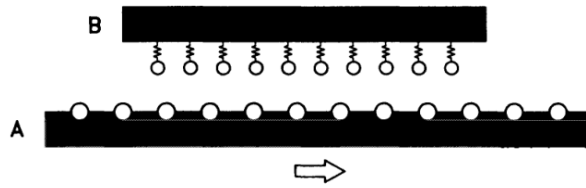


Figure 1.5: 独立振子模型

以理解为化学键) 连接到 B 的其余部分。而 A 固体在界面附近则是一些固定在固体上的规则排布的原子。现在，假定 B 固体固定，我们很缓慢地往右边，也就是 x 方向，推动 A 固体，使得两者相对滑动。

为了理解这个相对滑动过程，我们将注意力集中在 B 界面上的某个特定原子 B_0 ，如图 (1.6) 所示。 B_0 受到弹簧的弹力，因此有一个弹性势能 (作为其水平位置 x 的函数)，我们记为 $V_{BB} = \frac{1}{2}k_{BB}x^2$ 。但是， B_0 和 A 界面上规则排布的原子之间有排斥相互作用，相应的势能我们记为 V_{AB} 。显然 V_{AB} 在界面上每一个 A 原子位置都有一个峰值，如图 (1.6)(a) 图中的曲线 V_{AB} 所示。因此， B_0 原子的总势能为这两者的叠加，记为 $V_S = V_{BB} + V_{AB}$ ，其势能曲线如图 (1.6)(a) 所示。很显然，随着固体 A 往右边滑动，势能曲线 V_S 的形状会连续地变化，如图 (1.6) 所示。

我们注意到图 (1.6)(b),(c),(d) 中势能曲线 V_S 有两个局部极小， B_0 原子最初待在右边的那个极小位置。假定固体 A 滑动得足够缓慢，那 A 往右边滑动的过程就不会激起 B_0 原子的振动， B_0 就会一直待在右边的极小中。但是，从 (b) 到 (d)，右边这个极小的势能从比左边的极小更小，变成了比左边的极小更大。即是说，从 (b) 到 (d)， V_S 的整体最小位置从右边那个极小处切换到了左边。在图 (b) 中原子待在整体最小位置，因此它是“稳定”的，但在图 (d) 中整体最小切换到了左边，这样待在右边极小位置的原子就没有那么稳定了，这时候，只要稍微再推动一下固体 A (即图中从 (d) 到 (e) 的过程)，就能使得右边的这个极小变平，从而待在右边的 B_0 就会自发地掉落到左边的整体最小中，释放出来的势能将会转化为动能，使得连接 B_0 的弹簧振动起来，这个振动将传向 B 固体的其余部分，最终变成热能耗散掉。这就是摩擦力耗散能量的一个最简单机制。

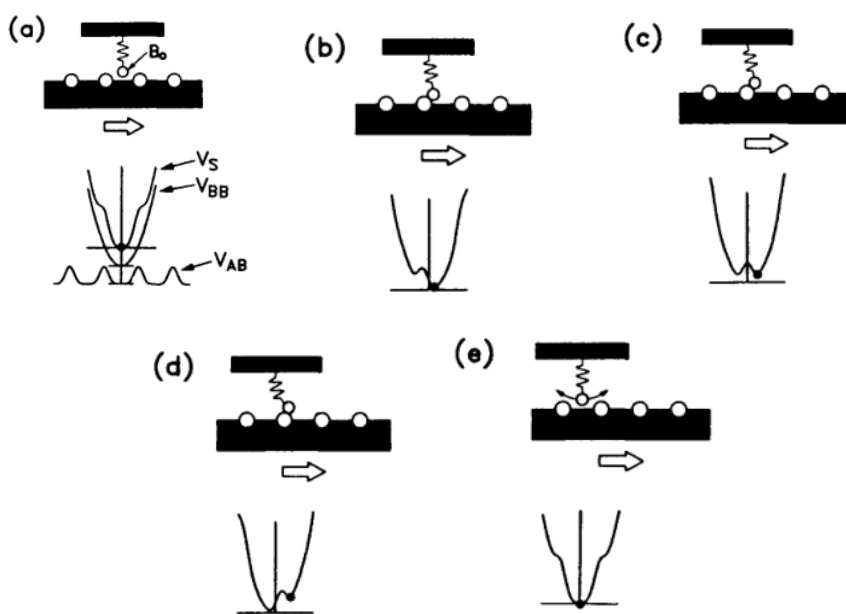


Figure 1.6: 独立振子模型

很显然，在上面的机制中，出现摩擦力的根本原因在于 V_S 出现了两个相互切换的极小，如果 V_S 始终只有一个极小，那就不会有摩擦力。特别的，从图 (1.6) 可以看得很清楚， V_S 有两个极小的这一要求等价于 (c) 图位置时 V_S 有两个极小，这也就是说，等价于在这个位置处 V_S 曲线中间要出现一个极大，即在势能曲线正中间要满足

$$0 > \frac{\partial^2 V_S}{\partial x^2} = \frac{\partial^2 V_{BB}}{\partial x^2} + \frac{\partial^2 V_{AB}}{\partial x^2} = k_{BB} + \frac{\partial^2 V_{AB}}{\partial x^2}. \quad (1.23)$$

这是独立振子模型中出现摩擦力的基本判据，如果这个条件被破坏，那 A 、 B 间的摩擦力就可能消失，出现超滑现象。

1.3 微分形式

前面我们用势能对位置的偏导给出了力的定义，这一定义抓住了自然界中基本相互作用力的本质，我们也看到，满足这一定义的力学系统自动是一个哈密顿系统，因此这样的定义对于理解自然界的基本规律来说非常有用。然而，这一定义却不能包括日常生活中的所有力，比如说，它不能包括宏观表现出来的摩擦力。实际上，这样定义出来的力常常也称之为**保守力**，而日常生活中的力有一些是非保守力，摩擦力就是非保守力，那么保守力有什么特性呢？这一节我们来回答这一问题。但本节更重要的一个目的是介绍一种重要的数学方法，称之为微分形式。

1.3.1 微分形式

微分形式是人们在研究多变量微积分的时候引入的,不妨让我们从两变量微积分开始。假设有一个二元函数 $f(x,y)$, 我们要研究它的二重积分

$$A = \iint_D f(x,y) dx dy, \quad (1.24)$$

D 表示积分区域。假设为了计算这个积分的方便,我们要做如下变量代换(坐标变换)

$$\begin{cases} x = x(x', y') \\ y = y(x', y') \end{cases} \quad (1.25)$$

(x', y') 为另外一组变量(另外一组坐标)。则坐标变换以后积分 A 就变成

$$A = \iint f(x,y) \left| \frac{\partial(x,y)}{\partial(x',y')} \right| dx' dy', \quad (1.26)$$

式中 $\left| \frac{\partial(x,y)}{\partial(x',y')} \right| = \frac{\partial x}{\partial x'} \frac{\partial y}{\partial y'} - \frac{\partial x}{\partial y'} \frac{\partial y}{\partial x'}$ 为坐标变换的雅可比行列式。即是说,坐标变换以后被积函数要多乘上一个雅可比行列式。

关于二元函数积分坐标变换这一结果的证明可以参看任何一本多变量微积分的教材。这里我们想介绍一种巧妙的代数办法来快速导出这一结果。我们将上面二元函数积分的积分微元 $dx dy$ 重写为 $dx \wedge dy$, 当这样写的时候我们实际上给 dx, dy 规定了一种巧妙的代数乘法,乘号写作 \wedge , 通常称之为外积,称相应的代数为外代数,这种代数乘法满足

$$dx \wedge dy = -dy \wedge dx. \quad (1.27)$$

即是说,外积不满足乘法交换律,交换以后会多出一个负号,有时候称之为反交换。由于反交换,我们显然有

$$dx \wedge dx = -dx \wedge dx = 0, \quad dy \wedge dy = -dy \wedge dy = 0. \quad (1.28)$$

有了这个外代数以后,上面二元函数积分的积分微元在坐标变换下就会有

$$\begin{aligned} dx \wedge dy &= \left(\frac{\partial x}{\partial x'} dx' + \frac{\partial x}{\partial y'} dy' \right) \wedge \left(\frac{\partial y}{\partial x'} dx' + \frac{\partial y}{\partial y'} dy' \right) \\ &= \frac{\partial x}{\partial x'} \frac{\partial y}{\partial y'} dx' \wedge dy' + \frac{\partial x}{\partial y'} \frac{\partial y}{\partial x'} dy' \wedge dx' \\ &= \left(\frac{\partial x}{\partial x'} \frac{\partial y}{\partial y'} - \frac{\partial x}{\partial y'} \frac{\partial y}{\partial x'} \right) dx' \wedge dy' \\ &= \left| \frac{\partial(x,y)}{\partial(x',y')} \right| dx' \wedge dy'. \end{aligned} \quad (1.29)$$

上面的推导过程利用了 $dx' \wedge dx' = dy' \wedge dy' = 0$ 。从推导的结果很容易看到,利用外代数,二元函数积分坐标变换多出来的雅可比行列式自动出现了。

以上的结果很容易推广到 n 元函数积分,这时候同样只要将 n 重积分微元 $dx^1 dx^2 \dots dx^n$ 理解为外代数 $dx^1 \wedge dx^2 \wedge \dots \wedge dx^n$, 它们满足

$$dx^i \wedge dx^j = -dx^j \wedge dx^i, \quad (1.30)$$

则坐标变换下多出来的雅可比行列式就会自动出现。我们将被积函数 $f(x^1, x^2, \dots, x^n)$ 和 $dx^1 \wedge dx^2 \wedge \dots \wedge dx^n$ 乘在一起称为一个 n 重微分形式, 简称 n 形式, 记为 ω ,

$$\omega = f(x^1, x^2, \dots, x^n) dx^1 \wedge dx^2 \wedge \dots \wedge dx^n. \quad (1.31)$$

则 n 元函数的 n 重积分实际上就是对 n 形式 ω 的积分, 记为

$$A = \int \omega. \quad (1.32)$$

这里有两点值得说明: 第一, 以后我们往往将多重积分的积分号简单地记为 \int , 然后根据上下文确定它是几重积分。第二, 上面我们是用上指标来区分不同的变量, 下指标我们一会儿也要用到, 读者需要注意上、下指标的区别。

对于 n 个变量的情形, 我们可以推广 n 形式的概念, 定义 k -形式 α , 即 k 重微分形式 α , $0 \leq k \leq n$, α 的定义是

$$\alpha = \frac{1}{k!} \alpha_{i_1 i_2 \dots i_k} dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_k}. \quad (1.33)$$

这里每一个指标的取值都是从 1 到 n , 而且我们既使用了上指标又使用了下指标, 这里就有一个所谓的**求和约定**, 即默认对一个表达式中同时作为上指标和下指标出现的那些指标进行求和, 从 1 求到 n , 而省略掉求和号。我们称 (1.33) 式中的 $\alpha_{i_1 i_2 \dots i_k}$ 为 k 形式 α 的一个分量。由于外代数的反交换性质, 很显然 (1.33) 式中的 k 个指标 i_1, i_2, \dots, i_k 取值必须两两不同, 否则对 α 的贡献将为零。特别的, 这意味着, n -形式是最高重的非零形式, 任何 $k > n$ 的 k 形式都必定为零, 因为这时候它的 k 个指标取值必定会出现重复, 不可能两两不同。

不妨以 2-形式为例进行进一步的考察。这时候 α 的表达式是 $\alpha = \frac{1}{2} \alpha_{ij} dx^i \wedge dx^j$ 。如果 α_{ij} 的两个指标对称, 即 $\alpha_{ij} = \alpha_{ji}$, 则我们就有

$$\alpha_{ij} dx^i \wedge dx^j = \alpha_{ji} dx^j \wedge dx^i = -\alpha_{ji} dx^i \wedge dx^j = -\alpha_{ij} dx^i \wedge dx^j, \quad (1.34)$$

式中第 3 个等号是将求和指标 i, j 重命名了一下 (这不影响求和结果)。即是说, 这时候有 $\alpha = -\alpha$, 从而 $\alpha = 0$ 。由此可见, 对于任何 α_{ij} , 其关于 i, j 对称的部分对于 2 形式 α 的贡献都必定为零, 有贡献的是 i, j 指标反对称的部分。所以, 我们可以自然地要求 α_{ij} 关于指标 i, j 反对称, 即满足 $\alpha_{ji} = -\alpha_{ij}$ 。推广到任意的 k 形式则有, 我们可以自然地要求 k 形式的分量 $\alpha_{i_1 i_2 \dots i_k}$ 关于 k 个下指标两两反对称, 称之为全反对称。

对于 3 维空间这种只有 3 个变量的情形。仅有 0, 1, 2, 3 四种非零的微分形式。0 形式就是一个 3 元的标量函数 $f(x, y, z)$ 。由于反对称性, 3-形式只有一个独立的非零分量, 即 $f(x, y, z) dx \wedge dy \wedge dz$, 所以, 去掉微元 $dx \wedge dy \wedge dz$, 3-形式其实和 0-形式标量函数是等价的。

1-形式可以写成 $a_1 dx + a_2 dy + a_3 dz$, 其三个分量 (a_1, a_2, a_3) 刚好构成一个 3 维空间的矢量场, 也可以记作 $\mathbf{a}(\mathbf{x})$, 所以 3 维空间的 1-形式又可以写作

$$a_1 dx + a_2 dy + a_3 dz = \mathbf{a}(\mathbf{x}) \cdot d\mathbf{x}. \quad (1.35)$$

3 维空间的 2-形式可以写成

$$a = \frac{1}{2}a_{ij}(\mathbf{x})dx^i \wedge dx^j = a_{12}dx \wedge dy + a_{23}dy \wedge dz + a_{31}dz \wedge dx. \quad (1.36)$$

很显然, 它也只有 3 个独立的非零分量, 假设我们记 $a_{12} = b_3$, $a_{23} = b_1$, $a_{31} = b_2$, 并且定义映射 $dx \wedge dy \rightarrow dz$, $dy \wedge dz \rightarrow dx$, $dz \wedge dx \rightarrow dy$ (注意指标 1,2,3 以及微分 dx, dy, dz 各自的轮换), 则有

$$a_{12}dx \wedge dy + a_{23}dy \wedge dz + a_{31}dz \wedge dx \rightarrow b_1dx + b_2dy + b_3dz. \quad (1.37)$$

可见, 3 维空间 2 形式和 1 形式之间能够建立 1-1 对映。因此, 2-形式的 3 个独立非零分量 $(a_{23}, a_{31}, a_{12}) = (b_1, b_2, b_3) = \mathbf{b}$ 刚好构成一个 3 维空间矢量。

同时, 上面这个对映也告诉我们, $(dy \wedge dz, dz \wedge dx, dx \wedge dy)$ 也完全类似于一个 3 维空间的矢量微元, 通常将之定义为面积元矢量 $d\mathbf{S}$,

$$d\mathbf{S} = (dy \wedge dz, dz \wedge dx, dx \wedge dy). \quad (1.38)$$

利用面积元矢量, 我们就可以将三维空间的 2-形式写成

$$a = \frac{1}{2}a_{ij}(\mathbf{x})dx^i \wedge dx^j = \mathbf{b} \cdot d\mathbf{S}. \quad (1.39)$$

类似于前面 0,3 形式以及 1,2 形式之间的对映关系可以推广到 n 维空间, 在 k -形式和 $n-k$ 形式之间建立 1-1 对映关系, 这种对映数学家常常称之为霍奇 (Hodge) 对偶。

1.3.2 外微分与斯托克斯公式

对于微分形式, 我们可以定义一种很巧妙的微分运算, 称作外微分。为了看清楚外微分如何定义, 我们不妨先考察 2 维空间 (以 x, y 为坐标) 中的 1 形式 $a = a_x dx + a_y dy$, 我们定义 a 的外微分 da 为

$$da = da_x \wedge dx + da_y \wedge dy, \quad (1.40)$$

很显然这样定义的外微分是微分运算和外代数运算的结合体。我们来看一下 da 等于什么

$$\begin{aligned} da &= da_x \wedge dx + da_y \wedge dy \\ &= (\partial_x a_x dx + \partial_y a_x dy) \wedge dx + (\partial_x a_y dx + \partial_y a_y dy) \wedge dy \\ &= \partial_y a_x dy \wedge dx + \partial_x a_y dx \wedge dy \\ &= (\partial_x a_y - \partial_y a_x) dx \wedge dy, \end{aligned} \quad (1.41)$$

式中 $\partial_i = \frac{\partial}{\partial x^i}$ 。从计算结果很容易看出来, da 只有一个分量, 即 $(\partial_x a_y - \partial_y a_x)$, 而它刚好是二维矢量 \mathbf{a} 的旋度。

另一方面, 我们知道 2 维空间有所谓的格林公式, 即

$$\oint_{\partial D} (a_x dx + a_y dy) = \int_D (\partial_x a_y - \partial_y a_x) dx dy, \quad (1.42)$$

式中 ∂D 为 2 维空间中的一条闭合回路, D 是这条回路所包围的区域。现在, 利用上面的外微分运算, 显然可以将格林公式用微分形式重写成

$$\int_{\partial D} a = \int_D da. \quad (1.43)$$

这里 ∂D 前面的 ∂ 符号不是表示偏导, 而是表示取 D 的边界。

下面我们来考察 3 维空间 1 形式 $a = \mathbf{a} \cdot d\mathbf{x} = a_x dx + a_y dy + a_z dz$ 的外微分 da , 其定义同样是

$$da = da_x \wedge dx + da_y \wedge dy + da_z \wedge dz. \quad (1.44)$$

经过与两维情形完全类似的计算, 我们可以得到

$$da = (\partial_x a_y - \partial_y a_x) dx \wedge dy + (\partial_y a_z - \partial_z a_y) dy \wedge dz + (\partial_z a_x - \partial_x a_z) dz \wedge dx.$$

这当然是一个 3 维空间中的 2 形式, 这个计算结果清楚地告诉我们, da 的 3 个独立分量刚好够成矢量 \mathbf{a} 的旋度 $\nabla \times \mathbf{a}$ 。从而我们就能把刚才对 da 的计算结果重写成

$$da = (\nabla \times \mathbf{a}) \cdot d\mathbf{S}. \quad (1.45)$$

再一次, 我们看到, 外微分自动给出了 3 维空间的旋度运算。

另一方面, 我们知道, 3 维空间有斯托克斯公式,

$$\oint_{\partial D} \mathbf{a} \cdot d\mathbf{x} = \int_D (\nabla \times \mathbf{a}) \cdot d\mathbf{S} \quad (1.46)$$

式中 ∂D 表示 3 维空间的一条闭合回路, D 是以这条回路为边界的一张曲面。现在, 利用刚才从外微分运算中得到的结果, 我们就能将这个公式用微分形式写成

$$\int_{\partial D} a = \int_D da. \quad (1.47)$$

看, 形式上这个式子和 2 维格林公式的微分形式完全一样!

下面考察 3 维空间中 2 形式 $a = a_{12} dx \wedge dy + a_{23} dy \wedge dz + a_{31} dz \wedge dx$ 的外微分 da , 其定义同样是,

$$da = da_{12} \wedge dx \wedge dy + da_{23} \wedge dy \wedge dz + da_{31} \wedge dz \wedge dx. \quad (1.48)$$

完全类似上面我们进行过的计算, 可以得到

$$da = (\partial_3 a_{12} + \partial_1 a_{23} + \partial_2 a_{31}) dx \wedge dy \wedge dz. \quad (1.49)$$

利用上一小节的定义 $(a_{23}, a_{31}, a_{12}) = (b_1, b_2, b_3) = \mathbf{b}$, 我们可以把这个结果重写成

$$da = (\partial_1 b_1 + \partial_2 b_2 + \partial_3 b_3) dx \wedge dy \wedge dz = (\nabla \cdot \mathbf{b}) dx \wedge dy \wedge dz. \quad (1.50)$$

很显然这是一个 3-形式。从这里我们也看到, 外微分也能给出 3 维矢量的散度。

另一方面，我们知道 3 维空间有所谓的高斯定理，它告诉我们

$$\oint_{\partial V} \mathbf{b} \cdot d\mathbf{S} = \int_V (\nabla \cdot \mathbf{b}) dV, \quad (1.51)$$

式中 3 维体积元 $dV = dx dy dz$ ， ∂V 为 3 维空间中的一个闭合曲面， V 为其包围的 3 维区域。很明显，利用上面关于 2 形式的外微分，以及上一小节的 (1.39) 式，我们可以用 2-形式 a 将高斯定理重写为

$$\int_{\partial V} a = \int_V da. \quad (1.52)$$

又一次，我们得到了与 (1.47) 式形式完全类似的公式。

实际上，我们完全可以将 (1.47) 式和 (1.52) 式综合写成

$$\int_{\partial D} \alpha = \int_D d\alpha, \quad (1.53)$$

式中 α 表示 3 维空间中的一个 $k-1$ 形式， D 表示 3 维空间中一个以 ∂D 为边界的 k 维曲面 (因此 ∂D 是 $k-1$ 维的，而 $d\alpha$ 则是一个 k -形式)。 $k=2$ 时，它就是斯托克斯公式， $k=3$ 时它就是高斯定理。可见，利用外微分运算，我们可以将矢量分析中那些著名的公式和定理统一起来。

不仅如此，我们还可以将上面的结果推广到 n 维空间，对于 n 维空间的一个 $k-1$ 形式 $\alpha = \frac{1}{(k-1)!} \alpha_{i_1 i_2 \dots i_{k-1}} dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_{k-1}}$ ，我们可以一般性地定义其外微分 $d\alpha$ 为

$$d\alpha = \frac{1}{(k-1)!} (\partial_j \alpha_{i_1 i_2 \dots i_{k-1}}) dx^j \wedge dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_{k-1}}. \quad (1.54)$$

显然， $d\alpha$ 是一个 k 形式。可以证明 (请参考相关数学书)，我们将依然有 (1.53) 式！和前面的唯一区别是，现在，它不再限于 3 维空间中了，因此 k 也不限于小于等于 3 了，而是 $k \leq n$ 。正因为如此，人们通常称 (1.53) 式为广义的斯托克斯公式，或者有时候就简称为斯托克斯公式。

外微分有一个非常优雅而重要的性质，即对任何微分形式进行两次外微分，结果恒等于零，通常将这个结果简写成

$$d^2 = 0. \quad (1.55)$$

它的含义就是对于任何 $k-1$ 形式 α ，必有

$$d^2 \alpha = d(d\alpha) = 0. \quad (1.56)$$

这个结果的证明非常简单，按照外微分的定义

$$d^2 \alpha = \frac{1}{(k-1)!} (\partial_i \partial_j \alpha_{i_1 i_2 \dots i_{k-1}}) dx^i \wedge dx^j \wedge dx^{i_1} \wedge dx^{i_2} \wedge \dots \wedge dx^{i_{k-1}}. \quad (1.57)$$

显然这是一个 $k+1$ 形式，但是这个 $k+1$ 形式的分量 $\partial_i \partial_j \alpha_{i_1 i_2 \dots i_{k-1}}$ 关于 i, j 指标恒对称，而不是反对称，按照我们前面的分析，这就意味着这个 $k+1$ 形式等于零，从而就完成了证明。

下面介绍两个常用概念。首先，一个微分形式 α ，如果它的外微分等于零，即 $d\alpha = 0$ ，我们就称它为闭形式。其次，一个微分形式 α ，如果它是另一个微分形式 β 的外微分，即有 $\alpha = d\beta$ ，我们就称这样的 α 为一个恰当形式。根据性质 (1.55)，很显然，任何恰当形式都必定是闭形式！反过来，闭形式却不一定是恰当形式，闭形式什么时候是恰当形式什么时候不是，这往往和空间的拓扑有关系，是所谓的 de Rahm 上同调研究的内容，不过这超出本书的范围了。

1.3.3 保守力的特性

微分形式的语言在我们后面的章节中还有重要的应用，不过现在还是让我们用它来研究一下保守力的特性吧。

让我们先回顾一下前面给出的保守力的定义

$$\sum_i \mathbf{F}_i \cdot d\mathbf{x}_i = -dV. \quad (1.58)$$

为了将这个定义写得更漂亮一点，我们引入指标 $\mu = 1, 2, 3, \dots, 3N$ ，令它的前 3 个分量表示第 1 个粒子的三个直角坐标分量，次 3 个分量表示第 2 个粒子的三个直角坐标分量，依次类推，直到最后 3 个分量代表第 N 个粒子的三个直角坐标分量。从而我们可以将上面这个保守力的定义式重写成

$$F_\mu dx^\mu = -dV(x^1, \dots, x^{3N}). \quad (1.59)$$

当然，在这个式子中我们用了求和约定。我们可以称这样定义出来的 $F_\mu dx^\mu$ 为力 1 形式，简记为 F ，即 $F = F_\mu dx^\mu$ ，则上面这个保守力的定义式告诉了我们，保守力 1 形式是一个恰当形式，满足

$$F = -dV, \quad (1.60)$$

势能 V 是一个标量函数，也就是 0 形式，也称作势能 0 形式。则由恰当形式必定是闭形式的数学定理，我们立马知道

$$dF = 0. \quad (1.61)$$

根据外微分的定义，我们很容易计算 dF (下面的 $\partial_\mu = \frac{\partial}{\partial x^\mu}$)，

$$dF = (\partial_\mu F_\nu) dx^\mu \wedge dx^\nu = \left[\frac{1}{2}(\partial_\mu F_\nu - \partial_\nu F_\mu) + \frac{1}{2}(\partial_\mu F_\nu + \partial_\nu F_\mu) \right] dx^\mu \wedge dx^\nu. \quad (1.62)$$

根据微分形式分量表达的指标对称部分没有贡献，我们立马有

$$dF = \frac{1}{2}(\partial_\mu F_\nu - \partial_\nu F_\mu) dx^\mu \wedge dx^\nu. \quad (1.63)$$

因此 $dF = 0$ 就等价于

$$\partial_\mu F_\nu - \partial_\nu F_\mu = 0. \quad (1.64)$$

特别的，对于单个质点的情形（指标只能取 1,2,3），读者容易知道，上面这个结果其实就是力矢量 \mathbf{F} 的旋度等于零，即

$$\nabla \times \mathbf{F} = 0. \quad (1.65)$$

另一方面，注意到力 1 形式 F 是 $3N$ 维坐标空间的 1 形式，我们可以在 $3N$ 维坐标空间取一条闭合回路 ∂D ， ∂D 是 $3N$ 维坐标空间中某个 2 维曲面 D 的边界，则由斯托克斯公式，我们有

$$\int_{\partial D} F = \int_D dF = 0. \quad (1.66)$$

即是说，保守力 1 形式在坐标空间任何闭合回路上的积分都等于零。人们通常称力 1 形式在某条路径上的积分为**功**，因此这个结果就相当于说，保守力在任何闭合回路上所做的功恒为零！

由此我们容易知道，摩擦力必定不是保守力，因为在摩擦力的作用下将粒子沿着闭合回路运动一圈，摩擦力是要做非零的负功的，正是这个负功将机械能耗散为了热量。



2. 最小作用量原理

本章是经典力学的核心章节之一。本章我们将引入相空间的最小作用量原理，并证明它与哈密顿正则方程等价。我们也会从相空间的最小作用量原理导出坐标空间的最小作用量原理，并引入拉格朗日量的概念。最后，我们将通过引入广义坐标和广义动量，讲述如何将这两章发展起来的理论框架应用于约束系统。

在数学方法上，本章将通过费马原理引入泛函和变分，讲述变分法和泛函导数的基本思想，推导变分法中的欧拉-拉格朗日方程。

2.1 物理学的局部视角与全局视角

上一章我们引入了相空间，我们说粒子在相空间中按照哈密顿正则方程演化，对于单粒子这个方程是

$$\frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{x}}, \quad \frac{d\mathbf{x}}{dt} = \frac{\partial H}{\partial \mathbf{p}}. \quad (2.1)$$

这一章我们将介绍一种表面看起来完全不同的观点，叫做最小作用量原理 (principle of least action)。按照最小作用量原理，粒子在相空间中不是按照哈密顿正则方程这样的微分方程演化，粒子是按“代价”最小的相空间路径演化。即是说，粒子的演化路径有无穷多种可能性，每一条可能的演化路径都要付出一个相应的“代价”，而粒子的真实演化路径是所有可能路径中“代价”最小的那条，严格一点说应该是“代价”取极值的那条（**更严格地说应该是“代价”取驻值的那条，但是驻值这个概念比较微妙，本书中为了使得读者更好理解，我们统一称作极值**）。每一条路径的“代价”就叫做这条路径的作用量，记为 S ，它由下式给出

$$S[\mathbf{x}(t), \mathbf{p}(t)] = \int [\mathbf{p} \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{p})] dt. \quad (2.2)$$

式中时间 t 的函数 $(\mathbf{x}(t), \mathbf{p}(t))$ 代表一条任意给定的相空间路径，注意，这条路径无须满足哈密顿正则方程¹，因为它只是一条可能路径，不一定是粒子的真实演化路径。记号 $S[\mathbf{x}(t), \mathbf{p}(t)]$ 是为了强调，作用量依赖于路径，每一条路径都有一个相应的作用量值，我们可以把作用量看成是路径的函数，路径是这种函数的自变量，很显然，路径这种自变量不同于通常微积分教科书里的自变量，因为它本身是一个函数，是 t 的函数。这种以函数为自变量的函数就称作**泛函**！所以作用量是路径的泛函，由 (2.2) 式给出。

很显然，最小作用量原理看起来与哈密顿正则方程完全不同，结果却可以证明，这两种描述粒子在相空间中如何演化的方式物理上完全等价。哈密顿正则方程是一种局部视角的描述方式，每一个时刻都只需用到当前相点（相空间点）局部邻域内的信息，因为微分方程中的求导运算只涉及邻域。而最小作用量原理是一种全局视角的描述方式，需要知道每一条可能路径的作用量这种全局信息。奇妙的是，这两种不同视角在物理上却是等价的。人们有时候将微分方程这样的局部视角称作蚂蚁视角（蚂蚁太小，每一只蚂蚁都只能看到一个很小的邻域），而将最小作用量原理这样的全局视角称作上帝视角。看待物理的蚂蚁视角源自于牛顿，正是牛顿想到用微分方程来描述物理规律。而上帝视角则源自于几何光学中的费马原理，然后经过莫培督 (Maupertuis) 和哈密顿等人推广到力学里来。实际上，最小作用量原理有时候也称作哈密顿原理。不可思议的是，我们可以用这样两种完全不同的视角来看待同样的物理。

不仅如此，这种局部视角和全局视角也都可以延伸到量子物理中，在量子力学中，局部视角大致会导致算符描述，而全局视角会导致路径积分描述。

¹因此，一般来说没有 $\mathbf{p} = m\dot{\mathbf{x}}$ 这个关系只在真实路径上成立。

2.2 从费马原理到变分法

既然最小作用量原理源自费马原理的推广，不妨让我们先来回顾一下费马原理。当然，这一节更主要的目的是引入变分法这种处理最小作用量原理的数学方法。

2.2.1 费马原理

1657 年，业余数学王子费马在古希腊数学家希罗 (Hero of Alexandria) 的最短路径原理基础上提出，光走时间最短的路径，这就是几何光学的费马原理。假设真空中的光速为 c ，介质的折射率为 $n(\mathbf{x})$ ，相邻两点之间的距离为 $dl = |d\mathbf{x}|$ ，则费马原理说的是，给定初末两点，光走时间 t 最小的路径，

$$t = \int \frac{dl}{v} = \frac{1}{c} \int n(\mathbf{x}) |d\mathbf{x}|. \quad (2.3)$$

式中 $v = c/n$ 为介质中的光速，式中的积分沿着光的路径进行。可以证明，从费马原理出发，能够导出几何光学的所有定律，包括光的反射定律、折射定律，以及透镜成像的规律等等。具体的证明可以参见《费恩曼物理学讲义》第一卷，第 26，27 章。由于 c 是 (2.3) 式中除以一个整体常数，所以费马原理也可以等价地说成是，光走光程最短的路径，光程 S 为

$$S = \int n(\mathbf{x}) |d\mathbf{x}|. \quad (2.4)$$

为了说明数学上该如何处理费马原理，不妨考虑一种特殊情况，设光线在 $x-y$ 平面内运动，再设折射率仅仅依赖于 y 坐标，则光程的表达式为

$$\begin{aligned} S &= \int n(y) |d\mathbf{x}| = \int n(y) \sqrt{(dx)^2 + (dy)^2} \\ &= \int n(y) \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx \\ &= \int n(y) \sqrt{1 + (y')^2} dx. \end{aligned} \quad (2.5)$$

式中 $y' = \frac{dy}{dx}$ ，在上式的最后两步中我们用函数 $y(x)$ 来描写光在 $x-y$ 平面上的可能路径。另外，由于光线的起末两点是给定的，假设起点的 x 坐标为 a ， y 坐标为 y_a ，终点的 x 坐标为 b ， y 坐标为 y_b ，则这意味着光的任何可能路径 $y(x)$ 都以这两点为端点，即满足

$$y(a) = y_a, \quad y(b) = y_b \quad \text{fixed} \quad (2.6)$$

根据费马原理，光真实所走的路径 $y_c(x)$ 应该是所有满足 (2.6) 式的两端固定的函数 $y(x)$ 中，使得 (2.5) 式给出的光程 $S[y(x)]$ 取极小值的那一个特定函数，如图 (2.1) 所示。光程 S 依赖于可能路径 $y(x)$ ，是路径 $y(x)$ 的泛函，所以记作 $S[y(x)]$ 。问题是，怎么找到这个特定函数呢？

从 (2.5) 式可以看出，光程 S 的被积函数的表达式既依赖于 $y(x)$ ，又依赖于 $y'(x)$ 。我们不妨抽象一步，将这个被积表达式写成 $L(y, y')$ ，在这个特定问题中 $L(y, y') = n(y) \sqrt{1 + (y')^2}$ ，

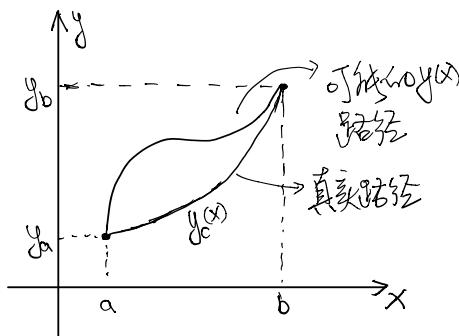


Figure 2.1: 光的路径

但是下文的讨论将不限于这一特定的 L 函数。所以我们要做的就是所有两端由 (2.6) 式给出的函数 $y(x)$ 中，找出使得下面的表达式 $S[y(x)]$ 取极值的那个特定函数，

$$S[y(x)] = \int_a^b dx L(y, y'). \quad (2.7)$$

这就把费马原理变成了一个纯粹的数学问题，这个数学问题最早是欧拉和拉格朗日解决的。欧拉和拉格朗日所发展的方法就是我们即将介绍的变分法，它的本质其实就是多变量微分，只不过这个多变量是多到无穷个实变数！

2.2.2 变分和泛函导数

为了更好地理解变分法，让我们回顾一下多变量求导和微分。假设有多变量函数 $S(y_1, y_2, \dots)$ ，这里自变量 y_i 就是普通的实变数。假设我们要求这个函数的极值，显然这个极值由偏导等于零给出，即 $\frac{\partial S}{\partial y_i} = 0, i = 1, 2, \dots$ 。函数的偏导如何定义呢？一种定义方法是通过微分，即将

$$dS = S(y_1 + dy_1, y_2 + dy_2, \dots) - S(y_1, y_2, \dots), \quad (2.8)$$

展开到各无穷小量 $dy_i, i = 1, 2, \dots$ 的一阶，一阶无穷小量前面的展开系数就是 $\frac{\partial S}{\partial y_i}$ ，即是说，我们有

$$dS(y_1, y_2, \dots) = \sum_i \frac{\partial S}{\partial y_i} dy_i. \quad (2.9)$$

因此多变量函数的极值条件是

$$dS = 0. \quad (2.10)$$

也即是说，当且仅当在自变量的无穷小改变之下函数值保持不变时，函数才取到极值。

下面我们做两个很平凡的操作：第一，考虑到自变量很多，我们把函数 $S(y_1, y_2, \dots)$ 重记为 $S[y_i]$ ，当然这里的 i 不是一个固定的指标，而是要取遍所有的自变量。第二，我们把各自变量的指标 i 改写成 x ，则按照刚才的记号，这个多变量函数就应该记为 $S[y_x]$ ，而上面的两个方程 (2.8) 和 (2.9) 就应该重写成

$$dS[y_x] = S[y_x + dy_x] - S[y_x], \quad dS[y_x] = \sum_x \frac{\partial S}{\partial y_x} dy_x. \quad (2.11)$$

以上还是多变量求导和微分，下面是关键的一步，我们设想自变量如此之多，以至于将这些自变量放在一起它们布满了连续的 x 轴。换言之，我们假设上面的指标 x 是一个连续指标。就微积分的实质而言连续指标和离散指标并没有什么不同，但在数学形式上的确有些不同，比方说对离散变量的求和过渡到连续情形那就应该变成积分。为了强调连续指标的不同，我们将 (2.11) 式中所有的微分号 (包括偏导符号) 换一种写法，写成 δ (其本质含义依然是一阶无穷小)，然后将求和变成积分。这样一来，(2.11) 式就变成了

$$\delta S[y_x] = S[y_x + \delta y_x] - S[y_x], \quad \delta S[y_x] = \int dx \frac{\delta S}{\delta y_x} \delta y_x. \quad (2.12)$$

下面是最后一步，既然 x 是连续指标， y_x 当然也可以看成是 x 的一个函数，重写为 $y(x)$ 。即是说，连续多个自变量刚好可以看成是以函数 $y(x)$ 为自变量，而刚才的记号 $S[y_x]$ 也就成了 $S[y(x)]$ ，它表示 S 是依赖于未定函数 $y(x)$ 的，是以未定函数 $y(x)$ 为自变量的某种“函数”，我们知道，这也就是泛函。而 (2.12) 式中的第一个方程也就变成了，

$$\delta S[y(x)] = S[y(x) + \delta y(x)] - S[y(x)], \quad (2.13)$$

这称作泛函的变分，按照定义它要展开到无穷小函数 $\delta y(x)$ 的一阶项。这个式子的含义就是，当作为自变量的函数 $y(x)$ 发生无穷小改变 $\delta y(x)$ 时，泛函 $S[y(x)]$ 的改变量。当然，通过上面的叙述过程我们也看到，变分普通的微分本质是一样的，只不过现在的自变量按照普通的实变数来理解有连续统的无穷多个。当然，现在函数求极值的方程 (2.10) 也就变成了泛函取极值的方程，

$$\delta S[y(x)] = 0. \quad (2.14)$$

即是说，泛函取极值时，泛函的值在自变量函数的无穷小改变下保持不变。

同样的，(2.12) 式中的第二个方程现在就变成了

$$\delta S[y(x)] = \int dx \frac{\delta S}{\delta y(x)} \delta y(x). \quad (2.15)$$

式中无穷小量 $\delta y(x)$ 前面的系数 $\frac{\delta S}{\delta y(x)}$ 就叫做泛函导数，很显然，它和普通导数本质完全一样，只不过现在的自变量按照普通的实变数来理解有连续统的无穷多个。而上面式子中对 x 的积分本质不过就是对所有的变量求和。

既然变分的本质其实就是微分，那它当然也满足微分的那些规则，比方说莱布尼兹法则，即假设 f, g 为两个关于 $y(x)$ 的泛函，则

$$\delta(fg) = (\delta f)g + f(\delta g). \quad (2.16)$$

但是变分并不是普通的微分，普通的微分和积分是对变量 x 进行的，而 x 在变分中不过是一个指标，真正的变量是 $y(x) = y_x$ ，变分是对 y_x 进行的。这就产生了一些新的东西，因为我们当然可以同时指标 x 进行普通的微分和积分，由于这些操作是作用在指标上，它当然和对 y_x 本身的“微分” (即变分) 操作可交换顺序。由此我们知道，**变分可以和普通的微分与积分交换顺序！**

最简单的关于 $y(x)$ 的泛函就是 $y(x)$ 本身, 具体来说, 我们将 $y(x)$ 理解为函数 y 在某个特定 x 处的取值, 它当然依赖于函数 y 本身, 所以是一个泛函。对这个泛函进行变分, 即有

$$\delta y(x) = \int dx' \delta(x-x') \delta y(x'), \quad (2.17)$$

式中 $\delta(x-x')$ 为狄拉克的 δ 函数 (参见数理方法的书)。将这个结果与泛函导数的定义式 (2.15) 比较容易知道

$$\frac{\delta y(x)}{\delta y(x')} = \delta(x-x'). \quad (2.18)$$

我们想用这个例子说明泛函导数的计算都是怎么进行的, 这个例子说明, 除了可以应用泛函导数对应的求导规则 (比如链式法则) 之外, 往往都是通过变分进行的。

2.2.3 欧拉-拉格朗日方程

下面回到我们在费马原理的例子中提出来的变分问题, 我们要求的是下面泛函的极值

$$S[y(x)] = \int_a^b dx L(y, y'). \quad (2.19)$$

当然, 这里有一个额外的限制条件, 即所有可能函数 $y(x)$ 在 a, b 两端都是固定的, 也即是说, 当我们对 $y(x)$ 进行变分时, 这两端要固定不变, 即

$$\delta y(a) = \delta y(b) = 0. \quad (2.20)$$

上面我们说过, 泛函取极值的条件是变分等于零, 由此我们就可以解出费马原理提出的变分问题。下面我们来看这个过程具体如何进行。

首先我们注意到 $L(y, y')$ 关于 y 和 y' 是普通的函数关系, 因此按照变分的定义, 我们有

$$\begin{aligned} \delta L(y, y') &= L(y + \delta y, y' + \delta y') - L(y, y') \\ &= \frac{\partial L}{\partial y} \delta y + \frac{\partial L}{\partial y'} \delta y' \\ &= \frac{\partial L}{\partial y} \delta y + \frac{\partial L}{\partial y'} (\delta y)'. \end{aligned} \quad (2.21)$$

最后一个等号我们利用了变分可以和普通的求导交换顺序。由此我们有

$$\begin{aligned} \delta S[y(x)] &= \int_a^b dx \delta L(y, y') \\ &= \int_a^b dx \left[\frac{\partial L}{\partial y} \delta y + \frac{\partial L}{\partial y'} (\delta y)' \right] \\ &= \int_a^b dx \left[\frac{\partial L}{\partial y} \delta y + \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \delta y \right) - \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \right) \delta y \right] \\ &= \int_a^b dx \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \delta y \right) + \int_a^b dx \left[\frac{\partial L}{\partial y} - \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \right) \right] \delta y \\ &= \frac{\partial L}{\partial y'} \delta y \Big|_a^b + \int_a^b dx \left[\frac{\partial L}{\partial y} - \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \right) \right] \delta y \\ &= \int_a^b dx \left[\frac{\partial L}{\partial y} - \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \right) \right] \delta y. \end{aligned} \quad (2.22)$$

式中第三个等号我们利用了通常的分部积分，最后一个等号我们利用了 $y(x)$ 的两端固定，即利用了 (2.20) 式。所以由泛函取极值的条件 $\delta S = 0$ ，我们可以导出

$$\delta S = \int_a^b dx \left[\frac{\partial L}{\partial y} - \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \right) \right] \delta y = 0. \quad (2.23)$$

但是，变分 $\delta y(x)$ 作为一个普通函数来看的话是一个取无穷小函数值但是函数形式任意的函数，对于任意函数都有上式成立的话，那当且仅当

$$\frac{\delta S}{\delta y(x)} = \frac{\partial L}{\partial y} - \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \right) = 0. \quad (2.24)$$

这就是著名的欧拉-拉格朗日方程。通过上面的推导我们知道，使得泛函取极值的那个特定函数 $y_c(x)$ 必然要满足这个欧拉-拉格朗日方程，反过来也一样，满足欧拉-拉格朗日方程的函数必然使得泛函 $S[y(x)]$ 取极值。

比方说，对于前面费马原理中提到的特定函数 $L(y, y') = n(y) \sqrt{1 + (y')^2}$ ，我们代入欧拉-拉格朗日方程，就可以导出光的路径所要满足的方程，即

$$\begin{aligned} \frac{dn}{dy} \sqrt{1 + (y')^2} - \frac{d}{dx} \left(n(y) \frac{y'}{\sqrt{1 + (y')^2}} \right) &= 0. \\ \Rightarrow (1 + (y')^2) \frac{dn}{dy} - n(y) y'' &= 0. \end{aligned} \quad (2.25)$$

进而容易验证必有（不要试图去推导，反过来验证比较简单）

$$\frac{d}{dx} \left(\frac{n(y)}{\sqrt{1 + (y')^2}} \right) = 0 \Leftrightarrow \frac{n(y)}{\sqrt{1 + (y')^2}} = C. \quad (2.26)$$

其中 C 为常数。剩下的就是求解最后这个一阶微分方程，不过这不是我们这里关心的问题。考察一个最简单的特殊情况吧，假设光线是在真空中传播，从而 $n(y) = 1$ 。这时候上面最后的方程告诉我们

$$y' = k \Leftrightarrow y = kx + a, \quad (2.27)$$

式中 k, a 均为常数。可见，这时候光走的正是一条直线！不仅如此，随着折射率函数 $n(y)$ 的不同，我们还可以得到最速降线和悬链线等等。

但是，我们刚才处理的是一个特殊的费马原理问题，对于一般的费马原理问题，光程函数由 (2.4) 式给出。为了将这一问题转化为一个变分问题，我们引入参数 s 来将光的路径参数化，从而得到光程泛函 $S[\mathbf{x}(s)]$,

$$S[\mathbf{x}(s)] = \int_{s_i}^{s_f} n(\mathbf{x}) \left| \frac{d\mathbf{x}}{ds} \right| ds. \quad (2.28)$$

我们要做的就是用变分法求解这个泛函的极值，同样，这里要求所有可能路径的端点是固定的，即在起末两端有

$$\delta \mathbf{x}(s_i) = \delta \mathbf{x}(s_f) = 0. \quad (2.29)$$

我们发现，上面这个极值问题的泛函 $S[\mathbf{x}(s)]$ 具有如下结构

$$S[\mathbf{x}(s)] = \int_{s_i}^{s_f} ds L(\mathbf{x}, \dot{\mathbf{x}}), \quad (2.30)$$

式中 $\dot{\mathbf{x}} = \frac{d\mathbf{x}}{ds}$ ，函数 $L(\mathbf{x}, \dot{\mathbf{x}}) = n(\mathbf{x})|\dot{\mathbf{x}}|$ 。很显然，这个泛函和我们刚才在推导欧拉-拉格朗日方程时处理的泛函具有完全类似的结构，只要将刚才的 x 替换成这里的 s ，将之前的 $y(x)$ 替换成这里的 $\mathbf{x}(s)$ 就可以了。所以我们有完全类似的欧拉-拉格朗日方程，

$$\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{ds} \left(\frac{\partial L}{\partial \dot{\mathbf{x}}} \right) = 0. \quad (2.31)$$

2.3 相空间的最小作用量原理

我们知道所谓的相空间，就是粒子的位置坐标和动量放在一起所构成的空间，对于单粒子体系，相空间就是 (\mathbf{x}, \mathbf{p}) 空间，粒子在相空间中按照哈密顿正则方程演化。然而最小作用量原理提供了一种看起来完全不同的观点，它说，粒子在相空间的真实演化路径是所有可能路径中使得作用量 S 取极值的那条路径，作用量 S 由下式给出

$$S[\mathbf{x}(t), \mathbf{p}(t)] = \int_{t_i}^{t_f} [\mathbf{p} \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{p})] dt. \quad (2.32)$$

读者很容易看到，这个泛函极值问题具有 $S[\mathbf{x}(t), \mathbf{p}(t)] = \int_{t_i}^{t_f} dt L(\mathbf{x}, \mathbf{p}, \dot{\mathbf{x}})$ 的结构，看起来这正好是欧拉-拉格朗日方程能够处理的泛函极值问题。当然，需要我们将前面推导欧拉-拉格朗日方程时的 x 替换成这里的 t ，将 $y(x)$ 替换成这里的 $(\mathbf{x}(t), \mathbf{p}(t))$ 。事实上，如果直接代欧拉-拉格朗日方程的话的确也能得到正确结果，我们会发现这时候欧拉-拉格朗日方程正好给出了哈密顿正则方程。但这样做实际上掩盖了一个微妙的问题。

问题就是，最小作用量原理的这个泛函极值问题实际上和费马原理那一类泛函极值问题有一点微妙的区别，那就是，在这里可能的相空间路径 $(\mathbf{x}(t), \mathbf{p}(t))$ 的两个端点不能完全固定！实际上，如果我们任意固定 t_i 处的起端和 t_f 处的末端，任意指定这两端的 \mathbf{x} 和 \mathbf{p} ，那这个泛函极值问题将是无解的。基本原因在于，前面欧拉-拉格朗日方程给出的都是二阶微分方程，而在这里这个泛函极值路径所要满足的哈密顿正则方程则是一组一阶微分方程，要确定这样的一阶微分方程的解我们只能要么给定它在起端（即初始时）的值，要么给定它在末端的值，而不能同时任意指定它的初始值和末尾值！当然，在这里，完全固定起端或者完全固定末端都不是我们的正确选择，正确的做法是固定起末两端的 \mathbf{x} 坐标，但是，完全不限制这两端的 \mathbf{p} 坐标，也就是只要求

$$\delta \mathbf{x}(t_i) = \delta \mathbf{x}(t_f) = 0. \quad (2.33)$$

即是说，在现在的泛函极值问题中，可能的相空间路径如图 (2.2) 所示。

为了说明以上处理的确能解决问题，我们不妨按照变分法的步骤具体推导一下，由变

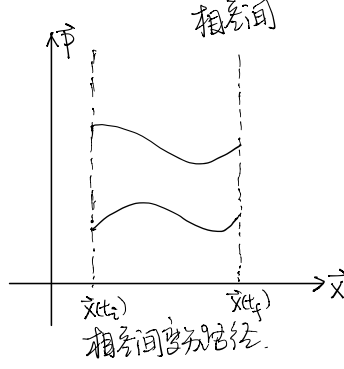


Figure 2.2: 可能的相空间路径。注意时间 t 我们没有画出。

分法

$$\begin{aligned}
 \delta S[\mathbf{x}(t), \mathbf{p}(t)] &= \int_{t_i}^{t_f} [\delta \mathbf{p} \cdot \dot{\mathbf{x}} + \mathbf{p} \cdot \delta \dot{\mathbf{x}} - \delta H(\mathbf{x}, \mathbf{p})] dt. \\
 &= \int_{t_i}^{t_f} [\delta \mathbf{p} \cdot \dot{\mathbf{x}} + \frac{d}{dt} (\mathbf{p} \cdot \delta \mathbf{x}) - \dot{\mathbf{p}} \cdot \delta \mathbf{x} - \delta H(\mathbf{x}, \mathbf{p})] dt \\
 &= \mathbf{p} \cdot \delta \mathbf{x} \Big|_{t_i}^{t_f} + \int_{t_i}^{t_f} [\delta \mathbf{p} \cdot \dot{\mathbf{x}} - \dot{\mathbf{p}} \cdot \delta \mathbf{x} - \delta H(\mathbf{x}, \mathbf{p})] dt \\
 &= \int_{t_i}^{t_f} [\delta \mathbf{p} \cdot \dot{\mathbf{x}} - \dot{\mathbf{p}} \cdot \delta \mathbf{x} - \frac{\partial H}{\partial \mathbf{x}} \cdot \delta \mathbf{x} - \frac{\partial H}{\partial \mathbf{p}} \cdot \delta \mathbf{p}] dt \\
 &= \int_{t_i}^{t_f} \left[\left(\dot{\mathbf{x}} - \frac{\partial H}{\partial \mathbf{p}} \right) \cdot \delta \mathbf{p} - \left(\dot{\mathbf{p}} + \frac{\partial H}{\partial \mathbf{x}} \right) \cdot \delta \mathbf{x} \right] dt \quad (2.34)
 \end{aligned}$$

最后由泛函极值方程 $\delta S = 0$ ，就能得到

$$\frac{\delta S}{\delta \mathbf{p}(t)} = \dot{\mathbf{x}} - \frac{\partial H}{\partial \mathbf{p}} = 0, \quad \frac{\delta S}{\delta \mathbf{x}(t)} = -\left(\dot{\mathbf{p}} + \frac{\partial H}{\partial \mathbf{x}} \right) = 0. \quad (2.35)$$

当然，这正是哈密顿正则方程，即是说，使得作用量取极值的相空间路径正是满足哈密顿正则方程的路径。由此我们就证明了，最小作用量原理这种看起来完全不同的全局视角，其结果和力学规律的微分方程视角完全等价！值得提醒读者的是，对于一般的哈密顿系统，动量 \mathbf{p} 与速度 $\dot{\mathbf{x}}$ 之间的关系将不一定是 $\mathbf{p} = m\dot{\mathbf{x}}$ ，两者的关系应该由以上哈密顿正则方程的第一个方程 $\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}}$ 去确定。

人们很容易将上面的最小作用量原理推广到多粒子情形。假设有 N 个粒子，每个粒子 3 个坐标，因此共有 $3N$ 个坐标，我们可以用指标 $\mu = 1, 2, \dots, 3N$ 来区分它们，相应的位置坐标就标记为 x^μ ，动量坐标就标记为 p_μ （注意动量是下指标，位置是上指标），利用上下指标的求和约定，就可以把多粒子的相空间作用量写成

$$S[x(t), p(t)] = \int_{t_i}^{t_f} dt [p_\mu \dot{x}^\mu - H(x, p)]. \quad (2.36)$$

式中 $H(x, p)$ 是 $H(x^1, \dots, x^{3N}, p_1, \dots, p_{3N})$ 的简记符号， $S[x(t), p(t)]$ 也是作类似理解。

完全类似的变分，就能得出多粒子情形的哈密顿正则方程，如下

$$\dot{x}^\mu = \frac{\partial H}{\partial p_\mu}, \quad \dot{p}_\mu = -\frac{\partial H}{\partial x^\mu}. \quad (2.37)$$

注意这两个式子等号左右两边的指标对应关系，即，对一个上指标的求导，其结果相当于一个下指标的求导，而对一个下指标的求导，其结果则相当于一个上指标的求导。

所以，只要写出相空间作用量，我们就能得到哈密顿正则方程，但是从相空间作用量的一般表达式可以知道，写出具体的作用量关键在于写出具体的哈密顿量，一个系统的哈密顿量怎么写呢？简单来说就是所有对系统能量有贡献的都要包括进哈密顿量。不过，也有很多系统，人们甚至无法判断有哪些项对系统能量有贡献，这时候想写出哈密顿量并不容易，因此写出相空间作用量的具体形式当然也不容易。好在，马上我们会介绍一个坐标空间（或者说位形空间）的最小作用量原理，它里面用到的作用量是表达在坐标空间的作用量，这时候，人们往往可以根据系统所满足的对称性直接写出它的具体表达式（即是说，**对称性限制作用量**），有了这个作用量在坐标空间的表达式，随之就能得到运动微分方程，甚至还能反过来得到系统哈密顿量的表达式。关于这些内容，我们会在后面有关对称性的章节中具体讨论。

下面的一个问题是，作为一个物理量，作用量的量纲是什么？回答很简单，**作用量的量纲等于能量量纲乘以时间量纲，或者说也可以说等于动量量纲乘以长度量纲**。简言之，作用量与角动量同量纲。为了看清楚这一点，只需注意到作用量等于 $\int p\dot{x} - H$ 对时间的积分， $p\dot{x}$ 和能量同量纲，而哈密顿量当然是能量量纲，所以 $p\dot{x} - H$ 也是能量量纲，再对时间积分，就是能量量纲乘以时间量纲。

推广到显含 t 情形

到此为止，我们所有的讨论都限于封闭系统，它的总能量是守恒的。但是，我们也很容易将相关的讨论推广到非封闭系统，这时候总可以把所关心的系统和与它相互作用着的其它部分看成一个整体，而这个整体当然是一个封闭系统，因此前面所有的讨论对这个整体都成立。现在，假设我们已经了解了其它部分的运动情况，而将注意力集中在所关心的系统上，这时候那些其它部分的力学变量就可以看成是一些已知的随时间变化的参数。由于有了这些含时参数，我们就需要假设哈密顿量 H 可能显含时间，即是说，系统之外其它部分的运动使得系统的能量随时间变化了！

但非封闭系统除了使得哈密顿量显含时间之外，它并不会改变相空间作用量的一般结构，从而由最小作用量原理，我们依然有哈密顿正则方程。当然，这时候 $\frac{dH}{dt} = \frac{\partial H}{\partial t} \neq 0$ ，因此一般来说系统的能量是不守恒的。

2.4 坐标空间的最小作用量原理

前面我们介绍了相空间的最小作用量原理，但是，通常人们在谈最小作用量原理时，指的其实是坐标空间的最小作用量原理！现在让我们来介绍它。值得强调的是，坐标空间的最小作用量原理并不是一个新的假定，实际上，它可以从相空间的最小作用量原理推导出来。

2.4.1 最小作用量原理与拉格朗日方程

为了理解如何从相空间的最小作用量原理走向坐标空间的最小作用量原理，我们来看一个二元函数求极值问题。假设我们要求一个二元函数 $S(x, y)$ 的极值，则需要求解下面两

个方程

$$\frac{\partial S}{\partial x} = 0, \quad \frac{\partial S}{\partial y} = 0. \quad (2.38)$$

我们当然可以同时求解这两个方程,但是,等价的,也可以先求解 $\frac{\partial S}{\partial y} = 0$, 得到一个关系式 $y = \phi(x)$, 然后将这个关系式代入原来的二元函数 $S(x, y)$ 中, 得到一个一元函数 $S(x, \phi(x))$, 然后再求解这个一元函数的极值,

$$\frac{dS}{dx} = \frac{\partial S}{\partial x} + \frac{\partial S}{\partial y} \frac{\partial \phi}{\partial x} = 0. \quad (2.39)$$

读者很容易证明, 也很容易理解, 这两种不同方法得出来的结果是一样的。

下面我们以单粒子情形为例, 说明以上思路如何可以导出坐标空间的最小作用量原理。我们的出发点是相空间的最小作用量原理, 即求泛函 $S[\mathbf{x}(t), \mathbf{p}(t)]$ 的极值, 我们知道, 这其实就是要求解下面两个方程

$$\frac{\delta S}{\delta \mathbf{x}(t)} = 0, \quad \frac{\delta S}{\delta \mathbf{p}(t)} = 0. \quad (2.40)$$

前面处理相空间最小作用量原理时 [(2.35) 式], 我们是同时给出这两个方程, 设想人们会同时求解它们。现在, 按照上一段的思路, 我们先求解第二个方程, 即

$$\frac{\delta S}{\delta \mathbf{p}(t)} = \dot{\mathbf{x}} - \frac{\partial H}{\partial \mathbf{p}} = 0. \quad (2.41)$$

假设由这个方程解出了一个 $\mathbf{p}(t)$ 关于 $\mathbf{x}(t)$ 的泛函表达式 (注意 $\dot{\mathbf{x}}(t)$ 也是一个 $\mathbf{x}(t)$ 的泛函), 比方说 $\mathbf{p}(t) = \phi[\mathbf{x}(t)]$, 则当将这个泛函关系代入原来的相空间作用量 $S[\mathbf{x}(t), \mathbf{p}(t)]$ 时, 得到的 $S[\mathbf{x}(t), \phi[\mathbf{x}(t)]]$ 就是所谓的坐标空间作用量泛函。剩下只要将这个坐标空间作用量泛函对坐标 $\mathbf{x}(t)$ 变分求极值就可以了! 这个求解坐标空间作用量泛函的极值问题, 就叫做坐标空间的最小作用量原理。值得说明的是, 通常我们将这个坐标空间作用量泛函 $S[\mathbf{x}(t), \phi[\mathbf{x}(t)]]$ 记为 $S[\mathbf{x}(t)]$, 希望读者不要和原来的相空间作用量泛函 $S[\mathbf{x}(t), \mathbf{p}(t)]$ 混淆。

为了将问题看得更清楚一点, 我们注意到, 原来的相空间作用量泛函具有 $S[\mathbf{x}(t), \mathbf{p}(t)] = \int dt L(\mathbf{x}, \mathbf{p}, \dot{\mathbf{x}})$ 的结构, 其中

$$L(\mathbf{x}, \mathbf{p}, \dot{\mathbf{x}}) = \mathbf{p} \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{p}). \quad (2.42)$$

我们注意到, 方程 (2.41) 实际上就是求这个函数 $L(\mathbf{x}, \mathbf{p}, \dot{\mathbf{x}})$ 关于变量 \mathbf{p} 的极值, 不妨定义一个新的函数, 记为 $L(\mathbf{x}, \dot{\mathbf{x}})$ (希望读者不要和原来的 $L(\mathbf{x}, \mathbf{p}, \dot{\mathbf{x}})$ 混淆)

$$L(\mathbf{x}, \dot{\mathbf{x}}) = \text{extrem}_{\mathbf{p}} L(\mathbf{x}, \mathbf{p}, \dot{\mathbf{x}}) = \text{extrem}_{\mathbf{p}} [\mathbf{p} \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{p})]. \quad (2.43)$$

则很明显, 新的坐标空间作用量泛函 $S[\mathbf{x}(t)]$ 其实就是

$$S[\mathbf{x}(t)] = \int_{t_i}^{t_f} L(\mathbf{x}, \dot{\mathbf{x}}) dt. \quad (2.44)$$

式中 $L(\mathbf{x}, \dot{\mathbf{x}})$ 由 (2.43) 式给出, 它有一个专门的名称, 叫做**拉格朗日量**。所谓坐标空间最小作用量原理, 其实就是求泛函 $S[\mathbf{x}(t)]$ 在坐标空间的极值路径, 当然在变分时我们要求

$$\delta \mathbf{x}(t_i) = \delta \mathbf{x}(t_f) = 0. \quad (2.45)$$

从前面关于二元函数极值问题的讨论中容易知道，**这个坐标空间的最小作用量原理和原来的相空间最小作用量原理是等价的！**

(2.43) 式给出的这个从哈密顿量 $H(\mathbf{x}, \mathbf{p})$ 得出拉格朗日量 $L(\mathbf{x}, \dot{\mathbf{x}})$ 的过程称之为勒让德变换，反过来可以证明，哈密顿量 $H(\mathbf{x}, \mathbf{p})$ 也是拉格朗日量 $L(\mathbf{x}, \dot{\mathbf{x}})$ 的勒让德变换，即

$$\text{extrem}_{\dot{\mathbf{x}}} [\mathbf{p} \cdot \dot{\mathbf{x}} - L(\mathbf{x}, \dot{\mathbf{x}})] = H(\mathbf{x}, \mathbf{p}). \quad (2.46)$$

证明过程如下，

$$\begin{aligned} & \text{extrem}_{\dot{\mathbf{x}}} [\mathbf{p} \cdot \dot{\mathbf{x}} - L(\mathbf{x}, \dot{\mathbf{x}})] \\ &= \text{extrem}_{\dot{\mathbf{x}}} [\mathbf{p} \cdot \dot{\mathbf{x}} - \text{extrem}_{\mathbf{p}'} [\mathbf{p}' \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{p}')]] \\ &= \text{extrem}_{\dot{\mathbf{x}}} \text{extrem}_{\mathbf{p}'} [(\mathbf{p} - \mathbf{p}') \cdot \dot{\mathbf{x}} + H(\mathbf{x}, \mathbf{p}')] \\ &= \text{extrem}_{\mathbf{p}', \dot{\mathbf{x}}} [(\mathbf{p} - \mathbf{p}') \cdot \dot{\mathbf{x}} + H(\mathbf{x}, \mathbf{p}')] \\ &= H(\mathbf{x}, \mathbf{p}). \end{aligned} \quad (2.47)$$

所以，拉格朗日量和哈密顿量互为对方的勒让德变换。

比方说，对于我们熟知的最简单的单粒子哈密顿量 $H(\mathbf{x}, \mathbf{p}) = \frac{\mathbf{p}^2}{2m} + V(\mathbf{x})$ ，由 $\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}} = \mathbf{p}/m$ ，我们容易解出 $\mathbf{p} = m\dot{\mathbf{x}}$ ，从而根据勒让德变换 (2.43)，容易得到单粒子的拉格朗日量

$$\begin{aligned} L(\mathbf{x}, \dot{\mathbf{x}}) &= \text{extrem}_{\mathbf{p}} [\mathbf{p} \cdot \dot{\mathbf{x}} - \frac{\mathbf{p}^2}{2m} - V(\mathbf{x})] \\ &= \frac{1}{2} m \dot{\mathbf{x}}^2 - V(\mathbf{x}). \end{aligned} \quad (2.48)$$

我们发现这个拉格朗日量碰巧等于动能减去势能（这个结论也可以推广到多粒子情形）。因此相应的坐标空间作用量泛函就是

$$S[\mathbf{x}(t)] = \int_{t_i}^{t_f} dt \left[\frac{1}{2} m \dot{\mathbf{x}}^2 - V(\mathbf{x}) \right]. \quad (2.49)$$

很显然的是，坐标空间的最小作用量原理完全可以用前面导出来的欧拉-拉格朗日方程来处理，从而有

$$\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{x}}} \right) = 0. \quad (2.50)$$

这就是所谓的拉格朗日力学的欧拉-拉格朗日方程，式中的 L 就是拉格朗日量 $L(\mathbf{x}, \dot{\mathbf{x}})$ ，因此这个方程通常**也称为拉格朗日方程**！比方说，对于单粒子拉格朗日量 $L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2} m \dot{\mathbf{x}}^2 - V(\mathbf{x})$ ，拉格朗日方程给出的是

$$m \frac{d^2 \mathbf{x}}{dt^2} = - \frac{\partial V}{\partial \mathbf{x}}. \quad (2.51)$$

这正是牛顿运动定律！所以我们看到，坐标空间的最小作用量原理能够正确地给出牛顿运动定律。

最后，读者容易明白的是，只要像以前一样适当地引入指标，那上面关于单粒子情形的所有讨论都可以容易地推广到多粒子情形，我们将这个推广留给读者自己去操作。

2.4.2 发现洛伦兹力

既然坐标空间的最小作用量原理和相空间的最小作用量原理等价，我们为什么还要专门讨论它呢？而且某种意义上甚至它的重要性还要超过相空间的最小作用量原理。这是因为，有时候推广拉格朗日量比直接写哈密顿量要简单，下面举一个例子说明。

我们知道，单粒子最简单的拉格朗日量是 $L = \frac{1}{2}m\dot{\mathbf{x}}^2 - V(\mathbf{x})$ ，它关于速度是一个二次型，人们禁不止想问，如果在拉格朗日量中加上一个速度的一次方项会怎么样呢？最简单的尝试是设这一项为 $\int dt \mathbf{A} \cdot \dot{\mathbf{x}}$ ，其中 \mathbf{A} 为一个常矢量，但事实上，这一项不会产生任何结果，原因在于它显然是一个全微分项，积分以后就成了 $\mathbf{A} \cdot (\mathbf{x}(t_f) - \mathbf{x}(t_i))$ ，但是 $\mathbf{x}(t)$ 在两个端点是固定的，所以变分的时候多出来的这一项贡献实际为零，因此不会产生任何影响。

那么，如果假设 \mathbf{A} 依赖于 \mathbf{x} 会怎么样呢？这时候拉格朗日量就是

$$L = \frac{1}{2}m\dot{\mathbf{x}}^2 + \mathbf{A}(\mathbf{x}) \cdot \dot{\mathbf{x}} - V(\mathbf{x}). \quad (2.52)$$

当然，我们可以使用拉格朗日方程来看清中间这一项的影响。但在最小作用量原理的使用中，有时候直接求作用量的变分并令其等于零，这和利用拉格朗日方程是一样方便的，这里我们想用这个例子说明这一点。这时候作用量 $S[\mathbf{x}(t)]$ 为

$$\begin{aligned} S[\mathbf{x}(t)] &= \int_{t_i}^{t_f} dt \left[\frac{1}{2}m\dot{\mathbf{x}}^2 + \mathbf{A}(\mathbf{x}) \cdot \dot{\mathbf{x}} - V(\mathbf{x}) \right] \\ &= \int_{t_i}^{t_f} dt \left[\frac{1}{2}m\dot{\mathbf{x}}^2 - V(\mathbf{x}) \right] + \int A_j(\mathbf{x}) \cdot dx^j. \end{aligned} \quad (2.53)$$

这里我们使用了矢量的分量形式，也使用了求和约定。下面要做的就是对这个作用量进行变分，前两项的变分是通常的，我们直接写出结果（用分量形式），我们将注意力集中在新增的这一项上，从而有（令 $\partial_i = \frac{\partial}{\partial x^i}$ ）

$$\begin{aligned} \delta S &= - \int_{t_i}^{t_f} dt [m\ddot{x}_i + \partial_i V(\mathbf{x})] \delta x^i + \delta \int A_j \cdot dx^j \\ &= - \int_{t_i}^{t_f} dt [m\ddot{x}_i + \partial_i V(\mathbf{x})] \delta x^i + \int (\partial_i A_j \delta x^i dx^j + A_i d\delta x^i) \\ &= - \int_{t_i}^{t_f} dt [m\ddot{x}_i + \partial_i V(\mathbf{x})] \delta x^i + \int d(A_i \delta x^i) + \int (\partial_i A_j \delta x^i dx^j - dA_i \delta x^i) \\ &= - \int_{t_i}^{t_f} dt [m\ddot{x}_i + \partial_i V(\mathbf{x})] \delta x^i + \int (\partial_i A_j - \partial_j A_i) dx^j \delta x^i \\ &= - \int_{t_i}^{t_f} dt [m\ddot{x}_i + \partial_i V(\mathbf{x}) - F_{ij} \dot{x}^j] \delta x^i. \end{aligned} \quad (2.54)$$

式中 $F_{ij} = \partial_i A_j - \partial_j A_i$ ，式中第 4 个等号利用了路径两端固定从而全微分项积出来实际为零。最小作用量原理告诉我们， $\delta S = 0$ ，由此可以得到

$$\begin{aligned} \frac{\delta S}{\delta x^i(t)} &= -[m\ddot{x}_i + \partial_i V(\mathbf{x}) - F_{ij} \dot{x}^j] = 0 \\ \Rightarrow m\ddot{x}_i &= -\partial_i V(\mathbf{x}) + F_{ij} \dot{x}^j. \end{aligned} \quad (2.55)$$

我们看到，在拉格朗日量中多加上这一项，其效果就是，牛顿定律右边多了一个力 $F_{ij} \dot{x}^j$ ，这个力初看起来很陌生，实际上，它正可以代表洛伦兹力！为了看清楚这一点，我

们注意到 $F_{ij} = \partial_i A_j - \partial_j A_i$, 很容易看出, 它实际上就是对矢量场 \mathbf{A} 求旋度的分量式写法, 比如 $\partial_x A_y - \partial_y A_x$ 其实就是 $(\nabla \times \mathbf{A})_z$. 因此, F_{ij} 和磁场强度 \mathbf{B} 的分量形式是一一对应的, 人们可以使用所谓的列维-席维塔符号 ε_{ijk} , 将这两者之间的对应关系写成 $F_{ij} = \varepsilon_{ijk} B^k$. 这里 ε_{ijk} 关于三个指标是全反对称的, 即任意两个指标交换顺序都会出一个负号 (因此当三个指标中有两个取值一样时就会得到 0), 比如说 $\varepsilon_{jik} = -\varepsilon_{ijk}$, 并且 $\varepsilon_{123} = 1$.

利用 $F_{ij} = \varepsilon_{ijk} B^k$, 我们就可以将 (2.55) 式重写成

$$m\ddot{x}_i = -\partial_i V(\mathbf{x}) + \varepsilon_{ijk} \dot{x}^j B^k. \quad (2.56)$$

读者容易验证, 这个式子的矢量形式正是

$$m\ddot{\mathbf{x}} = -\nabla V + \dot{\mathbf{x}} \times \mathbf{B}. \quad (2.57)$$

这和洛伦兹力的表达式已经几乎一样了。为了得到完全正确的洛伦兹力, 只需将拉格朗日量中的 \mathbf{A} 替换成 $q\mathbf{A}$ (q 是粒子的电荷), 并将这个 \mathbf{A} 理解为所谓的矢量势就可以了。矢量势和磁场 \mathbf{B} 的关系是 $\mathbf{B} = \nabla \times \mathbf{A}$ 。

所以, 只要在拉格朗日量中做一点小小的动作, 那么最小作用量原理就能自动导出洛伦兹力, 我们就这样“发现”了洛伦兹力! 描述洛伦兹力的正确拉格朗日量是

$$L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2} m \dot{\mathbf{x}}^2 + q \mathbf{A}(\mathbf{x}) \cdot \dot{\mathbf{x}} - V(\mathbf{x}). \quad (2.58)$$

为了将它勒让德变换得到哈密顿量, 我们需要求

$$\frac{\partial}{\partial \dot{\mathbf{x}}} [\mathbf{p} \cdot \dot{\mathbf{x}} - L(\mathbf{x}, \dot{\mathbf{x}})] = 0 \Leftrightarrow \mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{x}}} \quad (2.59)$$

将拉格朗日量 (2.58) 代进去, 可以得到

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{x}}} = m\dot{\mathbf{x}} + q\mathbf{A}. \quad (2.60)$$

(注意, 对于一般的力学系统, 动量 \mathbf{p} 与速度 $\dot{\mathbf{x}}$ 之间的关系并不天然就是 $\mathbf{p} = m\dot{\mathbf{x}}$, 这里就是一个例子。) 由此可以反解出 $\dot{\mathbf{x}} = (\mathbf{p} - q\mathbf{A})/m$, 进而可以算出

$$\begin{aligned} H(\mathbf{x}, \mathbf{p}) &= \text{extrem}_{\dot{\mathbf{x}}} [\mathbf{p} \cdot \dot{\mathbf{x}} - L(\mathbf{x}, \dot{\mathbf{x}})] \\ &= \frac{(\mathbf{p} - q\mathbf{A})^2}{2m} + V(\mathbf{x}). \end{aligned} \quad (2.61)$$

这就是带电粒子在磁场中的哈密顿量。很明显, 直接猜这个哈密顿量的表达式要比刚才对拉格朗日量进行的小动作难多了! 这个例子清楚地说明了, 有时候从拉格朗日量以及坐标空间的最小作用量原理出发, 比从哈密顿量出发要容易一些!

2.4.3 最小作用量原理能够导出任何方程吗?

一个自然的疑问是, 是否只要合适地选取拉格朗日量, 最小作用量原理就能导出任何方程? 如果是这样的话, 那最小作用量原理本身就不包含任何信息, 那我们还不如直接关心运动微分方程。好在, 答案是否定的, 实际上最小作用量原理能够导出来的方程相当特

殊，而经典物理系统的一个特殊之处正在于，它的运动微分方程可以由最小作用量原理导出。

最小作用量原理导不出什么呢？以一维运动为例，可以证明没有作用量能使下式成立

$$\frac{\delta S}{\delta x(t)} = \dot{x}(t). \quad (2.62)$$

读者不妨取不同的作用量泛函 $S[x(t)]$ 试验一下，就能说服自己这个结论是成立的。证明也很简单，我们对上面这个式子再求一次泛函导数，有

$$\frac{\delta^2 S}{\delta x(t') \delta x(t)} = \frac{\delta \dot{x}(t)}{\delta x(t')} = \frac{d}{dt} \left(\frac{\delta x(t)}{\delta x(t')} \right) = \frac{d}{dt} \delta(t-t'). \quad (2.63)$$

上式最左边不过就是二阶导数，二阶导数可以交换求导顺序，因此它关于 t, t' 对称。但是，上式最右边的 $\frac{d}{dt} \delta(t-t')$ 当交换 t 和 t' 时将成为

$$\frac{d}{dt'} \delta(t'-t) = -\frac{d}{dt} \delta(t'-t) = -\frac{d}{dt} \delta(t-t'). \quad (2.64)$$

即是说，最右边的式子关于 t, t' 是反对称的。这就相互矛盾了，这就说明 (2.62) 不可能成立。

2.4.4 真实路径的作用量是否取极小值？

下面我们来讨论一下，所谓的最小作用量之说是否真有其事，换言之，是否真实路径的作用量是取极小值？

回答是否定的。真实路径的作用量只是作用量的“极值”，不一定是极小值，也不一定是极大值，实际上在所有可能路径所构成的无穷维空间中，真实路径的作用量通常只是作用量泛函的稳定值，或者说驻定值，Saddle point value。

至于为什么是驻值而不是极小值，以及什么是驻值，下面我仅仅只讨论一个例子。在我看来，这例子足够清楚地说明了问题。

考虑一维谐振子，其作用量是

$$S = \int_{t_1}^{t_2} \frac{m}{2} (\dot{x}^2 - \omega^2 x^2) dt. \quad (2.65)$$

将这个作用量在真实路径附近进行二阶变分，根据最小作用量原理，一阶变分等于零，而二阶变分很容易算出来是

$$\delta^2 S = \int_{t_1}^{t_2} \frac{m}{2} [(\delta \dot{x})^2 - \omega^2 (\delta x)^2] dt. \quad (2.66)$$

式中 δx 在路径的两端 (即 t_1, t_2 时刻) 等于零。作用量是否取极小值就看二阶变分是否大于零，如果二阶变分大于零，那作用量就是极小值。

由于 δx 在 t_1, t_2 时刻等于零，所以可以作正弦级数展开：

$$\delta x(t) = \varepsilon \sum_n a_n \sin(\omega_n(t-t_1)), \quad \text{式中 } \omega_n = \frac{n\pi}{t_2-t_1} \quad (2.67)$$

式中 ε 为无穷小量。无穷维向量 $(a_1, a_2, \dots, a_n, \dots)$ 可以看作是路径空间 (即所有可能 $\delta x(t)$ 的空间) 的坐标, 真实路径位于这个无穷维向量空间的原点 $(0, 0, \dots, 0, \dots)$ 。根据 $\delta x(t)$ 的这一式子, 进而可以算得

$$\delta \dot{x}(t) = \varepsilon \sum_n a_n \omega_n \cos(\omega_n(t - t_1)). \quad (2.68)$$

代入 $\delta^2 S$ 的表达式, 利用三角函数的正交性, 不难得到

$$\delta^2 S = \varepsilon^2 \frac{m}{4} (t_2 - t_1) \sum_n a_n^2 (\omega_n^2 - \omega^2). \quad (2.69)$$

此式当 $\omega_1 > \omega$ 从而所有 $\omega_n > \omega$, 即 $\frac{\pi}{t_2 - t_1} > \frac{2\pi}{T}$, 即 $t_2 - t_1 < \frac{T}{2}$ 时为正 (T 为谐振子的周期), 即作用量取极小值。而当此条件不满足时, $\omega_n^2 - \omega^2$ 就会有些为负, 从而使得 $\delta^2 S$ 依赖于展开系数 a_n 的选择而可正可负, 从而作用量就不是取极小值, 而是取鞍点值, Saddle point value。所谓的鞍点, 即是指在所有可能路径的无穷维空间中, 真实路径是这空间中的这么“一个点”, 它沿着某些方向变动, 作用量是极小, 而沿着另一些方向变动, 作用量则是取极大, 犹如一个马鞍的鞍点。

上面的例子清楚地说明了, 对于任何多粒子系统, 当 $t_2 - t_1$ 足够小 (在上例中是 $t_2 - t_1 < \frac{T}{2}$), 从而真实路径足够短时, 作用量的确是取极小值。但是, 对于比较长的真实路径, 作用量通常只是取鞍点值, 也就是驻值。

2.5 广义坐标和广义动量

前面我们看到了, 最小作用量原理 (无论是相空间的还是坐标空间的) 以及与之等价的哈密顿正则方程或者拉格朗日方程, 均可以统一处理单粒子情形和多粒子情形, 形式上唯一的改动无非是指标的取值范围在单粒子情形是 $1, 2, 3$, 而在多粒子情形是 $1, 2, \dots, 3N$ 。但统一处理单粒子和多粒子仅仅只是我们这两章介绍的力学框架的优点之一, 这一框架的另一个重要优点是使得我们不限于直角坐标, 而可以使用任何坐标。这种对坐标的使用自由也使得我们可以方便地处理一大类约束系统。

理想约束与广义坐标

所谓的约束系统, 就是这样的 N 粒子系统, 它的 $3N$ 个坐标之间并不相互独立, 而是要满足一定的约束条件²。比方说, 两维平面内一个约束在圆环 $x^2 + y^2 = R^2$ 上的粒子, 它的 x, y 坐标就不相互独立。由于这些变量不相互独立, 所以当我们利用最小作用量原理进行变分的时候, 各变量的变分也就不相互独立, 从而就无法直接得出通常的哈密顿正则方程或者拉格朗日方程。

因此, 一般来说, 约束系统的处理很复杂, 但是有一大类约束系统, 称之为理想约束系统, 处理起来却很方便。值得说明的是, 当对系统的约束光滑, 从而不存在摩擦力时, 很多约束系统都是理想约束系统。一般地, 理想约束系统就是这样的约束系统, 它的约束力对系统所有可能的满足约束条件的运动 (不管它是否真实发生) 均不做功! 从而只要不

²这实际上只包括了所谓的完整约束 (holonomic constraint) 系统, 实际的约束系统比这还要复杂。

破坏约束条件，那约束力对这一系统的能量就没有贡献，也就是对哈密顿量没有贡献，从而我们就可以根本不去管约束力。这时候就可以使用这两章所发展起来的力学处理框架。

不仅如此，对于理想约束系统，我们还可以不使用直角坐标，而是使用一些描述系统的独立变量为坐标，比方说，对于约束在圆环 $x^2 + y^2 = R^2$ 上的粒子，我们可以不使用直角坐标，而是使用角度 θ 作为独立坐标。描述理想约束系统的这些独立变量就叫做**广义坐标**！独立变量的个数就叫做系统的**自由度**数。使用广义坐标的好处就在于，它们是自动满足约束条件的独立变量，这样就无需在问题的求解中额外再把约束条件强加进来。如此一来，在处理最小作用量原理时，对这些独立变量的变分就是相互独立的，这样就能够根据最小作用量原理写出哈密顿正则方程或者拉格朗日方程了。

与广义坐标对应的概念叫广义动量，它的定义要复杂一些，下面我们首先从相空间出发来给出定义，然后再讨论如何从位形空间以及拉格朗日量出发定义它。

直接从相空间出发

为了定义广义动量，我们首先注意到一个 N 粒子系统的相空间作用量为

$$\begin{aligned} S &= \int dt [p_\mu \dot{x}^\mu - H(x, p)] \\ &= \int p_\mu dx^\mu - \int H dt. \end{aligned} \quad (2.70)$$

注意第二行这个表达式，它的第二项很简单，就是哈密顿量对时间的积分，我们主要关心前一项，它是一个 1-形式 $p_\mu dx^\mu$ 沿着相空间路径的积分。这个 1-形式对于决定哈密顿正则方程无疑是至关重要的，因此它有一个专门的名称，叫做辛势，记作 Θ ，

$$\Theta = p_\mu dx^\mu. \quad (2.71)$$

以上是用通常的坐标表达问题，因此 $\mu = 1, 2, \dots, 3N$ 。但是通常坐标的坏处是，它们可能不相互独立。为此，我们转用广义坐标，假设这个系统有 s 个独立的广义坐标（自由度数是 s ），记作 $q^a, a = 1, 2, \dots, s$ 。则，为了保证哈密顿正则方程在广义坐标中看起来形式也一样（注意，当存在约束时，由于通常坐标的各变量不相互独立，所以其实并没有对于它们的哈密顿正则方程），我们**要求辛势 Θ 在广义坐标中也有和通常坐标完全类似的形式**，具体来说，我们要求

$$\Theta = p_\mu dx^\mu = p_a dq^a. \quad (2.72)$$

上面表达式中的 p_a （虽然我们用的是与 p_μ 几乎相同的符号，但请读者不要把这两者搞混）就是**广义动量**。即是说，广义动量可以通过通常的动量 p_μ 经由下面的坐标变换得来

$$p_a = p_\mu \frac{\partial x^\mu}{\partial q^a}. \quad (2.73)$$

很明显，利用广义坐标和广义动量，我们可以将相空间的作用量写成

$$S = \int dt [p_a \dot{q}^a - H(q, p)]. \quad (2.74)$$

利用相空间的最小作用量原理，对广义坐标 q 和广义动量 p 进行变分，就能得到最一般的哈密顿正则方程

$$\frac{dq^a}{dt} = \frac{\partial H}{\partial p_a}, \quad \frac{dp_a}{dt} = -\frac{\partial H}{\partial q^a}. \quad (2.75)$$

显然，它和通常坐标中的哈密顿正则方程有完全类似的形式，区别在于，如果存在约束的话，那这时候实际上没有通常坐标的哈密顿正则方程。

如果没有约束，那广义坐标和广义动量不过是相空间的一组不同坐标系。但是如果有约束，那通常的坐标和通常的动量由于不相互独立，它们就不够成系统的相空间，相反，这时候系统的相空间是指广义坐标和广义动量的空间。显然，系统在相空间中依然按照哈密顿正则方程进行演化。

不妨举一个例子，还是那个约束在圆环 $x^2 + y^2 = R^2$ 上的粒子。取广义坐标为角度 θ ，它和直接坐标的关系是

$$x = R\cos(\theta), \quad y = R\sin(\theta). \quad (2.76)$$

将辛势 Θ 分别在两种坐标中表达，即有

$$\begin{aligned} \Theta &= p_x dx + p_y dy = p_\theta d\theta \\ \Rightarrow p_\theta &= [-p_x \sin(\theta) + p_y \cos(\theta)]R \end{aligned} \quad (2.77)$$

再利用动量沿着圆周的法向分量为零，即

$$p_x \cos(\theta) + p_y \sin(\theta) = 0. \quad (2.78)$$

即可以推出

$$\Rightarrow \mathbf{p}^2 = p_\theta^2 / R^2.$$

读者容易验证 p_θ 其实刚好是粒子环绕圆环的角动量（所以广义动量不一定是通常的动量，它甚至可以是角动量）。我们也容易将粒子的哈密顿量用广义坐标和广义动量表达出来，

$$H = \frac{\mathbf{p}^2}{2m} + V = \frac{p_\theta^2}{2mR^2} + V(\theta). \quad (2.79)$$

进而可以写出这个例子的哈密顿正则方程

$$\begin{aligned} \dot{\theta} &= \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{mR^2} \\ \dot{p}_\theta &= -\frac{\partial H}{\partial \theta} = -\frac{\partial V}{\partial \theta}. \end{aligned} \quad (2.80)$$

θ 和 p_θ 的空间就是这个例子的相空间。 θ 的取值范围是一个圆周，构成这个例子的广义坐标空间，记为 S^1 ， p_θ 的取值范围是 $-\infty$ 到 $+\infty$ ，也就是取遍实数轴，记为 \mathbb{R}^1 。通常我们把广义坐标空间 S^1 画在平面上，再把广义动量空间 \mathbb{R}^1 画在垂直于这个平面的轴上。很显然，如此一来，上面这个例子的相空间就是以 S^1 为底，以 \mathbb{R}^1 为高的一个无穷长圆柱面，记为 $S^1 \times \mathbb{R}^1$ 。数学家常常称这个相空间为 S^1 的余切丛，并记为 T^*S^1 ，当然 $T^*S^1 = S^1 \times \mathbb{R}^1$ ，不过，我们不用管数学家们引入的这些奇怪名词。值得注意的是，这个例子告诉我们，一个哈密顿系统的相空间不一定是通常的欧几里德空间。

从位形空间出发

如果我们是从拉格朗日量以及坐标空间最小作用量原理出发处理问题，那情况就更简单，这时候只要先写出直角坐标中的拉格朗日量（通常是动能减去势能），然后再坐标变换成广义坐标就可以了。用广义坐标来表达的拉格朗日量通常记作 $L(q, \dot{q})$ ，这里 q 这样的符号只是一个抽象的记号，实际上它代表所有作为独立变量的广义坐标 q^a 。广义坐标的空间也叫做系统的位形空间，而函数组 $\{q^a(t), a = 1, 2, \dots, s\}$ 就代表位形空间的一条路径。相应的坐标空间的最小作用量原理就成了位形空间的最小作用量原理，作用量 $S[q(t)]$ 为

$$S[q(t)] = \int dt L(q, \dot{q}). \quad (2.81)$$

对位形空间的路径 $q(t)$ 进行变分，就能够得到熟知的拉格朗日方程

$$\frac{\partial L}{\partial q^a} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^a} \right) = 0. \quad (2.82)$$

同样，我们可以对拉格朗日量 $L(q, \dot{q})$ 进行勒让德变换，进而得到哈密顿量 $H(q, p)$ 。具体来说，

$$H(q, p) = \text{extrem}_{\{\dot{q}^a\}} [p_b \dot{q}^b - L(q, \dot{q})]. \quad (2.83)$$

显然，为了进行这个勒让德变换，需要求解

$$\frac{\partial}{\partial \dot{q}^a} [p_b \dot{q}^b - L(q, \dot{q})] = 0 \Rightarrow p_a = \frac{\partial L}{\partial \dot{q}^a}. \quad (2.84)$$

我们正是把后面这个式子当作从拉格朗日量出发对广义动量的定义。

下面以双摆为例来说明广义坐标表达下的拉格朗日方程如何方便我们求解力学问题的。所谓的双摆，就是两个质量分别为 m_1, m_2 的质点，用长度分别为 l_1, l_2 的轻杆连起来所组成的系统，如图 (2.3) 所示。假设连接处光滑且可自由活动，那这就是一个理想约束系统。这个系统如果直接用受力分析来处理，那还是比较复杂的。但用广义坐标和拉格朗

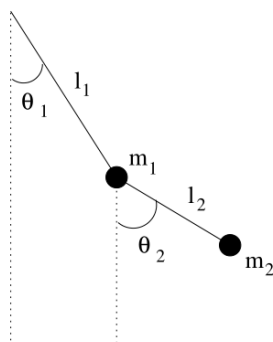


Figure 2.3: 双摆

日方程处理起来就很方便。

为此，我们引入图 (2.3) 中的两个广义坐标 θ_1, θ_2 。很显然，第一个质点的动能 T_1 和势能 V_1 (重力势能) 分别为

$$T_1 = \frac{1}{2} m_1 l_1^2 \dot{\theta}_1^2, \quad V_1 = -m_1 g l_1 \cos(\theta_1). \quad (2.85)$$

为了求出第二个质点的动能，我们假设系统在 $x-y$ 平面内摆动， x 轴水平向右， y 轴竖直向下。则第二个质点的坐标 (x_2, y_2) 可以表示为

$$x_2 = l_1 \sin(\theta_1) + l_2 \sin(\theta_2), \quad y_2 = l_1 \cos(\theta_1) + l_2 \cos(\theta_2). \quad (2.86)$$

则容易求出质点 2 的动能 T_2 为

$$\begin{aligned} T_2 &= \frac{1}{2} m_2 (\dot{x}_2^2 + \dot{y}_2^2) \\ &= \frac{1}{2} m_2 (l_1^2 \dot{\theta}_1^2 + l_2^2 \dot{\theta}_2^2 + 2l_1 l_2 \cos(\theta_1 - \theta_2) \dot{\theta}_1 \dot{\theta}_2). \end{aligned} \quad (2.87)$$

质点 2 的重力势能 V_2 为

$$V_2 = -m_2 g y_2 = -m_2 g (l_1 \cos(\theta_1) + l_2 \cos(\theta_2)). \quad (2.88)$$

由此可以写出整个系统的拉格朗日量 (等于总动能减去总势能)

$$\begin{aligned} L &= \frac{1}{2} (m_1 + m_2) l_1^2 \dot{\theta}_1^2 + \frac{1}{2} m_2 l_2^2 \dot{\theta}_2^2 + m_2 l_1 l_2 \cos(\theta_1 - \theta_2) \dot{\theta}_1 \dot{\theta}_2 \\ &\quad + (m_1 + m_2) g l_1 \cos(\theta_1) + m_2 g l_2 \cos(\theta_2). \end{aligned} \quad (2.89)$$

代入下面的拉格朗日方程，就可以得到系统的运动微分方程

$$\frac{\partial L}{\partial \theta_1} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}_1} \right) = 0, \quad \frac{\partial L}{\partial \theta_2} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}_2} \right) = 0. \quad (2.90)$$

不过，这是一个相当复杂的系统，实际上，当能量大到一定程度时，这个系统的运动是混沌的！

3. 对称性

设想你在一艘太空飞船里，离所有的星星都很远，船窗密闭，你看不到外面，也无法接收外界的信息，并且假设推进器已经不向外喷射物质了，总之，假设你和飞船成了一个真正的封闭系统。不过，只要局限于飞船之内，你可以做任何事情。请问，你是否能判断飞船在太空中的坐标，飞船的朝向，以及飞船是否在飞行？

答案当然是不能，你的这种不能就是物理定律的对称性。简言之，物理定律的对称性，就是对某些物理性质的无法辨测性。伽利略最早领悟了对称性在物理学中的重要性，伽利略的这种领悟后来被爱因斯坦发扬光大，最终成为现代物理学的基石。那么，为什么无法辨测呢？你为什么无法辨测飞船的坐标和方向呢？原因在于，即使把飞船平移一段距离并旋转一个角度，飞船之内的你也无法觉察到任何的不同。即是说，这种无法辨测性起因于你可以对系统进行一些操作，但操作前后你归纳出来的物理定律完全一样，有完全一样的数学形式。所以，物理定律的对称性其实就是物理定律的形式在操作前后的不变性。

关于物理定律的对称性的这一定义与几何对称性是一致的。比如，当我们说圆有很大的对称性时，我们指的是，圆的几何图形在绕圆心旋转任意角的操作下保持不变。当我们说人体外观左右对称时（并不严格），我们指的是将人体沿着中线左右对调，人体外观保持不变。

3.1 对称性与守恒定律

3.1.1 对称性就是作用量的不变性

前面的章节中我们讲了经典物理学的基本原理，最小作用量原理，上面又定义了何为物理定律的对称性，下面我们将这两者结合起来。结合对称性与最小作用量原理的关键在

于, 根据最小作用量原理, 基本力学定律由作用量的变分等于零给出。当然, 这个作用量可以是相空间作用量, 也可以是坐标空间作用量, 不过, 相对来说用坐标空间的作用量来分析对称性更加方便, 因此本章将集中于此。而在相空间中分析对称性需要的准备知识就稍微多一些, 我们将推迟到后面的章节中进行。

所谓的某系统有一个对称性, 我们是指 (为了公式简洁起见, 我们省略广义坐标上面的分量指标): 首先, 系统按照某条物理路径 $q(t)$ 在位形空间中演化, 这条物理路径满足基本方程 $\delta S[q(t)] = 0$ 。其次, 假设在对称操作的作用之下, 原来的 $q(t)$ 演化路径变换成了新的路径 $\tilde{q}(t) = F(q(t))$ 。所谓的对称操作, 我们是指这条新的路径满足同样形式的基本方程 $\delta S[\tilde{q}(t)] = 0$ 。即, 对于对称操作作用前后的两条路径 $q(t)$ 和 $\tilde{q}(t) = F(q(t))$, 有

$$\delta S[q(t)] = 0 \Leftrightarrow \delta S[\tilde{q}(t)] = 0. \quad (3.1)$$

很显然, 满足这一要求的最直接方式是

$$S[\tilde{q}(t)] = S[q(t)]. \quad (3.2)$$

即是说, 在对称操作的作用下, 作用量泛函保持不变, **对称性就是作用量的不变性**。

3.1.2 诺特定理

对称性可以分成两类, 像镜像对称这样的只包含有限个不同操作的对称性, 称为离散对称性。像空间旋转这样的对称性称为连续对称性, 因为它有无穷多不同的对称操作 (这里即是旋转操作), 这些对称操作依赖于某些连续的参数 (在这里就是旋转角度)。本章主要考察连续对称性。原因之一在于, 连续对称性出人意料地和所谓的守恒定律密切相关, 可以证明, 每一种连续对称性都对应一条守恒定律, 这就是著名的诺特定理。

诺特定理的证明并不复杂, 下面我们给出一个绝妙的证明。我不知道这个证明应该归功于谁, 我自己最早是从温伯格的量子场论书中学的, 但我也在好几个大师的书或者课程中或多或少见过这个证明的关键想法。

值得说明的是, 下面的证明也有一定的局限性, 即它假设了对称操作不改变时间 t 。不过, 即使将一般情形包括进来证明的关键也还是类似的, 只是如果想一般性地写出来那就要复杂一些。为了让读者看清楚证明的关键, 这里先处理简化情形, 然后, 我们将在下一节中用对时间平移对称性的讨论说清楚如何推广到更一般情形。

假设系统有某种连续对称性 (即是说, 系统的作用量在这样的连续操作之下保持不变), 记为 $G(\theta)$, θ 就是这种对称性所依赖的连续参数, $\theta = 0$ 表示对称操作还来不及加上, 即表示不进行任何操作, 也称恒等操作。下面考虑参数 $\theta = \varepsilon$ 的无穷小对称操作, ε 为无穷小量。假设在此对称操作之下, 时间 t 保持不变, 但 q 变为 \tilde{q}

$$q(t) \rightarrow \tilde{q}(t) = q(t) + \varepsilon F(q(t)), \quad (3.3)$$

式中 $F(q(t))$ 为某个表达式, 它的形式取决于对称操作的具体定义。则根据对称性的定义, 作用量泛函将在此对称操作下保持不变, 即

$$\delta S = S[\tilde{q}(t)] - S[q(t)] = \int dt L(\tilde{q}, \dot{\tilde{q}}) - \int dt L(q, \dot{q}) = 0. \quad (3.4)$$

下面是关键的证明技巧。前面的 ε 是一个无穷小常数，现在，设想我们将这个无穷小常数变成一个关于 t 的任意函数 $\varepsilon(t)$ (函数值为无穷小量)，并要求在路径的起末两端 (分别对应时间 t_i 和 t_f)，这个函数取值为零，即

$$\varepsilon(t_i) = \varepsilon(t_f) = 0. \quad (3.5)$$

我们考虑修改的变换

$$\begin{aligned} q(t) &\rightarrow \tilde{q}(t) = q(t) + \varepsilon(t)F(q(t)), \\ \text{即 } \delta q &= \tilde{q}(t) - q(t) = \varepsilon(t)F(q(t)). \end{aligned} \quad (3.6)$$

当然，现在因为 $\varepsilon(t)$ 不再是对称变换参数 (对称变换参数不依赖于 t)，所以这个变换并不是一个对称性，因此不能保持作用量不变，但是如果 $\varepsilon(t)$ 变回常数函数，那作用量将是不变的。这就意味着，在 (3.6) 的变换下，作用量的改变量必然具有如下形式

$$\delta S = \int_{t_i}^{t_f} dt (L(\tilde{q}, \dot{\tilde{q}}) - L(q, \dot{q})) = \int_{t_i}^{t_f} dt Q(q, \dot{q}) \dot{\varepsilon}(t). \quad (3.7)$$

式中 $Q(q, \dot{q})$ 为某个表达式。我们简单解释一下为什么是这样，原因有三：第一，作用量是某个函数 (拉格朗日量) 对时间的积分，所以作用量的改变量必然具有时间积分的形式。第二，拉格朗日量中只含有 $q(t)$ 的一阶导数，因此作用量的改变量中最多只含有 $\varepsilon(t)$ 的一阶导数。第三，将作用量的改变量计算到一阶无穷小项，那这个改变量必然不能含 $\varepsilon(t)$ ，因为这样的项在 $\varepsilon(t)$ 变回常数函数时不能变成零。所以这个改变量的一阶小量只能含 $\dot{\varepsilon}$ (因此当 $\varepsilon(t)$ 变回常数函数时自动得到零)。综上，我们必有表达式 (3.7)。

以上都是数学技巧，下面是整个证明的物理部分。上面讨论中的路径 $q(t)$ 是任意路径，不必是真实演化路径。现在，假设我们取 $q(t)$ 为真实的物理演化路径，那最小作用量原理将告诉我们，对于**任何**两端固定的无穷小改变 (变分)，将有 $\delta S = 0$ 。(3.6) 正是这样的一个两端固定的无穷小变分，因此这就意味着，**对于真实的物理路径** $q(t)$ ，(3.7) 式得出的作用量改变量必定恒为零，即

$$\delta S = \int_{t_i}^{t_f} dt Q(q, \dot{q}) \dot{\varepsilon}(t) = 0. \quad (3.8)$$

将上式分部积分，就可以得到

$$\delta S = - \int_{t_i}^{t_f} dt \dot{Q}(q, \dot{q}) \varepsilon(t) = 0. \quad (3.9)$$

但是， $\varepsilon(t)$ 的函数形式是任意的，因此这就说明，对于物理路径 $q(t)$ ，我们有

$$\frac{dQ}{dt} = 0. \quad (3.10)$$

即 $Q(q, \dot{q})$ 为一个守恒量！可见，相应于每一个连续的对对称性，都必定存在一个守恒量，这，就是诺特定理！

值得强调的是，以上诺特定理的证明过程也给我们指出了一条如何找到守恒量 $Q(q, \dot{q})$ 的道路，我们下面的物理讨论都基于这一道路。

3.1.3 动量守恒、角动量守恒、能量守恒

动量守恒、角动量守恒、还有能量守恒，这是最为人们熟知的守恒定律，也是最为基本的物理学定律之一。这些定律为什么如此基本呢？按照诺特定理的精神，原因应该在于导致这些守恒定律的对称性很基本，那么它们分别是什么对称性呢？结果表明，动量守恒是所谓的空间平移不变性的结果，角动量守恒是空间旋转不变性的结果，它们也就是本章一开始所谈的宇宙飞船位置坐标和空间方向的无法辨测性。而能量守恒则是时间平移不变性的结果，所谓的时间平移不变性也就是时间的零点是人们任意规定的，无法通过物理定律辨测时间的零点，因为明天做实验将和今天做实验得到同样的规律。

空间平移对称性与动量守恒

我们先来讨论空间平移对称性和动量守恒。考虑一个多粒子系统，记其拉格朗日量为 $L(\mathbf{x}_i, \dot{\mathbf{x}}_i)$ ($L(\mathbf{x}_1, \dots, \mathbf{x}_N, \dot{\mathbf{x}}_1, \dots, \dot{\mathbf{x}}_N)$ 的简记)，因此系统的作用量为 $S = \int dt L(\mathbf{x}_i, \dot{\mathbf{x}}_i)$ 。假设系统具有空间平移不变性，即在如下坐标平移下，作用量保持不变

$$\mathbf{x}_i \rightarrow \tilde{\mathbf{x}}_i = \mathbf{x}_i + \mathbf{a}. \quad (3.11)$$

很显然，这里作用量保持不变其实就相当于拉格朗日量在坐标平移下保持不变，即有 $L(\mathbf{x}_i + \mathbf{a}, \dot{\mathbf{x}}_i) = L(\mathbf{x}_i, \dot{\mathbf{x}}_i)$ 。考察无穷小平移，即取平移量 $\mathbf{a} = \boldsymbol{\varepsilon}$ 为无穷小量，从而有

$$0 = \delta L = L(\mathbf{x}_i + \boldsymbol{\varepsilon}, \dot{\mathbf{x}}_i) - L(\mathbf{x}_i, \dot{\mathbf{x}}_i) = \boldsymbol{\varepsilon} \cdot \sum_j \frac{\partial L}{\partial \mathbf{x}_j}. \quad (3.12)$$

下面使用关键技巧，即将无穷小参数 $\boldsymbol{\varepsilon}$ 变成 $\boldsymbol{\varepsilon}(t)$ ，即考察

$$\mathbf{x}_i \rightarrow \tilde{\mathbf{x}}_i = \mathbf{x}_i + \boldsymbol{\varepsilon}(t). \Leftrightarrow \delta \mathbf{x}_i = \boldsymbol{\varepsilon}(t). \quad (3.13)$$

从而容易计算出作用量的改变量，有

$$\begin{aligned} \delta S &= \int dt \delta L = \int dt \left[\sum_j \frac{\partial L}{\partial \mathbf{x}_j} \cdot \delta \mathbf{x}_j + \sum_j \frac{\partial L}{\partial \dot{\mathbf{x}}_j} \delta \dot{\mathbf{x}}_j \right] \\ &= \int dt \left[\sum_j \frac{\partial L}{\partial \mathbf{x}_j} \cdot \boldsymbol{\varepsilon} + \sum_j \frac{\partial L}{\partial \dot{\mathbf{x}}_j} \cdot \dot{\boldsymbol{\varepsilon}} \right] \end{aligned} \quad (3.14)$$

利用平移不变性的 (3.12) 式，即有

$$\delta S = \int dt \left(\sum_j \frac{\partial L}{\partial \dot{\mathbf{x}}_j} \right) \cdot \dot{\boldsymbol{\varepsilon}}. \quad (3.15)$$

根据上一节关于诺特定理的证明容易知道，上式说明 $(\sum_j \frac{\partial L}{\partial \dot{\mathbf{x}}_j})$ 是守恒量。根据上一章关于勒让德变换的讨论， $\frac{\partial L}{\partial \dot{\mathbf{x}}_j}$ 就是第 j 个粒子的动量 \mathbf{p}_j ，即 $\mathbf{p}_j = \frac{\partial L}{\partial \dot{\mathbf{x}}_j}$ 。所以，守恒的正是系统的总动量 \mathbf{p}

$$\mathbf{p} = \sum_j \mathbf{p}_j. \quad (3.16)$$

所以空间平移不变性会导致系统的总动量守恒！

空间旋转不变性与角动量守恒

下面讨论旋转不变性和角动量守恒。这一次我们考察单粒子情形 (推广到多粒子情形是直接了当的), 不过, 我们假设这个粒子在 d 维空间中运动 (即不限于通常的 3 维空间)。这里依然采用直角坐标, 从而粒子的位置矢量 $\mathbf{x} = (x^1, x^2, \dots, x^d)$, 我们将各分量记为 $x^\alpha, \alpha = 1, 2, \dots, d$ 。这里采用了上指标来标识各直角坐标分量, 但其实对于欧氏空间的直角坐标, 用上指标还是下指标并没有区别, 即有 $x_\alpha = x^\alpha$ 。

首先需要研究一下如何处理 d 维空间中的空间旋转, 核心是如何处理无穷小旋转 (因为诺特定理的关键在于无穷小旋转)。记无穷小旋转操作之下, 坐标的改变量为 δx^α , 很显然, δx^α 应该与原来的矢量 $\{x^\alpha\}$ 成线性关系 (因为一个矢量在无穷小旋转之下改变多少显然是正比于原来的矢量的), 我们设

$$\delta x_\alpha = \varepsilon_{\alpha\beta} x^\beta. \quad (3.17)$$

这里使用了求和约定, 式中的 $\varepsilon_{\alpha\beta}$ 为刻画旋转的无穷小量。很明显, 空间旋转不会改变一个矢量的长度, 从而 $\delta(\mathbf{x}^2) = 0$, 即有

$$0 = \delta(\mathbf{x}^2)/2 = \mathbf{x} \cdot \delta \mathbf{x} = x^\alpha \delta x_\alpha = \varepsilon_{\alpha\beta} x^\alpha x^\beta. \quad (3.18)$$

注意到 $x^\alpha x^\beta$ 关于指标 α, β 对称, 所以满足上式的唯一可能性是, $\varepsilon_{\alpha\beta}$ 关于两个指标反对称¹, 即

$$\varepsilon_{\alpha\beta} = -\varepsilon_{\beta\alpha}. \quad (3.19)$$

所以, d 维空间中独立的无穷小旋转有 $d(d-1)/2$ 个, 对于 $d=3$ 的三维空间, 独立的无穷小旋转数目刚好是 3!

因此, 在无穷小旋转操作之下, 我们有

$$\delta x_\alpha = \varepsilon_{\alpha\beta} x^\beta, \quad \delta \dot{x}_\alpha = \varepsilon_{\alpha\beta} \dot{x}^\beta. \quad (3.20)$$

如果系统具有旋转不变性, 从而在上面的无穷小旋转之下拉格朗日量保持不变, 即有 $\delta L(\mathbf{x}, \dot{\mathbf{x}}) = 0$ 。

下面将 $\varepsilon_{\alpha\beta}$ 变成 $\varepsilon_{\alpha\beta}(t)$, 那现在 $\delta \dot{x}_\alpha$ 就会多出一项, 变成 (注意我们说过, 上下指标其实没有区别)

$$\delta \dot{x}^\alpha = \varepsilon^{\alpha\beta} \dot{x}_\beta + \dot{\varepsilon}^{\alpha\beta} x_\beta. \quad (3.21)$$

从而现在拉格朗日量就不再是不变了, 其改变量会多出一个非零项, 变成

$$\begin{aligned} \delta L &= \frac{\partial L}{\partial \dot{x}^\beta} \dot{\varepsilon}^{\beta\alpha} x_\alpha = p_\beta x_\alpha \dot{\varepsilon}^{\beta\alpha} \\ &= -\frac{1}{2} (x_\alpha p_\beta - x_\beta p_\alpha) \dot{\varepsilon}^{\alpha\beta}. \end{aligned} \quad (3.22)$$

¹ 因为反对称和对称乘在一起才等于零, 证明如下: 设 $A_{\alpha\beta}$ 关于两指标反对称, $B^{\alpha\beta}$ 关于两指标对称, 即 $A_{\alpha\beta} = -A_{\beta\alpha}$, $B^{\alpha\beta} = B^{\beta\alpha}$ 。则有 $A_{\alpha\beta} B^{\alpha\beta} = -A_{\beta\alpha} B^{\alpha\beta} = -A_{\beta\alpha} B^{\beta\alpha} = -A_{\alpha\beta} B^{\alpha\beta}$, 从而 $A_{\alpha\beta} B^{\alpha\beta} = 0$ 。以上推导的最后一步是重命名了一下指标。

式中已经利用了动量的定义式 $p_\beta = \frac{\partial L}{\partial \dot{x}^\beta}$ 。当然，现在作用量也不是不变的，而是会改变为

$$\delta S = -\frac{1}{2} \int dt (x_\alpha p_\beta - x_\beta p_\alpha) \dot{\epsilon}^{\alpha\beta} \quad (3.23)$$

根据上一节诺特定理证明的有关讨论可以知道，上式意味着 $x_\alpha p_\beta - x_\beta p_\alpha$ 是一个守恒量，习惯上记为 $J_{\alpha\beta}$ ，

$$J_{\alpha\beta} = x_\alpha p_\beta - x_\beta p_\alpha. \quad (3.24)$$

从而 d 维空间的旋转不变性意味着存在 $d(d-1)/2$ 个守恒量，为 $J_{\alpha\beta}$ ，这些守恒量就是所谓的角动量。如果推广到多粒子情形，那守恒的就是总角动量。为了看清楚 $J_{\alpha\beta}$ 的确是角动量，我们取 $d=3$ ，这时候有 3 个守恒量，分别为

$$J_{12} = xp_y - yp_x, \quad J_{23} = yp_z - zp_y, \quad J_{31} = zp_x - xp_z. \quad (3.25)$$

显然这正是通常的角动量，通常分别记作 J^3, J^1, J^2 。利用上一章引入的列维-西维塔记号 ϵ^{ijk} (请不要和上文的各无穷小量混淆)，我们可以将 3 维空间角动量这两种写法之间的关系写为

$$J^i = \frac{1}{2} \epsilon^{ijk} J_{jk}. \quad (3.26)$$

不过，很显然，高维空间的角动量只能记为 $J_{\alpha\beta}$ 。

时间平移对称性与能量守恒

我们知道，封闭系统的能量守恒，这条守恒定律源自于哪种对称性呢？结果表明能量守恒源自于封闭系统所具有的时间平移不变性，即你是今天观测这个系统还是明天观测这个系统得到的物理规律是一样的。从作用量和拉格朗日量来看的话，时间平移不变性来源于拉格朗日量不显含时间 t ，从而在 $t \rightarrow \tilde{t} = t + a$ 的时间平移下（相当于将整个系统往过去方向移动了 a ，所以其时间坐标增加了 a ），我们有 $q(t) \rightarrow \tilde{q}(t) = q(\tilde{t})$ ，从而

$$\begin{aligned} S[q(t)] &= \int_{t_i}^{t_f} dt L(q(t), \dot{q}(t)) \rightarrow \\ S[q(\tilde{t})] &= \int_{t_i-a}^{t_f-a} dt L(q(\tilde{t}), \frac{d}{dt} q(\tilde{t})) = \int_{t_i}^{t_f} d\tilde{t} L(q(\tilde{t}), \dot{q}(\tilde{t})) \\ &= \int_{t_i}^{t_f} dt L(q(t), \dot{q}(t)) = S[q(t)], \end{aligned}$$

式中最后一个等号只是将 \tilde{t} 重命名为了 t 。可见，正因为 $L(q, \dot{q})$ 不显含 t ，所以作用量在时间平移下保持不变。对于开放系统，通常拉格朗日量是要显含 t 的，即要写成 $L(q, \dot{q}, t)$ ，这时候当然就没有时间平移不变性了。

下面考虑无穷小时间平移，并且和前面一样，将无穷小参数变为 $\epsilon(t)$ ，即考虑如下时间变换

$$t \rightarrow \tilde{t}(t) = t + \epsilon(t), \quad (3.27)$$

注意 \tilde{t} 为 t 的函数。现在，变换以后的作用量 $S[q(\tilde{t})]$ 为

$$\begin{aligned}
 S[q(\tilde{t})] &= \int dt L(q(\tilde{t}), \frac{d}{dt}q(\tilde{t})) = \int d\tilde{t} \left(\frac{dt}{d\tilde{t}} \right) L\left(q(\tilde{t}), \left(\frac{d\tilde{t}}{dt} \right) \dot{q}(\tilde{t})\right) \\
 &= \int d\tilde{t} \left(\frac{1}{1+\dot{\epsilon}} \right) L\left(q(\tilde{t}), (1+\dot{\epsilon})\dot{q}(\tilde{t})\right) \\
 &= \int d\tilde{t} L(q(\tilde{t}), \dot{q}(\tilde{t})) - \int d\tilde{t} (\dot{\epsilon}) L(q(\tilde{t}), \dot{q}(\tilde{t})) + \int d\tilde{t} \frac{\partial L}{\partial (\dot{q}(\tilde{t}))} \dot{q}(\tilde{t}) (\dot{\epsilon}) \\
 &= \int dt L(q(t), \dot{q}(t)) + \int dt \left[\frac{\partial L}{\partial (\dot{q}(t))} \dot{q}(t) - L(q(t), \dot{q}(t)) \right] \dot{\epsilon}, \tag{3.28}
 \end{aligned}$$

式中最后一行是将变量 \tilde{t} 重命名成了 t 。从上面计算可以得到，变换前后作用量的改变量为

$$\delta S = S[q(\tilde{t})] - S[q(t)] = \int dt \left[\frac{\partial L}{\partial \dot{q}} \dot{q} - L(q, \dot{q}) \right] \dot{\epsilon}. \tag{3.29}$$

完全类似于上一节证明诺特定理时的推理可以知道，上式意味着存在下面的守恒量

$$E(q, \dot{q}) = \frac{\partial L}{\partial \dot{q}} \dot{q} - L(q, \dot{q}). \tag{3.30}$$

上面的推导过程忽略了广义坐标的指标，如果恢复指标，那守恒量 $E(q, \dot{q})$ 的表达式就将是

$$E(q, \dot{q}) = \frac{\partial L}{\partial \dot{q}^a} \dot{q}^a - L(q, \dot{q}) \tag{3.31}$$

很明显，如果我们利用 $p_a = \frac{\partial L}{\partial \dot{q}^a}$ 将这个守恒量中的变量 \dot{q}^a 用广义动量 p_a 来表示，那这个守恒量就正是哈密顿量 $H(q, p)$ 。这说明，与时间平移对称性对应的守恒量 $E(q, \dot{q})$ 在物理上就代表能量。所以，时间平移对称性意味着能量守恒！

上面这个将时间平移对称性和能量守恒定律联系起来的推导过程正好完善了我们上一小节对诺特定理的证明。因为这里涉及到的对称性正好是需要将时间 t 变换为 \tilde{t} 的那种对称性，而上面的推导正好可以说明在这种情形下如何处理关于诺特定理的证明。

以上关于能量守恒定律的证明看起来有些不必要的复杂。但是，这一证明的好处是，它最具有诺特定理的精神，这同样也是我们关于动量守恒以及角动量守恒的讨论过程的优点，正因为如此，以上关于这些守恒定律的讨论过程完全可以平行地照搬到角动量中。而且，由于现在的物理系统比场要简单一些，所以这里的讨论可以写得更加清晰，从而更能凸显诺特定理相关推理和推导的本质，我想，读者日后学习量子场论时回头再来看这里的推导一定更能领悟这一推导的巧妙。

举例

不妨举一例子。上一章在讲费马原理时我们引入过这样一个光程泛函 $S[y(x)] = \int dx L(y, y')$ ，式中

$$L(y, y') = n(y) \sqrt{1 + (y')^2}. \tag{3.32}$$

显然,从数学上来说,这个例子的变量 x 就相当于我们上面的 t , 函数 $y(x)$ 就相当于上面的 $q(t)$ 。如此来看的话,这个例子显然也有“时间”平移不变性,即光程泛函在 $x \rightarrow \tilde{x} = x + a$ 的变换下不变。因此这个例子也有一个“守恒能量”,为

$$E(y, y') = \frac{\partial L}{\partial y'} y' - L(y, y') = -\frac{n(y)}{\sqrt{1 + (y')^2}}. \quad (3.33)$$

也即是说, $E(y, y')$ 是一个不依赖于 x 的积分常数,称之为欧拉-拉格朗日方程的首积分。显然,这正是我们上一章中找到了的那个首积分。

3.1.4 宇宙的总能量是零

既然已经讲到费马原理的例子,不妨让我们看一下最一般的光程泛函 $S[\mathbf{x}(s)]$ 有什么对称性,又会导致什么守恒定律。这里,

$$S[\mathbf{x}(s)] = \int n(\mathbf{x}) \left| \frac{d\mathbf{x}}{ds} \right| ds. \quad (3.34)$$

很明显,这里有参数 s 的平移不变性。即在 $s \rightarrow \tilde{s} = s + a$ 的变换下, $S[\mathbf{x}(s)]$ 保持不变。因此按照前面关于能量守恒的处理可以知道,必定存在一个守恒的“能量” $E(\mathbf{x}(s), \dot{\mathbf{x}}(s))$ 。

下面要做的就是找到 $E(\mathbf{x}(s), \dot{\mathbf{x}}(s))$ 的表达式。为此,考察如下无穷小变换

$$s \rightarrow \tilde{s}(s) = s + \varepsilon(s), \quad (3.35)$$

这里 \tilde{s} 是原参数 s 的函数。在此变换之下, $S[\mathbf{x}(s)]$ 将变换为

$$\begin{aligned} S[\mathbf{x}(s)] &\rightarrow S[\mathbf{x}(\tilde{s})] = \int n(\mathbf{x}(\tilde{s})) \left| \frac{d\mathbf{x}(\tilde{s})}{d\tilde{s}} \right| d\tilde{s} \\ &= \int n(\mathbf{x}(\tilde{s})) \left| \frac{d\mathbf{x}(\tilde{s})}{ds} \right| ds \\ &= \int n(\mathbf{x}(s)) \left| \frac{d\mathbf{x}(s)}{ds} \right| ds = S[\mathbf{x}(s)]. \end{aligned} \quad (3.36)$$

式中最后一行是将变量 \tilde{s} 重命名为 s 。我们发现,即使在这种修改以后的无穷小变换下, $S[\mathbf{x}(s)]$ 仍然不变!事实上,上面的推导过程和 $\tilde{s}(s)$ 的具体形式没有多大关系,也即是说,其实 $S[\mathbf{x}(s)]$ 在相当任意的 $s \rightarrow \tilde{s}(s)$ 这种重新参数化之下都会保持不变, $S[\mathbf{x}(s)]$ 具有关于 s 的重参数不变性!它意味着,在与“时间平移”有关的 (3.35) 变换下, $S[\mathbf{x}(s)]$ 的改变量将是零,我们写作

$$0 = \delta S = S[\mathbf{x}(\tilde{s})] - S[\mathbf{x}(s)] = \int ds (0) \cdot \dot{\varepsilon}(s). \quad (3.37)$$

因此,这也即是说,“守恒能量” $E(\mathbf{x}(s), \dot{\mathbf{x}}(s))$ 其实为零,即

$$E(\mathbf{x}(s), \dot{\mathbf{x}}(s)) = 0. \quad (3.38)$$

从这个例子我们学到的就是,如果一个物理系统的作用量不仅在时间平移下保持不变,而且实际上在 $t \rightarrow \tilde{t}(t)$ 的任意时间重参数化之下都保持不变。那这个物理系统的总能量将不仅守恒,而且实际上这个总能量必定等于零!宇宙就是一个这样的系统,对宇宙作

用量有贡献的包括引力场和物质场，是广义相对论的一个作用量，它具有所谓的微分同胚不变性，特别的，它具有 $t \rightarrow \tilde{t}(t)$ 的时间微分同胚不变性。因此，根据上面的推理，宇宙的总能量必定是零。这是因为，引力势能是负的，而物质场的能量是正的，正负抵消刚好得到零！

但是，上面的论证有一个漏洞，那就是宇宙毕竟是一个场论系统，不是一个点粒子系统，而场是一种空间分布。这使得当我们对场论系统进行类似上面那样的推导时，系统的渐近空间边界需要额外的特殊对待，要加上边界条件，它使得即使体内的时间有重参数不变性，这种重参数不变性也不能延展到渐近边界上去，这样一来渐近空间边界上就可能对总能量产生贡献。所以，如果宇宙在空间上有渐近边界，那它的总能量就不一定是零。不过，现在人们更倾向于认为宇宙在空间上并没有渐近边界，所以，总能量还是零，至少我们倾向于这么认为。

3.1.5 * 伽利略不变性

在本章的一开始我们设想了一艘在太空中孤独地飞行的宇宙飞船，我们说了，飞船中的人不仅不能辨测空间坐标和方向，他也不能辨测自己是否在飞行。也即是说，可以给飞船一个任意常数速度 \mathbf{v} ，使得 $\mathbf{x} \rightarrow \mathbf{x} + \mathbf{v}t$ ，而飞船中的人不会辨测到任何异常。

$$\mathbf{x} \rightarrow \mathbf{x} + \mathbf{v}t, \quad (3.39)$$

就是所谓的伽利略变换，它将一个惯性系变换到另外一个相对它以速度 \mathbf{v} 匀速飞行的惯性系。伽利略告诉我们，这种变换不会带来力学规律的变化，一切惯性系都是平权的。这种力学规律的伽利略变换不变性就是所谓的相对性原理。那么，伽利略不变性能给出什么守恒定律呢？

为了回答这个问题，我们首先构造一个具体的多粒子作用量，如下

$$S[\mathbf{x}(t)] = \int dt \left[\frac{1}{2} \sum_i m_i \dot{\mathbf{x}}_i^2 - \sum_{i < j} V(\mathbf{x}_i - \mathbf{x}_j) \right]. \quad (3.40)$$

但是，这个作用量在 $\mathbf{x}_i \rightarrow \tilde{\mathbf{x}}_i = \mathbf{x}_i + \mathbf{v}t$ 的伽利略变换下并非不变，而是会变化为

$$\begin{aligned} S[\mathbf{x}(t)] &\rightarrow S[\tilde{\mathbf{x}}(t)] = \int_{t_i}^{t_f} dt \left[\frac{1}{2} \sum_i m_i (\dot{\mathbf{x}}_i + \mathbf{v})^2 - \sum_{i < j} V(\mathbf{x}_i - \mathbf{x}_j) \right] \\ &= S[\mathbf{x}(t)] + \int_{t_i}^{t_f} dt \left[\sum_i m_i \dot{\mathbf{x}}_i \cdot \mathbf{v} \right] + \frac{1}{2} m \mathbf{v}^2 (t_f - t_i) \\ &= S[\mathbf{x}(t)] + \sum_i m_i [\mathbf{x}_i(t_f) - \mathbf{x}_i(t_i)] \cdot \mathbf{v} + \frac{1}{2} m \mathbf{v}^2 (t_f - t_i). \end{aligned}$$

式中 $m = \sum_i m_i$ 为系统的总质量。我们发现，作用量在伽利略变换下多出来了两项。但是，多出来的这两项仅仅影响路径的起末两个端点，而在应用最小作用量原理时这两个端点却是固定的，因此多出来的这两项不会对变分产生任何影响。也即是说，虽然作用量并非伽利略不变，但是 $\delta S = 0$ 的运动微分方程却是伽利略不变的。所以，这里虽然需要略微推广一下我们关于对称性的定义，但伽利略变换的确是我们所构造的这个力学系统的对称性。

下面我们考察无穷小伽利略变换，即设变换速度 $\mathbf{v} = \boldsymbol{\varepsilon}$ 。和前面一样，我们令 $\boldsymbol{\varepsilon}$ 为依赖于时间 t 的任意无穷小函数，记为 $\boldsymbol{\varepsilon}(t)$ 。即是说，我们考虑如下变换

$$\mathbf{x}_i \rightarrow \tilde{\mathbf{x}}_i = \mathbf{x}_i + \boldsymbol{\varepsilon}(t)t. \quad (3.41)$$

在这个变换的作用下，作用量 $S[\mathbf{x}(t)]$ 变换为

$$\begin{aligned} S[\mathbf{x}(t)] \rightarrow S[\tilde{\mathbf{x}}(t)] &= \int_{t_i}^{t_f} dt \left[\frac{1}{2} \sum_i m_i (\dot{\mathbf{x}}_i + \boldsymbol{\varepsilon} + \dot{\boldsymbol{\varepsilon}}t)^2 - \sum_{i < j} V(\mathbf{x}_i - \mathbf{x}_j) \right] \\ &= S[\mathbf{x}(t)] + \int_{t_i}^{t_f} dt \left[\sum_i m_i \dot{\mathbf{x}}_i \right] \cdot \boldsymbol{\varepsilon}(t) + \int_{t_i}^{t_f} dt \left[\sum_i m_i \dot{\mathbf{x}}_i \right] t \cdot \dot{\boldsymbol{\varepsilon}} \end{aligned}$$

对第二行中间那项分部积分，就可以得到

$$\begin{aligned} \delta S &= S[\tilde{\mathbf{x}}(t)] - S[\mathbf{x}(t)] \\ &= - \int_{t_i}^{t_f} dt \left[\sum_i m_i \mathbf{x}_i \right] \cdot \dot{\boldsymbol{\varepsilon}} + \int_{t_i}^{t_f} dt \left[\sum_i m_i \dot{\mathbf{x}}_i \right] t \cdot \dot{\boldsymbol{\varepsilon}} \\ &= \int_{t_i}^{t_f} dt \left[\left(\sum_i m_i \dot{\mathbf{x}}_i \right) t - \sum_i m_i \mathbf{x}_i \right] \cdot \dot{\boldsymbol{\varepsilon}}. \end{aligned} \quad (3.42)$$

上面这个式子意味着， $(\sum_i m_i \dot{\mathbf{x}}_i)t - \sum_i m_i \mathbf{x}_i$ 是守恒量，不妨记为 \mathbf{K} 。很显然 $\sum_i m_i \dot{\mathbf{x}}_i$ 就是系统的总动量 \mathbf{p} ， $\mathbf{p} = \sum_i m_i \dot{\mathbf{x}}_i$ 。由于伽利略不变的系统一定具有空间平移不变性，所以这个总动量 \mathbf{p} 本身就是守恒量。另外，我们可以引入质心坐标 \mathbf{x}_c ，从而有 $\sum_i m_i \mathbf{x}_i = m\mathbf{x}_c$ 。因此守恒量 \mathbf{K} 其实就是

$$\mathbf{K} = \mathbf{p}t - m\mathbf{x}_c. \quad (3.43)$$

不过， \mathbf{K} 守恒其实是显然的，因为 $\frac{d\mathbf{K}}{dt} = \mathbf{p} - m\frac{d\mathbf{x}_c}{dt} = 0$ 显然成立，我们上面推导的关键不过是指出这个守恒量和伽利略对称性密切相关。

3.2 对称性决定作用量

到现在为止，我们应该已经比较清楚作用量的重要性了，有了作用量的具体表达式，利用最小作用量原理就能得出运动微分方程。有了作用量里的拉格朗日量，经过勒让德变换就是得到描述系统能量的哈密顿量。更一般的，作用量的不变性意味着物理定律的对称性。总之，从作用量出发讨论物理问题非常简洁，也非常方便。问题是，如何写出一个物理系统的作用量呢？回答是，基本靠猜。其实，牛顿运动定律也是猜的，麦克斯威方程组也是猜的，从根本上来说，要得到一个物理系统的运动微分方程，我们都是靠猜。而面对一个全新的物理系统，猜它的作用量可比猜运动微分方程容易多了！我想，上一章我们关于洛伦兹力的讨论足以说明这一点。

更何况，通常来说物理学家猜作用量可不是漫无目的的，物理学家们作猜测时有一些基本的原则，其中最重要的就是根据对称性。只要我们能预先猜测出一个物理系统的对称性，那这些对称性往往就能决定（至少是极大地限制）作用量的形式，这是因为，作用量必

须得在对称变换下保持不变，作用量泛函不是一个任意的泛函，而是一种具有对称性的特殊泛函。

在非相对论物理中对称性对作用量有所限制，但这个限制还不够强，这就是为什么爱因斯坦之前的物理学很少谈对称性的原因。但是，相对论有很大的不同，相对论意味着一种很强的时空对称性。相对论不变性再结合一些其它的对称性，往往就能决定作用量。所以杨振宁先生曾经说：“对称性决定相互作用”，指的就是对称性往往能决定作用量。为了给读者示例对称性如何决定作用量，下面我们从相对论不变性开始谈起。

3.2.1 相对论不变性

本小节讨论相对论不变性，这里我们合适地选取时间的单位，使得光速 $c = 1$ 。因此质量的量纲（记为 $[m]$ ）和能量量纲（记为 $[E]$ ）相同，因为 $[m] = [mc^2] = [E]$ 。同样，时间的量纲和空间长度的量纲也相同。

如大家知道的，爱因斯坦基于两条基本的原理建立起了整个狭义相对论。这两条基本原理分别是：1. 相对性原理。即物理定律在所有的惯性系中有相同的形式。这当然是伽利略相对性原理的简单推广。2. 光速不变原理。即在任何给定的惯性系中，光速 c 总是一样的，并与光源的运动无关。

现在让我们考虑两个惯性系，分别称为 S 系和 S' 系，假定同一个事件在这两个参考系中的时空坐标分别是 (t, x, y, z) 和 (t', x', y', z') 。那么根据相对性原理，我们可以知道，这两个参考系之间的坐标变换一定是一个线性变换。这是因为，麦克斯韦方程作为一组特定的物理学定律必须在这两个参考系中取相同的形式，而麦克斯韦方程是一组关于时间和空间坐标的一阶偏微分方程，它如果要保持形式不变的话，除非两组时空坐标之间的变换是线性变换才可能。

另外，根据光速不变原理我们可以知道，在从 S 系到 S' 系的变换中， t' 不仅依赖于原来的时间坐标 t ，而且还一定同时依赖于原来的空间坐标 \mathbf{x} （与牛顿的绝对时间观念是完全不同的）。这是因为，根据光速不变原理，在 S 系中不同地点同时发生的两个事件在 S' 看来是不同时的，这就是所谓的的同时的相对性。

间隔不变性

现在，设想在 S 系中，从 (t, x, y, z) 点发出一束光到达 $(t + dt, x + dx, y + dy, z + dz)$ 点，由于 $c = 1$ ，显然我们有

$$-dt^2 + dx^2 + dy^2 + dz^2 = 0. \quad (3.44)$$

根据光速不变原理，同样的两个事件在 S' 系看来也得满足

$$-dt'^2 + dx'^2 + dy'^2 + dz'^2 = 0. \quad (3.45)$$

换言之，两个邻近的事件在两个参考系中的时空坐标必得满足一个约束关系，即当 (3.44) 成立时必有 (3.45) 成立，反之亦然。又由于相对性原理所告诉我们的，两参考系之间的坐标变换是线性变换，因此，对任意的两个邻近事件，我们必有

$$-dt'^2 + dx'^2 + dy'^2 + dz'^2 = D(\mathbf{v})(-dt^2 + dx^2 + dy^2 + dz^2), \quad (3.46)$$

式中 \mathbf{v} 是 S' 系相对于 S 系的速度。同样的, 由于 S 与 S' 地位平等, 如果从 S' 变换到 S , 我们就有

$$-dt'^2 + dx'^2 + dy'^2 + dz'^2 = D(-\mathbf{v})(-dt^2 + dx^2 + dy^2 + dz^2). \quad (3.47)$$

换言之, 我们必有 $D(-\mathbf{v})D(\mathbf{v}) = 1$ 。又由于空间的各向同性可知, D 对相对速度 \mathbf{v} 的依赖只能是依赖于其大小 v , 而必定和其方向无关, 因此我们必定有 $D(-\mathbf{v}) = D(\mathbf{v}) = D(v)$ 。因此, $(D(v))^2 = 1$, 即 $D(v) = \pm 1$ 。又由于 $D(v)$ 是 v 的连续函数, 而且 $D(0) = 1$ (对应 S 和 S' 变成同一个参考系), 因此我们必有 $D(v) = 1$ 。因此, 对于任意两个邻近事件我们必有

$$-dt'^2 + dx'^2 + dy'^2 + dz'^2 = -dt^2 + dx^2 + dy^2 + dz^2. \quad (3.48)$$

通常将 $-dt^2 + dx^2 + dy^2 + dz^2$ 称为两个邻近事件的时空间隔, 简称间隔, 并记为 $-d\tau^2$, 即

$$-d\tau^2 = -dt^2 + dx^2 + dy^2 + dz^2 = -dt^2 + d\mathbf{x}^2. \quad (3.49)$$

用这个记号, 方程 (3.48) 就可以简记成

$$d\tau'^2 = d\tau^2, \quad (3.50)$$

称为两个事件的间隔不变性。假如有一个粒子从 (t, x, y, z) 运动到 $(t + dt, x + dx, y + dy, z + dz)$, 那么其间隔 $d\tau^2$ 就等于一个固定在粒子上的钟在这个过程中所走过的时间的平方 (因为在粒子本身的参考系中这个钟的空间坐标改变量为零), 所以这时候 $d\tau$ 就是这个钟走过的时间, 称为粒子的固有时²。

人们通常约定 $t = x^0, x = x^1, y = x^2, z = x^3$ 这样就把四个时空坐标统一地记成了 $x^\mu, \mu = 0, 1, 2, 3$ 。利用这个记号, 我们就可以将间隔的计算公式重写为

$$-d\tau^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \quad (3.51)$$

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

dx^μ 这样的四维时空的上指标量就称为四矢量, 一个一般的四矢量可以记为 A^μ 。但是, 我们也可以利用 $\eta_{\mu\nu}$ 将 A^μ 的指标降下来, 即约定

$$A_\mu = \eta_{\mu\nu} A^\nu. \quad (3.52)$$

反过来, 人们还会引入一个 $\eta^{\mu\nu}$, 它同样满足 $-\eta^{00} = \eta^{11} = \eta^{22} = \eta^{33} = 1$, 其它指标分量都等于零。利用 $\eta^{\mu\nu}$ 我们也可以反过来将 A_μ 的指标升上去, 即约定

$$A^\mu = \eta^{\mu\nu} A_\nu. \quad (3.53)$$

²从这里的讨论读者可以看到, 间隔的定义和是否有粒子运动无关, 而固有时的定义却依赖于一个运动的粒子。因此, 这是两个不同的概念, 人们常常用不同的记号表示它们。

人们也可以定义两个四矢量 A^μ, B^μ 之间的内积, 记为 $A \cdot B$, 其定义为

$$A \cdot B = A^\mu B_\mu = A_\mu B^\mu = \eta_{\mu\nu} A^\mu B^\nu. \quad (3.54)$$

两个四矢量的内积是参考系变换不变的, 称为洛伦兹标量。

洛伦兹变换

狭义相对论中, 两个参考系之间的时空坐标变换称作洛伦兹变换, 根据前面所说, 它是一个保持间隔不变性的线性变换, 通常写成

$$x'^\mu = \Lambda^\mu_\nu x^\nu, \quad (3.55)$$

式中 Λ^μ_ν 构成变换矩阵 Λ 的分量形式。

下面我们来构造一个具体的洛伦兹变换。为此, 让我们考虑一种特殊情况, 假如 S' 系沿着 S 系的 x 轴正方向以匀速 v 运动。假设我们这么选取 S' 的坐标轴, 以使得初始时两个坐标系的坐标轴完全重合。这样一来, 由于两坐标系的相对运动只发生在 x 方向上, 与 y, z 方向无关, 所以显然有 $y' = y, z' = z$ 。因此, 我们只需要考虑事件的 t 坐标和 x 坐标在参考系变换下如何变换, 为此不妨暂时忽略间隔公式 (3.49) 中的 y, z 坐标, 进而将间隔公式简化为

$$-d\tau^2 = -dt^2 + dx^2 = dx^+ dx^-, \quad (3.56)$$

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

$$x' = \cosh(\omega)x - \sinh(\omega)t, \quad (3.57)$$

$$t' = -\sinh(\omega)x + \cosh(\omega)t. \quad (3.58)$$

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

$$t' = \frac{t - vx}{\sqrt{1-v^2}}, \quad (3.59)$$

$$x' = \frac{x - vt}{\sqrt{1-v^2}}, \quad (3.60)$$

再加上前面的 $y' = y, z' = z$, 这四个式子就是最常用的一组洛伦兹变换。

3.2.2 相对论粒子的作用量

相对论不变性告诉我们, 相对论粒子的作用量得是洛伦兹变换不变的! 唯一的和粒子坐标有关的这种不变量就是粒子的固有时 $d\tau$ 。所以, 粒子的作用量必定正比于 $d\tau$ 沿着粒

子运动路径的积分。固有时具有时间量纲，而作用量的量纲为能量量纲乘以时间量纲，刚好粒子质量 m 是能量量纲，从而我们知道，相对论粒子的作用量必定可以写成

$$S[x(s)] = -m \int d\tau = -m \int ds \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds}}. \quad (3.61)$$

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

值得指出的是，上述作用量 (3.61) 显然具有重参数不变性，即在 $s \rightarrow \tilde{s}(s)$ 的参数变换下保持不变。这意味着我们可以在一定意义上任意选择参数 s 。最常见的选择有两种，第一种是，取 $s = \tau$ ，即取固有时本身为路径的参数，从而作用量 (3.61) 可以写成

$$S[x(\tau)] = -m \int d\tau = -m \int d\tau \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}}. \quad (3.62)$$

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

$$d\tau^2 = dt^2 - d\mathbf{x}^2 = dt^2(1 - \mathbf{v}^2). \quad (3.63)$$

式中 $\mathbf{v} = \frac{d\mathbf{x}}{dt}$ 是粒子的速度。从而就可以将作用量 (3.61) 写成

$$S[\mathbf{x}(t)] = -m \int dt \sqrt{1 - \mathbf{v}^2}. \quad (3.64)$$

假设粒子的运动速度远低于光速，即 $\mathbf{v} \ll 1$ ，那这时候就可以利用关于 \mathbf{v}^2 的泰勒展开将相对论的作用量 (3.64) 近似成

$$S = -m \int dt + \int dt \frac{1}{2} m \mathbf{v}^2 + \dots \quad (3.65)$$

省略号表示 \mathbf{v}^2 的高阶项。很显然，除了相差一个对变分没有影响的常数项 $-m \int dt$ 之外，这个近似作用量正是非相对论的自由粒子的作用量！

另外，上述相对论粒子作用量显然具有 $x^\mu \rightarrow \tilde{x}^\mu = x^\mu + a^\mu$ (a^μ 为常矢量) 的时空坐标平移不变性。不妨以固有时参数的 (3.62) 来进一步讨论。为了考察这一时空平移对称性所对应的守恒量，我们考虑如下无穷小变换

$$x^\mu \rightarrow \tilde{x}^\mu = x^\mu + \varepsilon^\mu(\tau) \Leftrightarrow \delta x^\mu = \varepsilon^\mu(\tau). \quad (3.66)$$

在此变换下，作用量的改变量为

$$\begin{aligned} \delta S &= -m \int \delta(d\tau) = -m \int \frac{1}{2} \frac{\delta(d\tau^2)}{d\tau} \\ &= m \int \frac{1}{2} \delta(\eta_{\mu\nu} dx^\mu dx^\nu) / d\tau \\ &= m \int \frac{dx^\mu}{d\tau} \eta_{\mu\nu} \delta(dx^\nu) \\ &= m \int d\tau \left(\frac{dx^\mu}{d\tau} \right) \dot{\varepsilon}_\mu. \end{aligned} \quad (3.67)$$

这意味着与时空坐标平移对应的守恒量是 $m \frac{dx^\mu}{d\tau}$, 记作 p^μ ,

$$p^\mu = m \frac{dx^\mu}{d\tau}. \quad (3.68)$$

利用固有时 τ 和坐标时 t 之间的关系式 (3.63) 容易得出

$$p^0 = \frac{m}{\sqrt{1-\mathbf{v}^2}}, \quad \mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-\mathbf{v}^2}}. \quad (3.69)$$

另一方面, 我们知道, 时间坐标平移对应能量守恒, 空间坐标平移对应动量守恒。因此 p^0 就是相对论粒子的能量, \mathbf{p} 就是相对论粒子的动量。这两者一起构成了一个四矢量 p^μ , 称作四动量。很容易看出, 四动量满足如下关系

$$-p^\mu p_\mu = (p^0)^2 - \mathbf{p}^2 = m^2. \quad (3.70)$$

如果允许引入一个辅助性的力学变量 $e(s)$, 那我们还可以将作用量 (3.61) 写成一个更加顺眼的形式,

$$S[e(s), x(s)] = \frac{1}{2} \int ds \left(e^{-1} \eta_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} - em^2 \right). \quad (3.71)$$

为了证明这个作用量与前面的 (3.61) 相等价, 人们只要首先利用下式消去辅助力学变量 $e(s)$ 即可

$$\frac{\delta S}{\delta e(s)} = 0. \quad (3.72)$$

这个过程就和我们上一章从相空间最小作用量原理推导坐标空间最小作用量原理时所做的类似。

3.2.3 与电磁场的耦合

上一小节所讨论的只不过是一个自由的相对论粒子。而实际的相对论粒子会和诸如引力场或者电磁场这样的基本力场发生相互作用, 如何写出一个包含了相互作用的作用量呢? 这一小节我们以电荷为 q 的带电粒子与电磁场的相互作用为例来说明这一点。要点依然是通过考虑对称性, 不过, 这时候我们要考虑的是电磁场的规范对称性。

正如上一章涉及过的, 在电动力学中, 我们会引入所谓的矢量势 \mathbf{A} , 它和电势 ϕ 一起构成了一个四矢量 A^μ , 它的定义是 $A^0 = \phi, A^i = A_i (i = 1, 2, 3)$ 。但是 A^μ 本身不是物理的, 真正物理上可测量的是电场强度 \mathbf{E} 和磁场强度 \mathbf{B} , 这两者可以统一成一个两指标的所谓二阶反对称张量 $F_{\mu\nu}$, 具体来说即是 $F_{i0} = E_i, F_{ij} = \epsilon_{ijk} B^k (i, j = 1, 2, 3)$, 其中 F_{ij} 我们已经在上一章碰到过了。 $F_{\mu\nu}$ 与 A_μ 的关系是

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \quad (3.73)$$

式中 $\partial_\mu = \frac{\partial}{\partial x^\mu}$ 。

从 $F_{\mu\nu}$ 的关系式 (3.73) 可以看出, 它在如下变换下保持不变

$$A_\mu \rightarrow A_\mu + \partial_\mu \epsilon. \quad (3.74)$$

式中 $\varepsilon(x)$ 为时空坐标 x^μ 的任意函数。这个变换就称之为电磁场的规范变换，从这个变换可以知道， A_μ 本身不是物理的，因为不同的相差一个规范变换的 A_μ 描述了同样的电磁场强度。

我们也可以用微分形式的语言，将四矢量 A_μ 写成四维时空上的电磁场 1-形式 A ，

$$A = A_\mu dx^\mu. \quad (3.75)$$

用这套微分形式的语言，规范变换其实就是如下变换

$$A \rightarrow A + d\varepsilon. \quad (3.76)$$

带电粒子和电磁场相互作用，它的作用量也得在这个规范变换下保持不变，这就是所谓的规范对称性。

为了满足规范对称性，我们当然可以只用 $F_{\mu\nu}$ 来和带电粒子相耦合，因为它是规范不变的。但我们同时还要求作用量具有洛伦兹不变性，也就是说作用量得是一个洛伦兹标量，因此不能有没被内积掉的指标存在。 $F_{\mu\nu}$ 关于两个下标反对称，因此最简单的洛伦兹标量 $F_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds}$ 实际上等于零。用 $F_{\mu\nu}$ 构造出来的最简单非零洛伦兹标量是 $F_{\mu\nu} F^{\mu\nu}$ ，很显然它是电磁场场强的二次方项，如果电磁场不是非常强，那它就是一个二阶小量，那这样的项实际上贡献就非常小，因此可以从带电粒子的作用量中忽略！

看起来好像没什么不可忽略的既规范不变又洛伦兹不变的项可以加到相对论粒子的作用量中去！这当然有问题，因为带电粒子显然和电磁场有不可忽略的相互作用。仔细思考以后我们可能会发现下面这一项

$$+q \int_a^b A. \quad (3.77)$$

它就是电磁场 1-形式沿着带电粒子时空运动路径的积分，式中我们假设这条路径起于 a 点终止于 b 点。这一项显然是洛伦兹不变的，那它规范不变吗？很明显不是，因为在规范变换 (3.76) 的作用下，它会变为

$$q \int_a^b A \rightarrow q \int_a^b A + q \int_a^b d\varepsilon \quad (3.78)$$

但是，我们发现多出来的部分是一个全微分，它完全取决于路径的两个端点，即

$$q \int_a^b d\varepsilon = q(\varepsilon(b) - \varepsilon(a)). \quad (3.79)$$

而我们早就知道，在使用最小作用量原理时，路径的两个端点是固定不变的，因此规范变换下多出来的这一项对变分完全没有贡献。即是说，虽然给作用量加上的这一项 (3.77) 不是规范不变的，但是，由它导出来的运动微分方程却是规范不变的！所以，(3.77) 这一项实际上符合要求！

因此，我们可以写出相对论性带电粒子完整的作用量，为

$$\begin{aligned} S[x(s)] &= -m \int d\tau + q \int A \\ &= -m \int ds \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds}} + q \int ds A_\mu \frac{dx^\mu}{ds}. \end{aligned} \quad (3.80)$$

如果将参数 s 取成坐标时 t , 并考虑 $\mathbf{v} \ll 1$ 的非相对论极限, 那上式就可以近似成

$$S[\mathbf{x}(t)] = -m \int dt + \int dt \left[\frac{1}{2} m \mathbf{v}^2 - q\phi + q\mathbf{A} \cdot \mathbf{v} \right] + \dots \quad (3.81)$$

显然, 除了前面那个无关紧要的常数项外, 这个作用量正是我们上一章纯粹碰运气猜出来的带电粒子的作用量。但是现在, 在一定的意义上我们是通过对称性的考虑推导出了它!

3.2.4 对称性决定一切?

相对论不变性加上规范不变性不仅可以决定带电粒子的作用量, 它也可以决定电磁场本身的作用量, 从而也就决定了电磁场的麦克斯威方程。电磁场的规范对称性后来被杨振宁和米尔斯所推广, 变成所谓的非阿贝尔规范对称性。最后人们发现, 某种非阿贝尔规范对称性决定了强相互作用, 而另一种非阿贝尔规范对称性决定了弱相互作用。不仅如此, 爱因斯坦的另外一个伟大发现就是, 万有引力定律也遵循某种奇妙的“对称性”, 这种“对称性”既决定了处于引力场中的粒子的作用量, 也决定了引力场本身的作用量, 从而也就决定了引力场本身的爱因斯坦方程。

到此为止, 我们可以说, 自然界中的四种基本相互作用, 电磁相互作用、弱相互作用、强相互作用、以及万有引力相互作用, 它们的作用量都是由对称性所决定的。对称性是否决定了一切呢? 这个问题我们无法回答, 不过我们可以回顾一下, 决定了万有引力规律的是何种对称性, 并谈谈今天人们对自然界的对称性又有什么新的猜测。

万有引力和惯性力之间的对称性

爱因斯坦曾经向别人讲述过他所谓的“这辈子最幸福的想法”, 他说(大意): 有一天他坐在专利局里, 思考如何推广他的狭义相对论, 使之能自洽地包含万有引力³。突然, 他想到一个人在电梯里随着电梯自由下落的情形。爱因斯坦仿佛可以切身体会到这个人感受到的失重, 他意识到, 在这个自由下落的人看来, 地球重力不存在了, 因为它被加速运动的惯性力抵消了! 爱因斯坦抓住了这个念头的关键点, 即是说, 惯性力可以抵消万有引力, 即是说, 惯性力和万有引力其实是不可区分的!

这就是万有引力所遵循的奇妙对称性。根据这种对称性, 对于一个在外太空加速运动的人而言, 他会感受到一个加速运动的惯性力, 但是, 假设他不能向外看, 不知道飞船在加速, 他完全可以认为自己是在一个引力场中。所以伽利略发现, 只要不向外看, 飞船的速度就是不可辨测的, 而爱因斯坦进一步发现, 飞船的加速度也是不可辨测的, 因为它等效于一个引力场!

当然, 这里有一个微妙的关键, 即惯性力和万有引力之间的这种等价性只对局限在一个局部小区域(比如一部电梯)之内的观察者才成立。即是说, 万有引力的这种奇妙对称性是一种局域对称性(和电磁场的规范对称性有点类似)。爱因斯坦称他发现的这种万有引力的局域对称性为等效原理, 意即, 局域而言, 惯性力和万有引力等效。

³那时候人们只知道电磁力和万有引力, 强相互作用力和弱相互作用力还没发现。电磁力天然与狭义相对论自洽, 但是牛顿的万有引力理论是一种超距作用力, 即是说牛顿认为万有引力的传播不需要时间, 这和狭义相对论所说的相互作用的传播速度不能超过光速矛盾了!

将等效原理所描述的这种对称性和狭义相对论不变性结合起来，就能决定万有引力场的作用量，也就是所谓的希尔伯特-爱因斯坦作用量。对这个作用量应用最小作用量原理，就能导出引力场的爱因斯坦方程。而这一整个理论就是著名的广义相对论。

玻色子与费米子的对称性？

我们已经看到了对称性的神奇力量。人们当然想继续延伸这种力量的作用范围，使之无远弗届。在现代物理中，对称性观念的最重要推广也许要算超对称。读者也许知道，自然界中的基本粒子可以分成两大类，即费米子和玻色子，像电子这样的粒子就是费米子，而像光子这样的粒子就是玻色子。费米子和玻色子性质截然不同，两者在物理理论中的作用也不一样。然而，超对称说，也许费米子与玻色子之间也存在着一种对称性，也许每一个玻色子都有一个费米子伴侣，反之也一样，也许玻色子和费米子在理论中的作用是平等的。

当然，超对称依然是一个猜测，目前还没有被实验所验证。但是，物理学家们的确构造出了很多具有超对称性的理论模型，这些模型中有些能解决现代物理中一些重大的难题。不仅如此，超对称的理论模型由于具有更大的对称性，因此理论的性质也会更好，所以常常成为人们求解物理问题以及发展物理想法的玩具模型，当前数学物理研究的一个重要方向就是研究各种超对称场论模型。我们这里无法用简单的几句话描绘出超对称的美妙，因为这需要在量子场论中才能看清楚。



4. 两体问题和受限三体问题

前面几章初步发展了经典力学的理论框架，下面当然要用这个框架来处理具体问题。因此这一章我们将研究如何求解两体问题以及一类特殊的三体问题。需要说明的是，这一章的内容我想不出什么与众不同的讲述方式。但是作为一本正经的经典力学书，不讲这些内容无论如何是不行的。因此，这一章的处理可能和已有的理论力学书或讲义没有太大不同。

4.1 两体问题

太阳系的九大行星如何运动呢？这问题其实非常复杂，答案是未知的。因为这是一个多体问题，而这样的多体问题并没有一般性的解析解。不过我们知道，太阳的质量比行星大很多，任何行星主要是受太阳的万有引力，其它行星的影响往往可以忽略。如此一来，行星的运动问题就变成了一个只需考虑这颗行星本身和太阳的相互作用的两体问题。开普勒告诉我们，行星相对太阳是在作椭圆轨道运动。几百年前的牛顿和我们一样也知道这个结论，但是，牛顿面临的难题是，如何用理论推理得出行星必定作椭圆轨道运动，这就是所谓的开普勒问题，本章的核心之一就是要解决这一问题。

假设有两个质点，质量分别为 m_1, m_2 ，位置矢量分别为 $\mathbf{x}_1, \mathbf{x}_2$ 。两质点间有相互作用势能，它仅仅依赖于两质点间的距离 $|\mathbf{x}_1 - \mathbf{x}_2|$ ，记为 $V(|\mathbf{x}_1 - \mathbf{x}_2|)$ ，则这个系统的拉格朗日量可以写成

$$L = \frac{1}{2}m_1\dot{\mathbf{x}}_1^2 + \frac{1}{2}m_2\dot{\mathbf{x}}_2^2 - V(|\mathbf{x}_1 - \mathbf{x}_2|). \quad (4.1)$$

很显然, 这个系统有空间平移不变性, 因此总动量守恒, 即是说系统的质心作匀速直线运动。因此, 我们可以在质心参考系中考察问题。不妨设 $\mathbf{x}_1, \mathbf{x}_2$ 为两质点在这个质心系中的位置矢量, 从而有

$$m_1 \mathbf{x}_1 + m_2 \mathbf{x}_2 = 0, \quad (4.2)$$

另外, 我们再引入相对坐标 \mathbf{x} ,

$$\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{x}. \quad (4.3)$$

由这两个方程容易解出

$$\mathbf{x}_1 = \frac{m_2}{m_1 + m_2} \mathbf{x} = \frac{m}{m_1} \mathbf{x}, \quad \mathbf{x}_2 = -\frac{m_1}{m_1 + m_2} \mathbf{x} = -\frac{m}{m_2} \mathbf{x}. \quad (4.4)$$

式中 m 称为约化质量, 其定义为

$$m = \frac{m_1 m_2}{m_1 + m_2} \Leftrightarrow \frac{1}{m} = \frac{1}{m_1} + \frac{1}{m_2}. \quad (4.5)$$

将 $\mathbf{x}_1, \mathbf{x}_2$ 的表达式代入系统的拉格朗日量 (4.1), 就可以将这个拉格朗日量在质心系中简化为

$$L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2} m \dot{\mathbf{x}}^2 - V(|\mathbf{x}|). \quad (4.6)$$

这样一来, 就把原来的两体问题约化成了一个单体问题, 不过, 这个单体的质量是约化质量 m , 其位置矢量为原来两个质点的相对位置 \mathbf{x} 。

很明显, (4.6) 式不显含时间, 从而必有能量守恒! 另外, (4.6) 式显然在 \mathbf{x} 的空间旋转之下保持不变, 因此有角动量守恒, 记这个守恒的角动量为 \mathbf{J} , $\mathbf{J} = \mathbf{x} \times \mathbf{p}$ 。从矢量叉乘的性质容易知道, \mathbf{J} 必定始终和位置矢量 \mathbf{x} 相垂直。而 \mathbf{J} 是守恒的, 它指向三维空间的一个固定方向, 因此 \mathbf{x} 必定始终和这个固定方向垂直, 如此一来, \mathbf{x} 只能处于与这个方向垂直的平面上。即是说, 这个约化的单体运动必定是平面运动。

我们可以在这个运动平面上取极坐标, 从而将 $d\mathbf{x}^2$ 写成, $d\mathbf{x}^2 = dr^2 + r^2 d\phi^2$, 进而就可以将拉格朗日量 (4.6) 在极坐标中写成

$$L(r, \dot{r}, \dot{\phi}) = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\phi}^2) - V(r). \quad (4.7)$$

这个拉格朗日量显然不依赖于 ϕ , 即在 $\phi \rightarrow \phi + a$ 下保持不变 (由于 ϕ 是极角, 因此这其实就是旋转不变), 因此按照上一章关于对称性与守恒量的处理容易知道, $\frac{\partial L}{\partial \phi}$ 必定是守恒量。和旋转对称性对应的守恒量当然就是上面的角动量, 这里记为 J , 即

$$J = \frac{\partial L}{\partial \dot{\phi}} = m r^2 \dot{\phi} \quad (4.8)$$

是守恒量。

4.2 中心力场

4.2.1 中心力场的一般讨论

上一节出现的力场 $V(r)$ 就是中心力场，即是说，中心力场其实就是势能函数仅仅依赖于径向坐标 r ，从而具有空间旋转不变性的力场， $r=0$ 的坐标原点就是力心。上一节其实就是将两体问题约化为了一个在中心力场中运动的单体问题。我们已经看到，这样一个系统的角动量 $J = mr^2\dot{\phi}$ 是守恒的，另外系统的能量也是守恒的，即

$$E = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\phi}^2) + V(r) = \frac{1}{2}m\dot{r}^2 + V(r) + \frac{J^2}{2mr^2} \quad (4.9)$$

是一个常数。

中心力场的角动量守恒其实和著名的开普勒第二定律是一回事。根据开普勒第二定律，行星的掠面速度（即行星在单位时间内扫过的面积）是一个常数，记行星扫过的面积为 A ，即有

$$\frac{dA}{dt} = \frac{1}{2}|\mathbf{x} \times d\mathbf{x}|/dt = \frac{1}{2}|\mathbf{x} \times \mathbf{v}| \quad (4.10)$$

是一个常数。但是很显然，

$$\frac{dA}{dt} = \frac{1}{2}|\mathbf{x} \times \mathbf{v}| = \frac{1}{2}|\mathbf{x} \times m\mathbf{v}|/m = \frac{J}{2m}, \quad (4.11)$$

所以这正是角动量守恒。

所以中心力场真正需要仔细处理的其实是径向运动，即 r 如何随时间 t 演化。从 (4.9) 式容易看出，假设我们定义

$$V_{\text{eff}}(r) = V(r) + \frac{J^2}{2mr^2}, \quad (4.12)$$

(其中 $\frac{J^2}{2mr^2}$ 称为离心势能)，那么这一径向运动问题其实就是在有效势能 $V_{\text{eff}}(r)$ 中运动的一维问题。 r 就是这个一维运动的坐标，而这个“一维运动”粒子的守恒能量为

$$E = \frac{1}{2}m\dot{r}^2 + V_{\text{eff}}(r). \quad (4.13)$$

从这个式子很容易得到

$$\frac{dr}{dt} = \sqrt{\frac{2}{m}[E - V_{\text{eff}}(r)]}, \quad (4.14)$$

积分就可以得出 $r(t)$ 。而由角动量守恒，我们又有

$$\frac{d\phi}{dt} = \frac{J}{mr^2}, \quad (4.15)$$

进一步积分就能得出 $\phi(t)$ ，到此为止，原则上我们就可以完全解出中心力场问题了。

但是，很多时候我们更关心的其实是粒子的运动轨道。为此我们消去 (4.14) 式和 (4.15) 式中的 dt ，进而可以知道中心力场中粒子的轨道 $r(\phi)$ 由下面的积分给出

$$\phi = \phi_0 + \int^r dr' \frac{(J/r'^2)}{\sqrt{2m[E - V_{\text{eff}}(r')]}}, \quad (4.16)$$

很显然，只要知道了 $V_{\text{eff}}(r)$ ，就能求出粒子的运动轨道。

为此我们来研究一下 (4.12) 式定义的有效势能 $V_{\text{eff}}(r)$ 。首先很明显，由于离心势能在 $r \rightarrow 0$ 时趋于 $+\infty$ ，因此只要 $V(r)$ 在 $r \rightarrow 0$ 时不趋于 $-\infty$ 或者趋于 $-\infty$ 的速度不够快，那 $V_{\text{eff}}(r)$ 就会满足

$$r \rightarrow 0: V_{\text{eff}}(r) \rightarrow +\infty. \quad (4.17)$$

另外，合理的势能当然会随着 $r \rightarrow +\infty$ 而衰减为零，所以 $V_{\text{eff}}(r)$ 也应该满足，

$$r \rightarrow +\infty: V_{\text{eff}}(r) \rightarrow 0. \quad (4.18)$$

满足这两个要求的典型 $V_{\text{eff}}(r)$ 曲线如图 (4.1) 所示。图中能量水平线 E 和势能曲线的交点

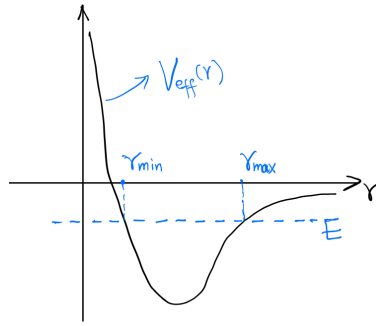


Figure 4.1: 有效势能曲线。

r_{\min} 和 r_{\max} 对应 $\dot{r} = 0$ ，粒子的运动束缚在 $r \in [r_{\min}, r_{\max}]$ 的区间上。 r_{\min} 和 r_{\max} 分别对应中心势场中粒子运动的近心点和远心点。但是从图中也可以看到，如果粒子的能量 E 过大，那 r_{\max} 点就不存在了，这时候粒子就不再做束缚运动，而是可以跑到无穷远处去了。

即使对于图 (4.1) 中所示的束缚运动，粒子的运动轨道一般也不闭合，而是如图 (4.2) 所示的那样，在半径区间为 $[r_{\min}, r_{\max}]$ 的一个环形区域内不断进动。从公式 (4.16) 可以知道，粒子半径每次从 r_{\max} 变到 r_{\min} 然后再回到 r_{\max} ，它的矢径都会扫过一个 $\Delta\phi$ 角，

$$\Delta\phi = 2 \int_{r_{\min}}^{r_{\max}} dr \frac{(J/r^2)}{\sqrt{2m[E - V_{\text{eff}}(r)]}}. \quad (4.19)$$

从图 (4.2) 中可以看出，这个角度通常并不等于 2π ，它和 2π 的差值就是所谓的近心点在一周之内的进动角。通常来说这个进动角并不等于零。关于进动角的近似计算，请读者参看朗道《力学》第三章的习题。

公式 (4.16) 中的平方根在近心点和远心点处改变符号，因此，如果将力心到近心点（或者远心点）的方向选为极轴， ϕ 角从这里开始算起，则很显然，在这个极轴的两侧，对于同一 r 值， ϕ 角的区别仅仅在于正负号不同。也即是说，有心力场中粒子的轨道关于这一从力心到近心点的轴是对称的。

什么情况下会出现闭合轨道呢？有一个伯特兰定理 (Bertrand Theorem) 说：只有两类中心力场，其中一切束缚运动都是闭合轨道，第一类是 $V(r) = -\frac{k}{r}$ ，第二类是 $V(r) = kr^2$ ，

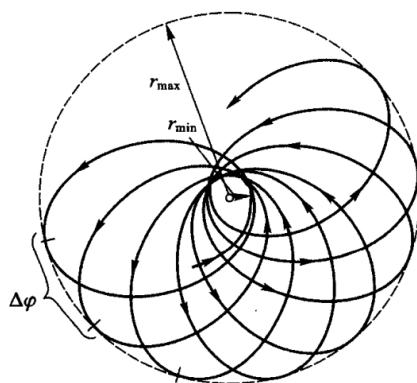


Figure 4.2: 不闭合的轨道。图片来自朗道《力学》。

这里系数 $k > 0$ 。第二类其实就是所谓的三维谐振子，而很幸运的是，万有引力属于第一类，正因为如此行星绕太阳才是一个闭合轨道。（注意，伯特兰定理并没有排除其它势能函数在某些特定情况下也可能出现闭合轨道。）

实际上，对于万有引力情形，行星的轨道不仅是闭合轨道，而且开普勒第一定律告诉我们，它还是椭圆轨道。怎么证明这是一个椭圆轨道呢？这就是我们下面要求解的开普勒问题。

4.2.2 开普勒问题

吸引力

求解开普勒问题的关键就在于对于 $V(r) = -\frac{k}{r}$ 的引力势能计算出公式 (4.16) 中的积分。这时候有效势能 $V_{\text{eff}}(r)$ 为，

$$V_{\text{eff}}(r) = -\frac{k}{r} + \frac{J^2}{2mr^2}, \quad (4.20)$$

这正好属于前面讨论过的典型势能曲线。而我们要计算的就是

$$\phi = \phi_0 + \int^r dr' \frac{(J/r'^2)}{\sqrt{2m[E + \frac{k}{r'} - \frac{J^2}{2mr'^2}]}}, \quad (4.21)$$

为此，进行 $r' = \frac{1}{u}$ 的变量代换，得积分

$$\begin{aligned} \phi &= \phi_0 - \int^{1/r} du \frac{1}{\sqrt{2mE/J^2 + u2mk/J^2 - u^2}} \\ &= \phi_0 - \int^{1/r} du \frac{1}{\sqrt{m^2k^2e^2/J^4 - (u - mk/J^2)^2}} \\ &= \phi_0 + \arccos \left[\frac{1}{e} \left(\frac{p}{r} - 1 \right) \right] \end{aligned} \quad (4.22)$$

式中

$$e = \sqrt{1 + \frac{2EJ^2}{mk^2}}, \quad p = \frac{J^2}{mk}. \quad (4.23)$$

由此即有

$$\frac{p}{r} = 1 + e \cos(\phi - \phi_0). \quad (4.24)$$

这是以力心为焦点的二次曲线的极坐标方程, e 就是所谓的离心率。从 (4.23) 式可知, 这个离心率依赖于粒子的总能量 E , 当 $E > 0$ 时, $e > 1$, 这时候 (4.24) 描述的是一条双曲线, 它说明粒子沿着双曲线的一支运动, 这显然不是束缚运动。当 $E = 0$ 时, $e = 1$, (4.24) 描述的是一条抛物线, 这也不是束缚运动。有意思的是 $E < 0$ 情形, 这时候 $e < 1$, (4.24) 描述的是椭圆, 很明显这是束缚轨道, 它也正是我们想要证明的, 行星绕太阳做椭圆轨道运动。

如果以力心到近日点的方向为极坐标的极轴, 则方程 (4.24) 可以进一步简化成

$$\frac{p}{r} = 1 + e \cos(\phi). \quad (4.25)$$

对于椭圆轨道, 从 (4.25) 式很容易看出, 近心点和远心点分别为

$$r_{\min} = \frac{p}{1+e}, \quad r_{\max} = \frac{p}{1-e}. \quad (4.26)$$

由 $r_{\min} + r_{\max} = 2a$, 可以求出椭圆的半长轴 a , 为

$$a = \frac{p}{1-e^2} = \frac{k}{2|E|}. \quad (4.27)$$

可见, 半长轴仅仅依赖于系统的总能量 E 和角动量 J 无关。由 $b = \sqrt{a^2 - c^2} = a\sqrt{1-e^2}$ 可以进一步求出半短轴, 和 a 不同, 它既依赖于 E 又依赖于 J 。

下面我们来证明开普勒第三定律。为此我们首先计算椭圆的面积

$$\begin{aligned} A &= \pi ab = \pi a^2 \sqrt{1-e^2} = \pi p^2 (1-e^2)^{-3/2} \\ &= \pi p^{\frac{1}{2}} \left(\frac{p}{1-e^2} \right)^{3/2} = \pi \frac{J}{(mk)^{\frac{1}{2}}} a^{3/2}. \end{aligned} \quad (4.28)$$

另一方面, 行星的掠面速度等于 $\frac{1}{2}r^2\dot{\phi} = \frac{J}{2m}$ 。将椭圆面积除以掠面速度, 就得到行星运动的周期 T

$$T = 2\pi \sqrt{\frac{m}{k}} a^{3/2}. \quad (4.29)$$

注意到对于万有引力, $k = GMm$ (M 为太阳质量), 代入上式即得 $T = 2\pi \sqrt{\frac{1}{GM}} a^{3/2}$, 这正是开普勒第三定律¹。

¹实际比这要稍微复杂一点, 其实我们的 m 是约化质量, 假设记太阳质量为 $M = m_1$, 行星质量为 m_2 , 那么其实 $k = Gm_1m_2$, 而 $m = \frac{m_1m_2}{m_1+m_2}$, 所以结果其实应该是 $T = 2\pi \sqrt{\frac{1}{G(m_1+m_2)}} a^{3/2}$ 。但是, 注意到 $m_1 \gg m_2$, 所以近似有 $T = 2\pi \sqrt{\frac{1}{GM}} a^{3/2}$ 。

排斥力

但是, 如果两个质点之间的中心力场不是一个吸引力, 而是排斥力 (比如同性带电粒子之间的库伦排斥力), 即 $V(r) = \frac{k}{r} (k > 0)$, 那情况会略微有些不同。这时候有效势能为

$$V_{\text{eff}}(r) = \frac{k}{r} + \frac{J^2}{2mr^2}, \quad (4.30)$$

这个势能恒大于零, 从而必有系统总能量 $E > 0$ 。另外, 与前面吸引力情形同样的算积分, 现在将会有

$$\frac{p}{r} = -1 + e \cos(\phi). \quad (4.31)$$

式中 e, p 依然由 (4.23) 式给出。由于现在 E 必定大于零, 从而必有 $e > 1$, 从而 (4.31) 描述的必定是双曲线, 粒子沿着双曲线的一支运动, 只不过, 现在力心位于这一支的外焦点上。当然, 这种情况粒子的运动必定是无界的, 不可能是束缚运动。近心点 r_{\min} 为

$$r_{\min} = \frac{p}{e-1} = a + c = a(1+e). \quad (4.32)$$

拉普拉斯-龙格-楞次矢量

前面说过, 闭合轨道是非常特殊的, 那么为什么行星绕太阳运动的轨道是一个闭合的椭圆轨道呢? 伯特兰定理告诉我们, 这是因为 $V(r) = -k/r$ 这种形式的势能的特殊性。那种势能特殊在哪儿呢? 回答是, 这种 $1/r$ 形式的势能不仅有空间旋转对称性, 它还有更高的隐藏的对称性, 通常称作动力学对称性。这种动力学对称性的存在意味着除了角动量守恒以外, 系统还有额外的守恒量。这个额外的守恒量就是所谓的拉普拉斯-龙格-楞次矢量, 记为 \mathbf{R} , 它的定义是

$$\mathbf{R} = \mathbf{v} \times \mathbf{J} - k \frac{\mathbf{x}}{r}. \quad (4.33)$$

为了证明 \mathbf{R} 的确是守恒量, 我们进行如下计算 (注意角动量是守恒的)

$$\begin{aligned} \frac{d}{dt}(\mathbf{v} \times \mathbf{J}) &= \dot{\mathbf{v}} \times \mathbf{J} = -k \frac{1}{m} \frac{\mathbf{x}}{r^3} \times \mathbf{J} = -k \frac{1}{r^3} \mathbf{x} \times (\mathbf{x} \times \mathbf{v}) \\ &= -k \frac{1}{r^3} [(\mathbf{x} \cdot \dot{\mathbf{x}})\mathbf{x} - (\mathbf{x}^2)\dot{\mathbf{x}}] = -k \frac{1}{r^3} [(r\dot{r})\mathbf{x} - r^2\dot{\mathbf{x}}] \\ &= k \left[-\frac{\mathbf{x}}{r^2} \dot{r} + \frac{\dot{\mathbf{x}}}{r} \right] = k \frac{d}{dt} \left(\frac{\mathbf{x}}{r} \right). \end{aligned} \quad (4.34)$$

从计算结果很明显可以看出 \mathbf{R} 是守恒量。

实际上, \mathbf{R} 始终从力心指向近心点。为了看出这一点, 我们选取 \mathbf{R} 方向为极轴, 从而

$$\mathbf{x} \cdot \mathbf{R} = r|\mathbf{R}|\cos\phi. \quad (4.35)$$

另一方面, $\mathbf{x} \cdot \mathbf{R} = \mathbf{x} \cdot (\mathbf{v} \times \mathbf{J}) - k \frac{\mathbf{x}^2}{r} = (\mathbf{x} \times \mathbf{v}) \cdot \mathbf{J} - kr = \frac{J^2}{m} - kr$ 。结合上面的式子, 就可以得到

$$\frac{J^2}{m} - kr = r|\mathbf{R}|\cos(\phi) \Leftrightarrow \frac{J^2}{mk}/r = 1 + \frac{|\mathbf{R}|}{k} \cos(\phi). \quad (4.36)$$

很显然, 最后这个式子正是椭圆轨道的 (4.25) 式。这个推导过程说明了两件事情: 第一, 由于 (4.25) 式正是以力心到近心点的方向为极轴, 所以这说明了 \mathbf{R} 的确指向近心点。第二, 它说明了闭合的椭圆轨道的确来源于额外的这个守恒量。

4.2.3 散射问题

开普勒问题更关心的是束缚轨道，散射问题则关心的是散射运动。这时候我们以一个粒子为散射粒子，另外一个粒子为靶粒子，我们要考察的是散射粒子相对于靶粒子的运动。我们知道，对于这一相对运动而言，只需要取质心参考系并将散射粒子质量取成约化质量 m ，则两体散射问题就约化成了一个单体问题。这时候可以将靶粒子的位置看成是“固定不动”的力心，初始时散射粒子从无穷远处朝着靶粒子入射，受靶粒子产生的中心力场作用以后，最后又以一个偏转角 θ 飞到无穷远处去（设 $r \rightarrow +\infty$ 时 $V(r) \rightarrow 0$ ），如图 (4.3) 所示。和所有的中心力场问题一样，散射过程当然也有能量守恒和角动量守恒，不过人们通常用粒子在无穷远处的速度 v_0 以及入射粒子的瞄准距离 b 来刻画这两个守恒量。所谓的瞄准距离，就是力心 O 到入射粒子初动量 \mathbf{v}_0 方向的垂直距离，如图 (4.3) 所示。很明显，能量 E 以及角动量 J 与 v_0, b 的关系为

$$E = \frac{1}{2}mv_0^2, \quad J = mv_0b. \quad (4.37)$$

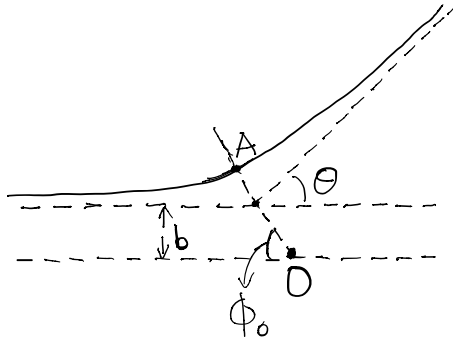


Figure 4.3: 散射

前面说过，对于这样的中心力场问题，粒子的运动轨道关于从力心到近心点的轴线（即图 (4.3) 中的 OA 线）对称。如果记散射粒子轨道的两条渐近线与此轴线的夹角为 ϕ_0 ，则很显然，粒子散射后的偏转角（也叫散射角） θ 满足

$$\theta = \pi - 2\phi_0. \quad (4.38)$$

从前面关于中心力场的一般讨论容易知道， ϕ_0 可以由下式算出

$$\phi_0 = \int_{r_{min}}^{\infty} dr \frac{(J/r^2)}{\sqrt{2m[E - V_{eff}(r)]}} = \int_{r_{min}}^{\infty} \frac{(b/r^2)dr}{\sqrt{1 - 2V(r)/(mv_0^2) - b^2/r^2}}. \quad (4.39)$$

由这两个式子原则上就能算出散射角 θ 对入射粒子速度 v_0 以及瞄准距离 b 的依赖关系。通过实验测量验证这个依赖关系就能验证我们的理论，特别是能得到很多关于中心力场 $V(r)$ 的信息。

卢瑟福通过他的 α 粒子散射实验最早表明了散射问题的重要性，卢瑟福的工作不仅是建立正确的原子模型的基础，而且在一定意义上它也是整个粒子物理的开端。不过，卢

瑟福不是直接去测量 θ 与 v_0, b 的关系, 而是引入了一个全新的更容易测量的量, 这就是散射截面。

具体来说, 就是考虑一束 (而不是单独的一个) 同样速度的入射粒子, 它们的瞄准距离不同且均匀分布, 然后测量单位时间有多少粒子被散射到 θ 方向的单位立体角内。很显然, 如果入射粒子越多, 那散射的就会越多, 散射粒子数一定正比于入射粒子流强 I (也就是单位时间单位横截面积上的入射粒子数), 因此为了抓住真正物理的信息, 实验上测的是散射粒子数与入射流强 I 的比值。如果记单位时间散射到 θ 方向的立体角 $d\Omega = 2\pi \sin \theta d\theta$ 之内 (整个立体角形成一个环绕水平方向的空心圆锥) 的粒子数为 dN , 那实验上测的就是下面这个量

$$d\sigma = \frac{dN}{I}. \quad (4.40)$$

很容易看出, 这个量有面积的量纲, 所以叫微分散射截面。

散射角 θ 是瞄准距离 b 的函数, 两者一一对应。假设空心圆锥立体角 $d\Omega = 2\pi \sin \theta d\theta$ 对应瞄准距离 $[b, b+db]$, 即是说瞄准距离在这个区间之内的入射粒子都被散射到圆锥立体角 $d\Omega = 2\pi \sin \theta d\theta$ 中了。则很显然, 这样的入射粒子构成一个以 b 为内径 $b+db$ 为外径的面积为 $2\pi b db$ 的圆环, 从而 $dN = I(2\pi b db)$, 从而

$$d\sigma = 2\pi b db = 2\pi b(\theta) \left| \frac{db}{d\theta} \right| d\theta = \frac{b(\theta)}{\sin \theta} \left| \frac{db}{d\theta} \right| d\Omega, \quad (4.41)$$

式中的绝对值符号来源于 $b(\theta)$ 通常是一个减函数, 而我们定义的微分散射截面始终是正的。很显然, 只要知道了 b 和 θ 之间的函数关系, 就能算出微分散射截面。反过来, 测量微分散射截面就可以间接验证 b 和 θ 之间的函数关系。

卢瑟福公式

假设散射粒子和靶粒子之间的力场为库伦排斥力场, $V(r) = \frac{k}{r} (k > 0)$, α 粒子和原子核的散射正是这种情形。则我们容易算出公式 (4.39) 中的积分, 为

$$\phi_0 = \arccos \left(\frac{\frac{k}{mv_0^2 b}}{\sqrt{1 + \left(\frac{k}{mv_0^2 b}\right)^2}} \right). \quad (4.42)$$

从而即有

$$b^2 = \left(\frac{k}{mv_0^2} \right)^2 \tan^2 \phi_0. \quad (4.43)$$

根据 (4.38) 式, 即有

$$b^2 = \left(\frac{k}{mv_0^2} \right)^2 \cot^2 \left(\frac{\theta}{2} \right). \quad (4.44)$$

将这个式子两边微分, 并代入微分散射截面的公式 (4.41), 就可以得到

$$d\sigma = \left(\frac{k}{2mv_0^2} \right)^2 \frac{1}{\sin^4 \left(\frac{\theta}{2} \right)} d\Omega. \quad (4.45)$$

这就是著名的卢瑟福公式。

然而, 我们知道, 对于 α 粒子这样的微观粒子, 正确的描述应该用量子力学。因此上面用经典物理推导出来的卢瑟福公式并不能保证一定对。但是, 这里碰巧的是, 量子力学推导出来的结果和上面经典物理的结果完全一样。这种巧合正是因为 $1/r$ 这种类型的势场是特殊的, 有隐藏着的更高对称性。

4.3 受限三体问题和拉格朗日点

前两节通过将两体问题约化为单体问题, 再通过能量守恒和角动量守恒, 我们已经完全求解出了两体问题。按照道理来说, 下面要研究的自然应该是三体问题, 即三个质量分别为 m_1, m_2, m_3 的质点通过万有引力相互作用的问题。但是三体问题是数学和天文学中著名的难题, 庞加莱等人曾经证明过, 一般性的三体问题不存在解析解。所以人们自然想到研究某种极端简化后的三体问题, 具体来说即是研究三体中的某一体质量相比来说非常小 (因此其引力场可以忽略) 的情形, 即

$$m_1 \gg m_3, \quad m_2 \gg m_3. \quad (4.46)$$

这就是所谓的受限三体问题, 或者有时候也称作限制性三体问题。

在实际应用中, 受限三体问题非常重要, 因为比方说, 太阳-地球-月亮就可以看成这样的一个受限三体系统, 因为月亮的质量相对很小。另外, 人们常常将航天器发射到日地系统或者地月系统中, 这时候航天器和日地系统或者和地月系统一起也构成了受限三体系统。然而, 即使对于受限三体问题, 一般性的求解也非常困难, 事实上, 1889 年庞加莱正是在研究受限三体问题时, 首次发现了动力系统混沌。所以下面我们也不是要一般性地求解受限三体问题, 而是将注意力集中于受限三体问题的一个重要概念, 叫做拉格朗日点。

从上面的论述可以知道, 受限三体问题就是 m_1 和 m_2 这两个大质量的物体形成两体运动, 然后 m_3 在这个两体的引力场中运动的问题。 m_3 质量很小, 因此对 m_1, m_2 的反作用可以忽略。这里进一步考虑一种简单情形, 我们假设, m_1, m_2 绕着它们的质心做匀速率圆轨道运动, 两者的旋转角速度为 ω , 很显然它满足

$$m\omega^2 R = G \frac{m_1 m_2}{R^2}, \quad (4.47)$$

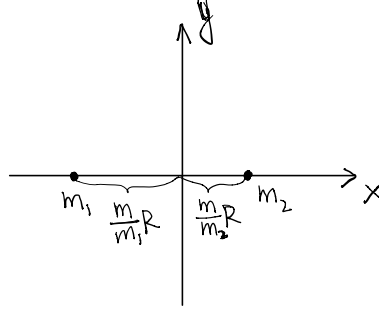
式中 m 为约化质量, R 为 m_1, m_2 两者之间的距离。由上式即有

$$\omega^2 = \frac{G(m_1 + m_2)}{R^3}. \quad (4.48)$$

另外, 由前面关于两体问题的分析可以知道, m_1, m_2 与质心的距离 r_1, r_2 分别为

$$r_1 = \frac{m}{m_1} R, \quad r_2 = \frac{m}{m_2} R. \quad (4.49)$$

下面进一步简化问题, 假设 m_3 的运动限制在 m_1, m_2 的旋转平面上。为了分析 m_3 的具体运动, 我们取 m_1, m_2 的质心为原点, 取旋转平面为 $x-y$ 平面, 取参考系 (坐标系)

Figure 4.4: m_1, m_2 所构成的旋转参考系。

在这个平面上以角速度 ω 旋转，并让 m_1, m_2 位于这个旋转参考系的 x 轴上，如图 (4.4) 所示。记 m_3 相对这个旋转参考系的速度为 $\mathbf{v}' = (\dot{x}, \dot{y})$ ，它在这个旋转参考系中的坐标为 $\mathbf{x} = (x, y)$ ，则 m_3 相对静止惯性系的速度 \mathbf{v} 应该等于 \mathbf{v}' 再加上参考系旋转的速度 $\omega \times \mathbf{x}$ ，即

$$\mathbf{v} = \mathbf{v}' + \omega \times \mathbf{x}. \quad (4.50)$$

从而 m_3 的拉格朗日量 L 为

$$\begin{aligned} L &= \frac{1}{2}m_3(\mathbf{v}' + \omega \times \mathbf{x})^2 - V \\ &= \frac{1}{2}m_3\mathbf{v}'^2 + m_3\mathbf{v}' \cdot (\omega \times \mathbf{x}) - \left[V - \frac{1}{2}m_3(\omega \times \mathbf{x})^2\right] \\ &= \frac{1}{2}m_3(\dot{x}^2 + \dot{y}^2) + m_3\omega(x\dot{y} - y\dot{x}) - V_{\text{eff}}. \end{aligned} \quad (4.51)$$

式中有有效势能 V_{eff} 的定义为

$$V_{\text{eff}} = V - \frac{1}{2}m_3(\omega \times \mathbf{x})^2 = V - \frac{1}{2}m_3\omega^2(x^2 + y^2), \quad (4.52)$$

其中的第二项代表的是离心势能。而 V 代表的是 m_3 受 m_1, m_2 的万有引力势能，即

$$V = -G\frac{m_1m_3}{r_{13}} - G\frac{m_2m_3}{r_{23}}, \quad (4.53)$$

式中

$$r_{13}^2 = \left(x + \frac{m}{m_1}R\right)^2 + y^2, \quad r_{23}^2 = \left(x - \frac{m}{m_2}R\right)^2 + y^2. \quad (4.54)$$

利用拉格朗日方程容易得出 m_3 在旋转参考系中的运动微分方程，为

$$\begin{aligned} m_3\ddot{x} - 2m_3\omega\dot{y} &= -\frac{\partial V_{\text{eff}}}{\partial x} \\ m_3\ddot{y} + 2m_3\omega\dot{x} &= -\frac{\partial V_{\text{eff}}}{\partial y}. \end{aligned} \quad (4.55)$$

上述方程左边的第二项代表的就是所谓的科里奥利力。这组方程的一般性求解不是一件容易的事情，下面我们考虑它的平衡解，即 $\dot{x} = \dot{y} = \ddot{x} = \ddot{y} = 0$ 的解。相应的解就是所谓的拉格朗日点，显然它们相应于有效势能的极值点，满足如下方程

$$\begin{aligned} 0 &= \frac{\partial V_{\text{eff}}}{\partial x} = Gm_1m_3 \frac{x + Rm/m_1}{r_{13}^3} + Gm_2m_3 \frac{x - Rm/m_2}{r_{23}^3} - m_3\omega^2 x \\ 0 &= \frac{\partial V_{\text{eff}}}{\partial y} = Gm_1m_3 \frac{y}{r_{13}^3} + Gm_2m_3 \frac{y}{r_{23}^3} - m_3\omega^2 y. \end{aligned} \quad (4.56)$$

下面要做的就是求解方程 (4.56)。首先求 $y = 0$ 的解，这时候 (4.56) 式的第二个方程自动满足，而第一个方程相当于

$$\omega^2 x = Gm_1 \frac{x + Rm/m_1}{|x + Rm/m_1|^3} + Gm_2 \frac{x - Rm/m_2}{|x - Rm/m_2|^3}. \quad (4.57)$$

这个方程有三个解，图 (4.5) 分别画出了这个方程左边和右边的函数曲线，从交点可以清楚地看到这三个解。也即是说，在 m_1, m_2 的连线上有三个拉格朗日点，分别位于 m_1 的左

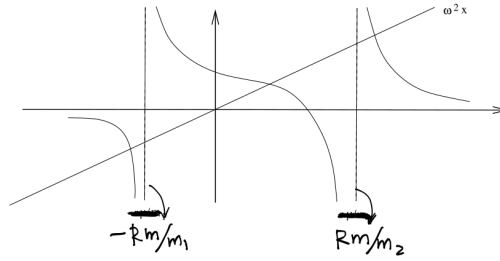


Figure 4.5: $y = 0$ 对应的 x 的三个解。图片来自 David Tong 的在线讲义 classical dynamics.

侧、 m_2 的右侧、以及 m_1, m_2 的中间，记为 L_3, L_2, L_1 。这三个拉格朗日点最早是欧拉发现的。

下面求方程 (4.56) $y \neq 0$ 的解。这时候由 (4.56) 的第二个方程可以得到

$$G \frac{m_2}{r_{23}^3} = \omega^2 - G \frac{m_1}{r_{13}^3}. \quad (4.58)$$

代入 (4.56) 的第一个方程，就可以得到

$$\omega^2 = \frac{G(m_1 + m_2)}{r_{13}^3} = \frac{G(m_1 + m_2)}{r_{23}^3}. \quad (4.59)$$

与前面的 (4.48) 式比较，即有

$$r_{13} = r_{23} = R. \quad (4.60)$$

即 m_3 要和 m_1, m_2 构成等边三角形，显然，有两个点满足这样的要求，分别记为 L_4, L_5 。这就是拉格朗日发现的两个拉格朗日点。

所以，总共有五个不同的拉格朗日点，其分布如图 (4.6) 所示。研究表明， L_1, L_2, L_3 这

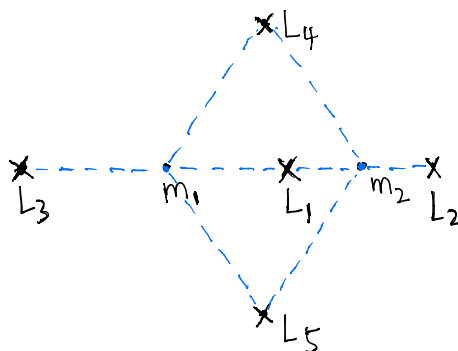


Figure 4.6: 五个不同的拉格朗日点。

三个拉格朗日点是不稳定平衡点，而只要 m_2 远远小于 m_1 ，那么由于科里奥利力的存在， L_4, L_5 这两个拉格朗日点将是稳定平衡点。

所有的拉格朗日点都很有用。比如， L_1, L_2, L_3 这三个拉格朗日点虽然不稳定，但是在它们附近的航天器只需要很小的能量消耗就能回复到平衡点附近。在日地系统中（地球为 m_2 ），由于距离地球比较近， L_1 这个拉格朗日点常被用来放置太阳观测卫星，例如 NASA(美国国家航空航天局)1997 年发射的先进成分探测器 (Advanced Composition Explorer, ACE)，就在 L_1 点运行。NASA 和 ESA(欧洲空间局) 联合发射的太阳和日光层探测器 (Solar and Heliospheric Observatory, SOHO) 同样在日地系统的 L_1 处运行。日地系统的 L_2 点同样离地球比较近，因此也被 NASA 和 ESA 用来放置太空天文台。宇宙学领域两颗著名的用于探测微波背景辐射的探测器，WMAP 和 PLANCK 就被放置在日地系统的 L_2 处。类似的，地月系统的这三个拉格朗日点也很有用，比如，2018 年中国发射的登月探测器玉兔二号就利用了地月系统中（月球为 m_2 ，地球为 m_1 ）的 L_2 点来实现其在月背时和地球的中继通信。

太阳和行星系统的 L_4, L_5 点天然就是稳定平衡点，因此可以想见，这两个位置会吸引很多小行星。这些小行星通常被称为特洛伊星体。据统计，木星系统的 L_4, L_5 位置有几千颗这样的小行星。同样，地月系统的 L_4, L_5 点也是稳定平衡点，不令人意外，这两个位置被观测到吸引有星际尘埃。

5. 再谈哈密顿力学

本章讲述哈密顿力学的系统理论，内容包括泊松括号，正则变换，以及哈密顿-雅可比方程。这一章和通常教材的不同在于，处理方式更为现代，语言也更为现代，因此更有利于读者日后阅读现代文献。

这一章中包含了一些漂亮的定理，包括诺特定理的重新回顾，刘维尔定理，还有庞加莱回归定理。只要不是单纯的数学定理，而是理论力学框架中的结论，我们都给出了证明，并且我采用的都是一些在我看来最为简洁优雅的证明。

5.1 泊松括号

5.1.1 辛结构

前面的章节中我们引入了哈密顿正则方程，它是一组关于粒子如何在广义坐标和广义动量所构成的相空间中演化的一阶常微分方程，其形式是

$$\dot{q}^a = \frac{\partial H}{\partial p_a}, \quad \dot{p}_a = -\frac{\partial H}{\partial q^a}. \quad (5.1)$$

式中假设系统有 n 个自由度，从而指标 $a = 1, 2, \dots, n$ 。我们看到，广义坐标 q 和广义动量 p 在哈密顿正则方程中地位非常平等，正因为如此，人们也常将 q^a, p_a 称为一对正则变量，同时也称广义坐标为正则坐标，称广义动量为正则动量。但是我们也看到，正则坐标的方程和正则动量的方程**相差了一个负号**！这个负号从哪儿来的呢？

为了理解哈密顿正则方程中这个额外负号的来源，我们可以回想一下相空间的最小作用量原理，我们知道，哈密顿正则方程的形式是由相空间作用量决定的，而这个作用量除了取决于哈密顿量之外，还有一项就是取决于辛势 $\Theta = p_a dq^a$ (默认对 a 求和)。哈密顿正则方程中的负号显然和哈密顿量没什么关系，所以它必定来源于辛势 Θ 。但是，辛势的表达式 $\Theta = p_a dq^a$ 中 p, q 甚至都不完全平等，为了让这两者更平等一点，我们不妨考察 Θ 的外微分，称之为辛形式，记为 ω ，它的具体定义是

$$\omega = d\Theta = dp_a \wedge dq^a. \quad (5.2)$$

辛形式 ω 是一个 2-形式，而且由于它是辛势的外微分，所以是一个恰当形式，从而当然也一定是闭形式，即满足

$$d\omega = 0. \quad (5.3)$$

辛形式 ω 关于正则坐标和正则动量非常对称，但是，由于 $dq^a \wedge dp_a = -dp_a \wedge dq^a$ ，所以 q, p 互换的时候正好会多出一个负号。这其实就是哈密顿正则方程中额外的那个负号的来源。辛形式对于哈密顿正则系统非常重要，原因在于，它是相空间的一个基本结构，因此带有辛形式 ω 的相空间就称之为定义了**辛结构**的相空间。

既然 q, p 地位平等，都是相空间的坐标，我们不妨把它们写得更平等一点。具体来说，即引入 $2n$ 维相空间的坐标 $\mathbf{x} = (x^1, x^2, \dots, x^{2n})^T = (q^1, \dots, q^n, p_1, \dots, p_n)^T$ ，式中 T 表示转置操作，所以 \mathbf{x} 是一个具有 $2n$ 个分量的列矢量，**注意它不是三维坐标矢量**，请读者注意根据上下文区分 \mathbf{x} 这个符号的含义。 \mathbf{x} 的各分量我们记为 $x^i, i = 1, 2, \dots, 2n$ ，前 n 个 i 指标代表正则坐标 q ，后 n 个 i 指标代表正则动量 p 。

有了更为对称的相空间坐标 \mathbf{x} 之后，我们就可以把辛形式重新表示为

$$\omega = dp_a \wedge dq^a \equiv \frac{1}{2} \omega_{ij} dx^i \wedge dx^j. \quad (5.4)$$

式中和本章的所有地方一样使用了**求和约定**，对于重复指标 a ，约定从 1 到 n 求和，对于重复指标 i, j 则约定从 1 到 $2n$ 求和。我们可以把辛形式在坐标 \mathbf{x} 中的分量 ω_{ij} 理解为一个 $2n \times 2n$ 矩阵的第 i 行第 j 列，很明显这个矩阵 (记为 J^{-1}) 为，

$$J^{-1} = (\omega_{ij}) = \begin{pmatrix} 0_{n \times n} & -1_{n \times n} \\ 1_{n \times n} & 0_{n \times n} \end{pmatrix}. \quad (5.5)$$

这是一个反对称矩阵，满足 $\omega_{ji} = -\omega_{ij}$ 。式中 $1_{n \times n}$ 表示 $n \times n$ 的单位矩阵。 J^{-1} 的逆矩阵当然就记为 J (请不要和角动量相混淆)，很容易看出 $J = -J^{-1}$ ，为

$$J = (\omega^{ij}) = \begin{pmatrix} 0_{n \times n} & 1_{n \times n} \\ -1_{n \times n} & 0_{n \times n} \end{pmatrix}, \quad (5.6)$$

这里我们已经把 J 的第 i 行第 j 列记为 ω^{ij} ，即有

$$\omega_{ij}\omega^{jk} = \delta_i^k, \quad \omega^{ij}\omega_{jk} = \delta_k^i. \quad (5.7)$$

式中 δ_j^i 表示单位矩阵的第 i 行第 j 列，即 $i = j$ 时， $\delta_j^i = 1$ ， $i \neq j$ 时， $\delta_j^i = 0$ 。

很容易验证，有了上面由辛形式导出的矩阵 J ，我们就可以将哈密顿正则方程重写为

$$\dot{x}^i = \omega^{ij} \partial_j H, \quad (5.8)$$

式中 $\partial_j = \frac{\partial}{\partial x^j}$ 。或者写成矩阵形式

$$\dot{\mathbf{x}} = J \frac{\partial H}{\partial \mathbf{x}}. \quad (5.9)$$

哈密顿正则方程的这种形式完美地解释了额外的那个负号如何来源于相空间的辛结构。

5.1.2 泊松括号

哈密顿力学如何刻画物理量呢？回答是，很简单，任何物理量都是相空间中的函数，具有 $A(q, p)$ 这样的形式 (这里 q, p 只是示意性的符号，分别代表所有的正则坐标和所有的正则动量)，或者也可以用相空间的 \mathbf{x} 坐标，写作更抽象的 $A(\mathbf{x})$ 。那么随着粒子在相空间中演化，物理量该如何演化呢？为了回答这个问题，我们进行下面的推导

$$\frac{dA}{dt} = \frac{\partial A}{\partial q^a} \dot{q}^a + \frac{\partial A}{\partial p_a} \dot{p}_a = \frac{\partial A}{\partial q^a} \frac{\partial H}{\partial p_a} - \frac{\partial A}{\partial p_a} \frac{\partial H}{\partial q^a}. \quad (5.10)$$

人们通常将上面最后的表达式记为 $[A, H]$ ，称之为泊松括号。即是说，泊松括号的定义为

$$[A, B] = \frac{\partial A}{\partial q^a} \frac{\partial B}{\partial p_a} - \frac{\partial A}{\partial p_a} \frac{\partial B}{\partial q^a}. \quad (5.11)$$

很明显，我们也可以利用前面引入的 \mathbf{x} 坐标以及 ω^{ij} ，进而将这个定义式写为

$$[A, B] = \frac{\partial A}{\partial x^i} \omega^{ij} \frac{\partial B}{\partial x^j} = (\partial_i A) \omega^{ij} (\partial_j B). \quad (5.12)$$

利用泊松括号，就可以将物理量如何随时间演化的方程写成

$$\frac{dA}{dt} = [A, H]. \quad (5.13)$$

特别的，我们可以将哈密顿正则方程写成

$$\dot{q}^a = [q^a, H], \quad \dot{p}_a = [p_a, H]. \quad (5.14)$$

或者更简洁地写成

$$\dot{\mathbf{x}} = [\mathbf{x}, H]. \quad (5.15)$$

上面考察的是最常见的，表达式本身不显含时间的物理量，如果推广到显含时间的情形，即表达式具有 $A(q, p, t)$ 的形式，则物理量的演化方程就要推广成

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + [A, H]. \quad (5.16)$$

按照泊松括号的定义，人们很容易计算出如下基本泊松括号

$$[q^a, q^b] = [p_a, p_b] = 0, \quad [q^a, p_b] = \delta_b^a. \quad (5.17)$$

或者也可以用更抽象的 \mathbf{x} 坐标，写成

$$[x^i, x^j] = \omega^{ij}. \quad (5.18)$$

利用这个基本泊松括号，我们也可以将泊松括号的定义式 (5.12) 重写成

$$[A, B] = (\partial_i A)(\partial_j B)[x^i, x^j]. \quad (5.19)$$

从这里可以看到，泊松括号的定义其实就是普通的偏导运算加上基本变量 \mathbf{x} 的泊松括号。

可以证明，泊松括号有如下性质：

- $[A, B] = -[B, A]$ 。
- 线性性，即 $[A, c_1 B_1 + c_2 B_2] = c_1 [A, B_1] + c_2 [A, B_2]$ ，式中 A, B_1, B_2 为三个任意物理量， c_1, c_2 为两个任意常数。
- 莱布尼兹法则，即 $[A, BC] = [A, B]C + B[A, C]$ 。
- 雅可比恒等式，即 $[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$ 。注意这个式子中的 A, B, C 是轮换的。

以上，1,2 两条性质比较显然。3,4 两条性质需要证明。首先看性质 3，为了证明它，我们注意到，根据泊松括号的定义式 (5.12)， $[A, B]$ 可以看作是对函数 B 的特定偏导运算（当然也可以看作对 A 的偏导运算），因为 $[A, B]$ 具有 $(A^i \partial_i)B$ 的形式，式中系数 A^i 与 A 有关，具体可以从定义式 (5.12) 中读出来。我们记偏导算符 $A^i \partial_i = L_A$ ，则

$$L_A B = [A, B] = (A^i \partial_i)B. \quad (5.20)$$

而性质 3 其实就是偏导算符的莱布尼兹法则，即 $L_A(BC) = (L_A B)C + B(L_A C)$ 。

下面证明性质 4，即证明雅可比恒等式。为此我们注意到，对于任何一个物理量来说，每将它进行一次泊松括号就相当于对它进行一次偏导运算。因此 $[B, C]$ 既是 B 的一阶偏导项，也是 C 的一阶偏导项，而 $[A, [B, C]]$ 很容易看清是 B 的二阶偏导项与 C 的二阶偏导项之和。同样 $[B, [C, A]]$ 是 C 的二阶偏导项与 A 的二阶偏导项之和，而 $[C, [A, B]]$ 则是 A 的二阶偏导项与 B 的二阶偏导项之和。即是说，雅可比恒等式 $[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$ 左边三项全部展开算出来的每一项必定要么是 A 的二阶偏导项，要么是 B 或 C 的二

阶偏导项，即雅可比恒等式的左边是 A, B, C 的二阶偏导项之和。下面我们证明，表达式 $[A, [B, C]] + [B, [C, A]] + [C, [A, B]]$ 中关于 A, B, C 的二阶偏导项均是零，很显然，如果完成了这一证明，那就必定有雅可比恒等式。

这个证明并不难，举例来说，要证明 C 的二阶偏导项为零，我们只需注意， $[C, [A, B]]$ 显然是关于 C 的一阶偏导（ C 只进行了一次泊松括号运算），因此可以不用管，剩下的两项显然有

$$\begin{aligned} [A, [B, C]] + [B, [C, A]] &= [A, [B, C]] - [B, [A, C]] \\ &= L_A L_B C - L_B L_A C = (L_A L_B - L_B L_A) C. \end{aligned} \quad (5.21)$$

但是， $(L_A L_B - L_B L_A)$ 其实是一阶偏导运算，原因在于

$$\begin{aligned} L_A L_B - L_B L_A &= (A^i \partial_i)(B^j \partial_j) - (B^i \partial_i)(A^j \partial_j) \\ &= [A^i (\partial_i B^j) - B^i (\partial_i A^j)] \partial_j. \end{aligned} \quad (5.22)$$

所以 $[A, [B, C]] + [B, [C, A]]$ 其实也是关于 C 的一阶偏导项。因此雅可比恒等式左边关于 C 的二阶偏导项其实为零，由于 A, B, C 的轮换性质，我们同样可以证明，关于 A, B 的二阶偏导项也是零，如此就完成了整个证明！

利用雅可比恒等式，我们有

$$\begin{aligned} \frac{d}{dt}([A, B]) &= [[A, B], H] = -[[B, H], A] - [[H, A], B] \\ &= [[A, H], B] + [A, [B, H]] = \left[\frac{d}{dt} A, B \right] + \left[A, \frac{d}{dt} B \right]. \end{aligned} \quad (5.23)$$

也即是

$$\frac{d}{dt}([A, B]) = \left[\frac{d}{dt} A, B \right] + \left[A, \frac{d}{dt} B \right]. \quad (5.24)$$

如果物理量显含时间，那我们还可以在泊松括号的性质中增加一条，**性质 5**， $\frac{\partial}{\partial t}([A, B]) = \left[\frac{\partial}{\partial t} A, B \right] + \left[A, \frac{\partial}{\partial t} B \right]$ 。这条性质并不难证明，这里留给读者作练习。结合这条性质和雅可比恒等式就可以看出，对于物理量显含时间的情形，依然有 (5.24) 式。

如果某个物理量 A 满足 $\frac{d}{dt} A = 0$ ，它当然就是一个守恒量。因此根据 (5.24) 式可以知道，如果 A 是一个守恒量， B 也是一个守恒量，那么它们的泊松括号 $[A, B]$ 将是一个新的守恒量。另外，从物理量的时间演化方程 (5.13) 容易知道，对于一个不显含时间的物理量 A 来说，它是守恒量的充要条件是

$$[A, H] = 0. \quad (5.25)$$

如果两个物理量的泊松括号等于零，人们就称这两个物理量**泊松对易**。因此，不显含时间的物理量成为守恒量的充要条件是，与哈密顿量泊松对易！

在量子力学中，物理量都是算符（或者说矩阵），这时候两个物理量的泊松括号就变成了算符的对易子，它的定义甚至比泊松括号更为简单。对于两个算符（可以按照矩阵来理解） \hat{A}, \hat{B} ，其对易子就是 $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$ 。不难证明，算符对易子满足泊松括号的所有那些

性质, 甚至更简单, 读者有兴趣的话不妨试着证明一下算符对易子的雅可比恒等式 (提示: 直接按定义展开证明)。算符对易子和泊松括号的这种相似性并非偶然, 它说明, 经典物理和量子物理是有联系的, 经典物理可以看成量子物理在普朗克常数 $\hbar \rightarrow 0$ 时的极限。的确, 可以证明, 在 $\hbar \rightarrow 0$ 时算符对易子自动就会给出经典物理的泊松括号。

下面利用泊松括号的性质算一个例子, 算角动量的泊松括号。为了避免混淆, 这里我们将三维空间位置矢量记为 \mathbf{r} , $\mathbf{r} = (x, y, z)$ 。根据角动量的定义 $\mathbf{J} = \mathbf{r} \times \mathbf{p}$, 可以知道它的分量形式

$$J_1 = yp_z - zp_y, \quad J_2 = zp_x - xp_z, \quad J_3 = xp_y - yp_x. \quad (5.26)$$

从而易有

$$\begin{aligned} [J_1, J_2] &= [yp_z - zp_y, zp_x - xp_z] \\ &= [yp_z, zp_x] + [zp_y, xp_z] - [yp_z, xp_z] - [zp_y, zp_x] \\ &= y[p_z, z]p_x + x[z, p_z]p_y = xp_y - yp_x = J_3. \end{aligned} \quad (5.27)$$

上面的计算过程利用了泊松括号的性质以及基本泊松括号。类似的, 也可以算得

$$[J_1, J_2] = J_3, \quad [J_2, J_3] = J_1, \quad [J_3, J_1] = J_2. \quad (5.28)$$

习惯上, 人们常常利用列维-西维塔符号, 将这个结果概括成

$$[J_i, J_j] = \varepsilon_{ijk} J_k. \quad (5.29)$$

从这个计算可以知道, 如果一个系统的 J_1, J_2 均守恒, 那么它的 J_3 也必定守恒。另外, 利用上面的结果, 也不难证明

$$[\mathbf{J}^2, J_i] = 0. \quad (5.30)$$

5.1.3 涡旋的运动

前面处理的哈密顿系统都是一些通常的单粒子或者多粒子系统, 但是哈密顿系统的适用范围比这要广泛得多。这一节我们将表明, 两维平面流体中多个涡旋所构成的系统是一个哈密顿系统, 而且对于这个哈密顿系统而言, 涡旋运动的正则坐标和正则动量天然就是一回事。

考察一种平面流体 (我们用 $\mathbf{x} = (x, y)$ 表示两维平面的坐标), 假设其中有一个涡旋, 涡旋中心位于坐标原点, 简称有一个位于坐标原点的涡旋。当这样说时, 我们指的是流体围绕着坐标原点在打转, 或者说流速场像图 (5.1) 中所示的那样分布。更精确地说, 位于坐标原点的涡旋指的是, 流速场 $\mathbf{v} = (v_x, v_y)$ 的旋度满足,

$$\partial_x v_y - \partial_y v_x = 2\pi\gamma\delta(\mathbf{x}), \quad (5.31)$$

式中 $\delta(\mathbf{x}) = \delta(x)\delta(y)$ 为两维平面的狄拉克 δ 函数, 而式中的常数 γ 用于表示这个涡旋的强度。

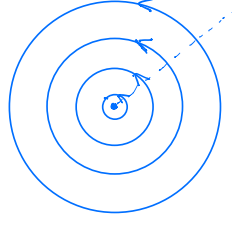


Figure 5.1: 位于坐标原点的一个涡旋。

换言之，位于原点的涡旋指的是这样一种速度场分布，它的旋度在坐标原点处是无穷大，而在其它地方都是零。因此将速度场围绕着半径为 $r = |\mathbf{x}|$ 的圆周 C 积分，将有

$$\begin{aligned}\oint_C (v_x dx + v_y dy) &= \int_D (\partial_x v_y - \partial_y v_x) dx \wedge dy \\ &= 2\pi\gamma \int_D \delta(\mathbf{x}) dx \wedge dy = 2\pi\gamma.\end{aligned}\quad (5.32)$$

式中 D 表示圆周 C 所包围的区域，式中的推导用了斯托克斯公式。

很容易验证，满足要求（即满足 (5.31) 式和 (5.32) 式）的速度场为，

$$v_x = -\gamma \frac{y}{r^2}, \quad v_y = \gamma \frac{x}{r^2}. \quad (5.33)$$

假设有一个质点跟随这个位于原点的涡旋运动，那它的运动方程当然就是

$$\dot{x} = v_x = -\gamma \frac{y}{|\mathbf{x}|^2}, \quad \dot{y} = v_y = \gamma \frac{x}{|\mathbf{x}|^2}. \quad (5.34)$$

下面假设有两个涡旋，分别位于 $\mathbf{x}_1 = (x_1, y_1)$ 和 $\mathbf{x}_2 = (x_2, y_2)$ ，强度分别为 γ_1 和 γ_2 ，并且每一个涡旋都跟随着另一个的速度场运动。那么我们就有方程

$$\begin{aligned}\dot{x}_1 &= -\gamma_2 \frac{y_1 - y_2}{|\mathbf{x}_1 - \mathbf{x}_2|^2}, & \dot{y}_1 &= \gamma_2 \frac{x_1 - x_2}{|\mathbf{x}_1 - \mathbf{x}_2|^2} \\ \dot{x}_2 &= \gamma_1 \frac{y_2 - y_1}{|\mathbf{x}_2 - \mathbf{x}_1|^2}, & \dot{y}_2 &= \gamma_1 \frac{x_2 - x_1}{|\mathbf{x}_2 - \mathbf{x}_1|^2}.\end{aligned}\quad (5.35)$$

但是，速度场是可以矢量叠加的，因此如果有多个涡旋，那每一个涡旋都将跟着所有其它涡旋共同产生的速度场而运动！这样一来，我们就可以将上面的方程推广到多个涡旋的情形，记第 i 个涡旋的位置为 $\mathbf{x}_i = (x_i, y_i)$ ，即有

$$\dot{x}_i = -\sum_{j \neq i} \gamma_j \frac{y_i - y_j}{|\mathbf{x}_i - \mathbf{x}_j|^2}, \quad \dot{y}_i = \sum_{j \neq i} \gamma_j \frac{x_i - x_j}{|\mathbf{x}_i - \mathbf{x}_j|^2}. \quad (5.36)$$

方程 (5.36) 所描述的这个多涡旋系统实际上是一个哈密顿正则系统。其相空间就是多涡旋的位形空间，即 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 所构成的空间。相空间的辛形式可以定义为

$$\omega = \sum_i \gamma_i dy_i \wedge dx_i. \quad (5.37)$$

这里有两点说明：第一，将 ω 定义成 $\sum_i \gamma_i dy_i \wedge dx_i$ 而不是 $\sum_i \gamma_i dx_i \wedge dy_i$ ，这并不是一件重要的事情，因为它仅仅是一种约定。第二，(5.37) 给出的并不是辛形式的标准形式，为了

得到标准形式, 我们只需重新定义一下变量, 具体来说即是令 $p_{x,i} = \gamma_i y_i$, 进而将 (5.37) 式写成 $\omega = \sum_i dp_{x,i} \wedge dx_i$ 。请注意, 在相差一个系数的意义上, y_i 坐标其实是正则动量 $p_{x,i}$ 。

有了辛形式 ω 以后, 我们就很容易写出基本泊松括号

$$[x_i, y_j] = \gamma_i^{-1} \delta_{ij}. \quad (5.38)$$

进而就可以计算任何两个物理量的泊松括号

$$[A, B] = \sum_i \frac{1}{\gamma_i} \left(\frac{\partial A}{\partial x_i} \frac{\partial B}{\partial y_i} - \frac{\partial A}{\partial y_i} \frac{\partial B}{\partial x_i} \right). \quad (5.39)$$

证明 (5.36) 构成哈密顿正则系统的另一个关键是定义如下哈密顿量

$$H = -\frac{1}{2} \sum_{i < j} \gamma_i \gamma_j \log |\mathbf{x}_i - \mathbf{x}_j|^2. \quad (5.40)$$

有了这个哈密顿量以后, 读者不难证明, (5.36) 式可以由如下哈密顿正则方程给出

$$\begin{aligned} \dot{x}_i &= [x_i, H] = \gamma_i^{-1} \frac{\partial H}{\partial y_i} \\ \dot{y}_i &= [y_i, H] = -\gamma_i^{-1} \frac{\partial H}{\partial x_i}. \end{aligned} \quad (5.41)$$

因此, 多个涡旋所构成的动力系统 (5.36) 的确是一个哈密顿正则系统。

不难证明, 下面这两个物理量 P_x, P_y 与 (5.40) 式给出的哈密顿量泊松对易 (即 $[P_x, H] = [P_y, H] = 0$), 从而是守恒量,

$$P_x = \sum_i p_{x,i} = \sum_i \gamma_i y_i, \quad P_y = \sum_i p_{y,i} = -\sum_i \gamma_i x_i. \quad (5.42)$$

(请注意, $p_{x,i}$ 和 $p_{y,i}$ 任何一个都能看作正则动量, 但两者不能同时都看作正则动量。) 从这两个守恒量的定义容易看出, 它们其实就是两个分量的总动量。不过, P_x, P_y 两者之间可不泊松对易, 容易算得, 它们的泊松括号实际上是, $[P_x, P_y] = \sum_i \gamma_i$, 结果是一个常数, 当然也是一个守恒量。

不仅如此, 我们还可以证明, 如下定义的总角动量 L 也是一个守恒量 (即满足 $[L, H] = 0$),

$$L = \frac{1}{2} \sum_i (x_i p_{y,i} - y_i p_{x,i}) = -\frac{1}{2} \sum_i \gamma_i (x_i^2 + y_i^2). \quad (5.43)$$

之所以这个角动量的定义多除了 2, 是因为在现在的系统中, 每个变量都是双重的, 既可以当成正则坐标, 也可以当成正则动量, 这样当我们按通常的方式定义角动量时, 就把这些变量重复考虑了两次。

不难验证, 守恒量 L 和守恒量 P_x, P_y 之间满足如下代数关系,

$$[L, P_x] = P_y, \quad [L, P_y] = -P_x. \quad (5.44)$$

由此可以进一步得到, L 和守恒量 $\mathbf{P}^2 = P_x^2 + P_y^2$ 泊松对易。因此, 再加上哈密顿量本身这个守恒量, 我们就得到了三个两两泊松对易的守恒量, 即 H, L, \mathbf{P}^2 。这三个守恒量使得 2 个

涡旋以及 3 个涡旋的系统都是所谓的可积系统 (关于可积系统的讨论我们放到了下一章)。但是, 4 个以及 4 个以上涡旋的系统就不再是可积系统了, 实际上, 这时候的系统是混沌的¹。

5.2 正则变换

本节将要讲述哈密顿正则系统的一个核心概念, 叫做正则变换, 它是一种保持相空间辛结构的数学变换。关于正则变换有两种不同的观点, 一是, 将之看作描述同一个物理系统的两组不同相空间坐标, 即看成相空间的一类保持辛结构的坐标变换。另一种观点是, 将正则变换看作相空间到其自身的一对一映射 (称之为微分同胚), 这种映射会将一个相空间点映射为另一个不同的相空间点, 因此它会将系统的一个物理状态映射为另一个状态, 类似的, 这种映射也要保持辛结构。**在数学上, 这两种观点完全等价**, 它们可以看成是源自于对相空间辛结构的两种不同理解方式, 但是, 不同的观点会带来对正则变换物理内涵的不同思考, 而且由于这两种思考各有其方便之处, 所以我们分开介绍它们。

5.2.1 正则变换作为相空间坐标变换

首先讨论正则变换的第一种观点。上一节我们看到, 相空间有坐标 $\mathbf{x} = (q, p)$ (q, p 均是示意性的写法), 人们自然想到对相空间进行一个坐标变换, 变到 \mathbf{x}' 坐标。但是, 如果这种坐标变换过于任意, 那 \mathbf{x}' 就可能没有正则坐标和正则动量的结构, 即是说新坐标下无法分别定义正则坐标和正则动量, 这就不是物理上我们想要的。所谓的正则变换, 大体来说就是对相空间坐标变换进行一定的限制, 使得变换以后的新坐标 \mathbf{x}' 依然可以写成 $\mathbf{x}' = (q', p')$ 的形式, 这就是所谓的保持辛结构。

具体来说, 辛形式 ω 作为一个微分形式, 它本身当然不依赖于坐标系, 因此在任何新坐标 \mathbf{x}' 下都必然有

$$\omega = \frac{1}{2} \omega_{ij} dx^i \wedge dx^j = \frac{1}{2} \omega'_{mn} dx'^m \wedge dx'^n. \quad (5.45)$$

式中 ω'_{mn} 为辛形式在新坐标下的分量。利用 $\omega_{ij} dx^i \wedge dx^j = \omega_{ij} \frac{\partial x^i}{\partial x'^m} \frac{\partial x^j}{\partial x'^n} dx'^m \wedge dx'^n$, 与上面的式子比较, 就能得出 ω'_{mn} 应该为

$$\omega'_{mn} = \omega_{ij} \frac{\partial x^i}{\partial x'^m} \frac{\partial x^j}{\partial x'^n}. \quad (5.46)$$

问题是, 上一节我们已经看到, ω_{ij} 不是任意的, 它是前面给出过的特定矩阵 J^{-1} 的第 i 行第 j 列。但如果这种相空间的坐标变换过于任意的话, 那 (5.46) 式给出来的 ω'_{mn} 就不可能依然是矩阵 J^{-1} 的第 m 行第 n 列了。如此一来就无法保持正则坐标和正则动量的定义了 (因为 J^{-1} 的形式正是由于有正则坐标和正则动量才有的), 为了保持正则坐标和正则动量始终有定义, 我们要求 $\omega'_{mn} = \omega_{mn}$ 。即要求坐标变换满足

$$\omega_{mn} = \omega_{ij} \frac{\partial x^i}{\partial x'^m} \frac{\partial x^j}{\partial x'^n}. \quad (5.47)$$

¹更多的讨论请参阅, H. Aref, “Integrable, Chaotic, and Turbulent Vortex Motion in Two-Dimensional Flows”, Ann. Rev. Fluid Mech. 15 345 (1983).

这就叫保持相空间的辛结构!

为了看清楚保持辛结构的条件 (5.47) 有什么重要的含义, 我们将 $\frac{\partial x^i}{\partial x'^m}$ 定义成一个矩阵 D 的第 i 行第 m 列, 记为 $D^i_m = \frac{\partial x^i}{\partial x'^m}$, 则不难看出条件 (5.47) 相当于如下矩阵方程

$$D^T J^{-1} D = J^{-1}. \quad (5.48)$$

式中 D^T 表示矩阵 D 的转置矩阵。将上面这个矩阵方程两边求逆, 即有

$$(D^{-1})J(D^{-1})^T = J. \quad (5.49)$$

由于 $\frac{\partial x^i}{\partial x'^m} \frac{\partial x'^m}{\partial x^j} = \frac{\partial x^i}{\partial x^j} = \delta^i_j$, 所以 D^{-1} 的第 m 行第 j 列就是 $\frac{\partial x'^m}{\partial x^j}$, 因此我们可以将矩阵方程 (5.49) 重写为

$$\frac{\partial x'^m}{\partial x^i} \omega^{ij} \frac{\partial x'^n}{\partial x^j} = \omega^{mn} \Leftrightarrow \partial_i x'^m \omega^{ij} \partial_j x'^n = \omega^{mn}. \quad (5.50)$$

根据上一节关于泊松括号的定义可以知道, 这个结果其实就是

$$[x'^m, x'^n]_{\mathbf{x}} = \omega^{mn}. \quad (5.51)$$

式中我们给泊松括号加上了下角标 \mathbf{x} , 这是为了强调我们是用老正则变量 \mathbf{x} 来定义的泊松括号。我们当然也可以用新正则变量 \mathbf{x}' 来定义泊松括号, 显然就有 $[x'^m, x'^n]_{\mathbf{x}'} = \omega^{mn}$, 将这个式子和上面导出的式子比较, 即有

$$[x'^m, x'^n]_{\mathbf{x}} = \omega^{mn} = [x'^m, x'^n]_{\mathbf{x}'}. \quad (5.52)$$

可见, **相空间坐标变换保持辛结构的要求其实就等价于保持基本泊松括号不变。**

进一步, 从泊松括号的性质可以知道, 如果基本泊松括号在坐标变换下保持不变, 那我们计算任何两个物理量的泊松括号时, 是用新坐标来计算还是用老坐标来计算就是一样的了, 即

$$[A, B]_{\mathbf{x}} = [A, B]_{\mathbf{x}'}. \quad (5.53)$$

即是说, 物理量的泊松括号在正则变换下是不变的! 特别的, 我们将有

$$\dot{\mathbf{x}}' = [\mathbf{x}', H]_{\mathbf{x}} = [\mathbf{x}', H]_{\mathbf{x}'} = J \frac{\partial H}{\partial \mathbf{x}'}. \quad (5.54)$$

这也即是说, **哈密顿正则方程的数学结构在正则变换下将保持不变!**

但是, 值得特别强调的是, 哈密顿正则方程的数学结构虽然在正则变换下保持不变, 但作为常微分方程, 两种不同坐标下的哈密顿正则方程其实是不同的方程, 有不同的常微分方程形式。这是因为, 哈密顿量 H 作为原来变量 \mathbf{x} 的函数, 和作为变量代换后新变量 \mathbf{x}' 的函数, 这两者其实是不同的! 只不过由于它们描写的是同一个物理量, 所以我们才用了同样的符号 H 。如果要强调哈密顿量的数学表达式的话, 那我们应该用一个新的函数 $H'(\mathbf{x}')$ 来表示新坐标下的哈密顿量, 当然由于它和老坐标 \mathbf{x} 下的 $H(\mathbf{x})$ 描写同一个量, 所以有

$$H'(\mathbf{x}') = H(\mathbf{x}). \quad (5.55)$$

也即是说, 正则变换其实改变了哈密顿量的函数形式, 从而改变了动力学方程作为微分方程的形式, 因此, **正则变换一般来说并不是一种对称性!**

5.2.2 正则变换作为相空间的微分同胚映射

现在来讨论正则变换的第二种观点, 这种观点将正则变换看成是相空间到其自身的一种连续映射, 并且它是可逆的, 称作相空间的微分同胚²。不妨将这种微分同胚映射记为 g , 它会将相空间的 \mathbf{x} 点映射为 \mathbf{x}' 点, 即

$$g: \mathbf{x} \rightarrow \mathbf{x}'. \quad (5.56)$$

在这一映射的作用下, \mathbf{x} 点描述的物理状态就映射为 \mathbf{x}' 点描述的物理状态, 因此, 物理量 A 作为相空间点的函数 $A(\mathbf{x})$, 在微分同胚映射的作用下, 其函数形式也要改变, 设改变以后的函数为 $A'(\mathbf{x})$, 记为 $g_*: A \rightarrow A'$, g_* 表示微分同胚映射对物理量的作用。很显然, 映射以后 \mathbf{x}' 点的物理量值其实来自于映射之前的 \mathbf{x} 点, 所以必有

$$A'(\mathbf{x}') = A(\mathbf{x}). \quad (5.57)$$

很显然, 这就是上一小节以哈密顿量为例得到的 (5.55) 式。

不过, 正则变换并非任意的微分同胚映射, 而是保持辛结构的微分同胚映射。为了说清楚这一点, 我们考察微分同胚映射对辛形式 ω 的作用, 记为 $g^*: \omega \rightarrow \omega'$ 。由于 $\omega = \frac{1}{2} \omega_{ij} dx^i \wedge dx^j$, ω_{ij} 是常数系数因此在映射下保持不变, 由此可知, 必有

$$\omega \rightarrow \omega' = \frac{1}{2} \omega_{ij} dx'^i \wedge dx'^j. \quad (5.58)$$

所谓的保持辛结构, 即要求辛形式在映射前后保持不变, 即

$$g^*: \omega \rightarrow \omega' = \omega. \quad (5.59)$$

由此即有 $\omega_{mn} dx'^m \wedge dx'^n = \omega_{ij} dx^i \wedge dx^j$, 也即是

$$\omega_{mn} = \omega_{ij} \frac{\partial x^i}{\partial x'^m} \frac{\partial x^j}{\partial x'^n}. \quad (5.60)$$

这正是上一小节所得到的保持辛结构的条件 (5.47)。

到此为止, 我们已经清楚地看到, 关于正则变换的这第二种观点虽然看起来与第一种观点不同, 但它们在数学上其实完全等价。

我们当然可以将上面的核心结果用更物理的 (q, p) 坐标写出来, 这时候, $\omega = dp_a \wedge dq^a$, 因此 $\omega' = dp'_a \wedge dq'^a$, 因此保持辛结构的要求 (5.59) 就相当于

$$dp'_a \wedge dq'^a = dp_a \wedge dq^a \Leftrightarrow d\Theta' = d\Theta. \quad (5.61)$$

式中 $\Theta' = p'_a dq'^a$ 是正则变换以后的辛势。

上面的结果意味着 $d(\Theta - \Theta') = 0$ 。即正则变换前后的辛势之差是一个闭形式。前面讲微分形式的时候我们说过, 恰当形式必定是闭形式, 但是反过来, 闭形式不一定是恰当形式。但是, 如果只关心空间的局部区域, 而不关心空间整体, 那就有一个重要的庞加莱引

²微分同胚的严格数学定义比这要复杂抽象一些, 感兴趣的读者请参阅拓扑学方面的数学书。

理说，局部上反过来也对，**闭形式局部上也必定是恰当形式**。甚至在理论力学的应用中，这一庞加莱引理成立的相空间局部常常能延展为整个相空间。因此，我们有

$$\Theta - \Theta' = dF \Leftrightarrow p_a dq^a - p'_a dq'^a = dF. \quad (5.62)$$

式中， F 为相空间的某个函数。这一段的讨论告诉我们，**如果某个相空间的微分同胚映射满足 (5.62)，那它必定是正则变换**。并且反过来常常也对！

5.2.3 无穷小正则变换与诺特定理

现在，假设某个正则变换将相空间点 \mathbf{x} 变到和它邻近的 \mathbf{x}' ，即 \mathbf{x}' 与 \mathbf{x} 的差值为无穷小量，称为无穷小正则变换。为了考察无穷小正则变换，不妨设

$$q^a \rightarrow q'^a = q^a + \varepsilon Q^a(q, p), \quad p_a \rightarrow p'_a = p_a + \varepsilon P_a(q, p). \quad (5.63)$$

式中 ε 为无穷小量， Q^a, P_a 均为相空间的函数（请注意本书符号与其它教材的不同）。我们要解决的问题是，满足什么条件才能让上面这个无穷小映射变成一个正则变换？

上一小节的知识告诉我们，条件就是

$$dp'_a \wedge dq'^a - dp_a \wedge dq^a = 0. \quad (5.64)$$

代入上面的无穷小映射 (5.63)，并展开到一阶无穷小，即有

$$(dP_a \wedge dq^a + dp_a \wedge dQ^a) \varepsilon = 0. \quad (5.65)$$

由此即有

$$\begin{aligned} 0 &= dP_b \wedge dq^b + dp_a \wedge dQ^a \\ &= \frac{\partial P_b}{\partial q^a} dq^a \wedge dq^b + \frac{\partial P_b}{\partial p_a} dp_a \wedge dq^b + \frac{\partial Q^a}{\partial q^b} dp_a \wedge dq^b + \frac{\partial Q^a}{\partial p_b} dp_a \wedge dp_b \\ &= \frac{1}{2} \left(\frac{\partial P_b}{\partial q^a} - \frac{\partial P_a}{\partial q^b} \right) dq^a \wedge dq^b + \frac{1}{2} \left(\frac{\partial Q^a}{\partial p_b} - \frac{\partial Q^b}{\partial p_a} \right) dp_a \wedge dp_b \\ &\quad + \left(\frac{\partial P_b}{\partial p_a} + \frac{\partial Q^a}{\partial q^b} \right) dp_a \wedge dq^b. \end{aligned} \quad (5.66)$$

由此即可得到 Q^a, P_a 必须满足的方程

$$\begin{aligned} \frac{\partial P_b}{\partial q^a} - \frac{\partial P_a}{\partial q^b} &= 0 \\ \frac{\partial Q^a}{\partial p_b} - \frac{\partial Q^b}{\partial p_a} &= 0 \\ \frac{\partial P_b}{\partial p_a} + \frac{\partial Q^a}{\partial q^b} &= 0. \end{aligned} \quad (5.67)$$

这三个方程看起来复杂，实际上它们的通解非常简单，不难验证，下面为它们的通解

$$Q^a = \frac{\partial G}{\partial p_a}, \quad P_a = -\frac{\partial G}{\partial q^a}. \quad (5.68)$$

式中 G 为某个任意的相空间函数。即是说，如果无穷小映射 (5.63) 具有 (5.68) 的形式，那它就是一个无穷小正则变换，称之为由物理量 $G(q, p)$ 生成的无穷小正则变换。

举一个简单的例子，假设取 $G = p_c$ ，则由 (5.68) 式可以得到， $Q^a = \delta_c^a$ ， $P_a = 0$ 。从而 G 生成的无穷小正则变换为

$$q^a \rightarrow q^a + \varepsilon \delta_c^a, \quad p_a \rightarrow p_a. \quad (5.69)$$

即是说，正则动量 p_c 可以生成正则坐标 q^c 的无穷小平移。

人们常常记无穷小正则变换前后正则坐标的改变量为 $\delta q^a = q'^a - q^a$ ，记正则动量的改变量为 $\delta p_a = p'_a - p_a$ ，则根据 (5.63) 式和 (5.68) 式，我们有

$$\delta q^a = \varepsilon \frac{\partial G}{\partial p_a}, \quad \delta p_a = -\varepsilon \frac{\partial G}{\partial q^a}. \quad (5.70)$$

按照这一节一直在使用的关于正则变换的第二种观点，无穷小正则变换前后，物理量的函数形式也必然发生了无穷小改变，通常定义 $A(q, p)$ 的无穷小改变为

$$\delta A = A(q', p') - A(q, p) = A(q', p') - A(q, p). \quad (5.71)$$

代入 (5.70) 式，可以算得这个无穷小改变量为

$$\begin{aligned} \delta A &= \frac{\partial A}{\partial q^a} \delta q^a + \frac{\partial A}{\partial p_a} \delta p_a \\ &= \varepsilon \left(\frac{\partial A}{\partial q^a} \frac{\partial G}{\partial p_a} - \frac{\partial A}{\partial p_a} \frac{\partial G}{\partial q^a} \right) \\ &= \varepsilon [A, G]. \end{aligned} \quad (5.72)$$

以上的讨论促使我们进一步考察一个正则变换的单参簇，

$$q^a \rightarrow q'^a(q, p, \lambda), \quad p_a \rightarrow p'_a(q, p, \lambda), \quad (5.73)$$

λ 为这一簇正则变换的参数，满足 $q'^a(q, p, \lambda = 0) = q^a$ ， $p'_a(q, p, \lambda = 0) = p_a$ 。显然，随着参数 λ 的连续变动，这簇正则变换会在相空间中画出一条以 λ 为参数的路径，路径上任何两个相差无穷小量 $d\lambda$ 的点都是用无穷小正则变换连接起来的。因此，根据上面关于无穷小正则变换的 (5.70) 式，我们容易得到

$$\frac{dq'^a}{d\lambda} = \frac{\partial G}{\partial p'_a}, \quad \frac{dp'_a}{d\lambda} = -\frac{\partial G}{\partial q'^a}, \quad (5.74)$$

式中 $G(q', p', \lambda)$ 称为这个正则变换单参簇的**生成元**！习惯上，人们会省略变量上的撇号，将上面的方程写成

$$\frac{dq^a}{d\lambda} = \frac{\partial G}{\partial p_a}, \quad \frac{dp_a}{d\lambda} = -\frac{\partial G}{\partial q^a}. \quad (5.75)$$

人们常常将 $(\frac{dq^a}{d\lambda}, \frac{dp_a}{d\lambda})$ 想像成某种在相空间中随参数 λ 而变化的流动，由于 G 是整个相空间上的函数，所以这种流动在相空间的每一点都有定义，因此 $(\frac{dq^a}{d\lambda}, \frac{dp_a}{d\lambda})$ 就形成相空间的

一个流速场。进而，上面的方程 (5.75) 就称为由生成元 G 生成的**相流**。如果我们使用更抽象的相空间坐标 $\mathbf{x} = (q, p)$ ，那就可以将上面的相流方程 (5.75) 重写成

$$\frac{dx^i}{d\lambda} = \omega^{ij} \partial_j G. \quad (5.76)$$

$\frac{dx^i}{d\lambda}$ 就是 G 生成的相流的流速场，可以记作 v_G^i ，即 $v_G^i = \frac{dx^i}{d\lambda} = \omega^{ij} \partial_j G$ 。

特别的，我们可以取参数 λ 为时间 t ，取生成元 G 为哈密顿量 H ，那么上面的相流方程 (5.75) 就变成了标准的哈密顿正则方程。正因为如此，人们常常称哈密顿量为时间演化的生成元。由于哈密顿量描述的就是系统能量，因此这就给出了物理系统能量的一般定义，即，**时间演化的生成元描述系统的能量**。通过这一段的论述我们也知道了，系统在相空间的时间演化过程实际上是一个持续进行的正则变换，是以时间为参数的正则变换单参簇。因此，**时间演化过程保持相空间的辛结构**。

进一步，假设我们有两个正则变换的单参簇，分别由 G_1 和 G_2 生成，相应的参数分别记为 λ_1, λ_2 ，则利用相流方程以及 $\omega_{ij} \omega^{jm} = \delta_n^m$ ，可以算得

$$\omega_{ij} v_{G_1}^i v_{G_2}^j = \omega_{ij} \frac{dx^i}{d\lambda_1} \frac{dx^j}{d\lambda_2} = (\partial_m G_2) \omega^{mn} (\partial_n G_1) = [G_2, G_1]. \quad (5.77)$$

这个结果本章用不到，但是下一章会用到。数学家们喜欢记左边的 $\omega_{ij} v_{G_1}^i v_{G_2}^j = \omega(v_{G_1}, v_{G_2})$ ，称为辛形式在速度场 v_{G_1}, v_{G_2} 上的值。这样一来，上面的结果就可以重写为

$$\omega(v_{G_1}, v_{G_2}) = [G_2, G_1]. \quad (5.78)$$

它告诉我们，辛形式在两个速度场上的值，可以由生成这些速度场的物理量的泊松括号算出。

诺特定理

原则上正则变换可以是任何保持辛结构的数学变换。不过，对一个物理系统进行的物理操作（比如将系统旋转一个角度）常常在相空间诱导出一个正则变换。这是因为，对系统的物理操作会将一个物理状态变到另一个物理状态，从而在相空间诱导出一个微分同胚映射，而且由于物理操作之后系统依然是一个哈密顿正则系统，满足同样的哈密顿正则方程，所以这一诱导出来的微分同胚映射必定是一个正则变换。特别的，如果物理操作本身依赖于一个连续参数（比如旋转操作中的旋转角度），那它就会在相空间诱导出一个正则变换的单参簇，这时候我们就称这个单参簇的生成元为相应物理操作的生成元。比如，前面我们已经看到，**正则动量就是对相应正则坐标的平移操作的生成元**。

再比如说，对于一个在 $x-y$ 平面上运动的粒子，我们可以对系统进行绕原点的旋转操作。实际上，这一旋转操作的生成元就是角动量 $J = xp_y - yp_x$ 。这是因为， J 可以生成如下无穷小正则变换

$$\begin{aligned} \delta x &= \varepsilon[x, J] = \varepsilon[x, xp_y - yp_x] = -\varepsilon y, \\ \delta y &= \varepsilon[y, J] = \varepsilon[y, xp_y - yp_x] = \varepsilon x, \\ \delta p_x &= \varepsilon[p_x, J] = \varepsilon[p_x, xp_y - yp_x] = -\varepsilon p_y, \\ \delta p_y &= \varepsilon[p_y, J] = \varepsilon[p_y, xp_y - yp_x] = \varepsilon p_x. \end{aligned} \quad (5.79)$$

不难看出,这正是绕原点的无穷小旋转, ε 为旋转角度。所以, **角动量是旋转操作的生成元**。

以上所谈是任意的连续物理操作。但是,如果这一连续操作本身是一个对称操作,即在操作前后系统的运动微分方程保持形式不变。那就会有更进一步的结论,也就是诺特定理,它说,这时候这一**对称操作的生成元必定是守恒量**。证明非常简单,对称操作由于要保持运动微分方程(由哈密顿正则方程给出)的形式不变,因此必定要保持哈密顿量的函数表达式 $H(q, p)$ 不变³, 设对称操作的生成元为 G , 从而即有

$$0 = \delta H = \varepsilon [H, G] \Leftrightarrow [H, G] = 0. \quad (5.80)$$

进而根据物理量 G 的时间演化方程,有

$$\frac{dG}{dt} = [G, H] = 0. \quad (5.81)$$

因此, G 必定是守恒量。诺特定理得证。

在前面第三章中,我们在最小作用量原理的框架中证明过诺特定理。以上我们又在相空间中,在哈密顿正则系统的框架中再次证明了诺特定理。不同的理论框架能够让我们看清诺特定理的不同侧面,从而增加对这一深刻结论的理解。

不妨举一个简单的例子。考察一个在 $x-y$ 平面上运动的粒子,假设系统处在一个中心力场中,势能为 $V(r) = V(\sqrt{x^2 + y^2})$, 从而系统的哈密顿量为

$$H = \frac{1}{2m}(p_x^2 + p_y^2) + V(\sqrt{x^2 + y^2}). \quad (5.82)$$

很显然,这个哈密顿量是旋转不变的,因此根据诺特定理,必定有旋转操作的生成元 $J = xp_y - yp_x$ 守恒,即角动量守恒。

再比方说,在上一节关于涡旋运动的讨论中,我们曾经得出,系统的总动量 P_x, P_y , 以及总角动量 L , 这三者均守恒。现在我们看到,原因就在于, P_x, P_y 是平移操作的生成元, L 是旋转操作的生成元,而涡旋系统的哈密顿量既平移不变,又旋转不变。所以, P_x, P_y, L 的守恒其实是诺特定理的必然结论。

5.2.4 刘维尔定理和庞加莱回归定理

刘维尔定理

前面说过,物理系统在相空间的时间演化过程可以看成一种相空间流动,记相空间坐标为 \mathbf{x} , 那么这种流动的流速 $\dot{\mathbf{x}}$ 由哈密顿正则方程 $\dot{\mathbf{x}} = [\mathbf{x}, H]$ 给出。而且这一流动过程可以看成是一种持续进行的正则变换。进一步,我们可以考察所有可能初始状态的演化过程,由于每一个初始状态对应一个相空间点,因此所有可能相空间点的“流动”就构成了一种相空间“流体”。刘维尔定理说的就是,这种相空间“流体”是一种不可压缩“流体”。

³这里考察的是不显含时间的对称操作,即其生成元不显含时间。对于显含时间的情形,情况会比较复杂一些,虽然诺特定理依然是成立的。不过,不显含时间的情形推理更为简洁,且更能抓住论证的本质,而显含时间的情况又比较少见,所以我们就作讨论了。

具体来说就是, 假设初始 $t = 0$ 时刻在相空间任取一个区域 D_0 , 假设这个区域随时间演化到 t 时刻变成了区域 D_t , 即 D_0 内的点在 t 时刻“流到”了 D_t 。则刘维尔定理说, 这种流动的体积不可压缩, 即 D_t 的体积等于 D_0 的体积, 记为

$$\text{Vol}(D_t) = \text{Vol}(D_0). \quad (5.83)$$

如图 (5.2) 所示, 区域 D_t 相对于 D_0 可以发生很大的形变, 但是体积不变。

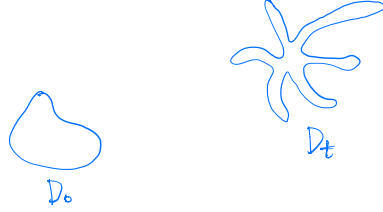


Figure 5.2: $\text{Vol}(D_t) = \text{Vol}(D_0)$.

为了证明刘维尔定理, 我们首先仔细看一下相空间体积元的定义。以一个 $2 \cdot 2 = 4$ 维相空间为例, 相空间体积元 Ω 可以定义为 $\Omega = dp_1 dq^1 dp_2 dq^2$, 用微分形式来写即是 $\Omega = dp_1 \wedge dq^1 \wedge dp_2 \wedge dq^2$ 。而容易验证的是,

$$\begin{aligned} & dp_1 \wedge dq^1 \wedge dp_2 \wedge dq^2 \\ &= \frac{1}{2} (dp_1 \wedge dq^1 + dp_2 \wedge dq^2) \wedge (dp_1 \wedge dq^1 + dp_2 \wedge dq^2) \\ &= \frac{1}{2!} \omega^{\wedge 2}. \end{aligned} \quad (5.84)$$

式中 ω 为辛形式, $\omega^{\wedge 2}$ 定义为 $\omega^{\wedge 2} \equiv \omega \wedge \omega$ 。推广到 $2n$ 维相空间的一般情形, 我们可以定义相空间体积元为 $\Omega = dp_1 dq^1 dp_2 dq^2 \dots dp_n dq^n$, 写成微分形式即是 $\Omega = dp_1 \wedge dq^1 \wedge dp_2 \wedge dq^2 \dots \wedge dp_n \wedge dq^n$ 。同样可以证明, 有

$$\Omega = \frac{1}{n!} \omega^{\wedge n}. \quad (5.85)$$

式中 $\omega^{\wedge n}$ 代表将 ω 用外积 \wedge 自乘 n 次。

(5.85) 式告诉我们, 相空间体积元唯一由辛形式 ω 决定。而辛形式 ω 在正则变换的微分同胚映射下是保持不变的 (即, 正则变换保持辛结构)! 所以, 相空间体积元, 进而任何区域的相空间体积, 都在正则变换的微分同胚映射下保持不变, 简称**相空间体积在正则变换下保持不变**。进一步, 物理系统的时间演化也是一种正则变换, 因此, **相空间体积在时间演化下保持不变**。这就证明了刘维尔定理。

[插入一个注记: 体积元 Ω 在正则变换下保持不变的一个推论是, 正则变换的雅可比行列式 $|\frac{\partial \mathbf{x}'}{\partial \mathbf{x}}|$ 等于 1, 即

$$\left| \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} \right| = 1. \quad (5.86)$$

这是因为，体积元在坐标变换下会多出一个雅可比行列式，如果体积元保持不变，那就说明这个行列式等于 1.]

同样的论证也可以用于相空间的任何 $2k(k \leq n)$ 维曲面 S_{2k} ，其“面积元”定义为 $\frac{1}{k!} \omega^{\wedge k}$ 。我们同样有：随时间演化的 S_{2k} ，其“面积”将保持不变！

特别的，在相空间任取一条闭合回路 C ，让 C 随着时间演化，考察辛势 Θ 沿着这条闭合回路的积分 $I_C = \frac{1}{2\pi} \oint_C \Theta$ ，由于

$$\oint_C \Theta = \int_{S_2} d\Theta = \int_{S_2} \omega. \quad (5.87)$$

式中 S_2 为回路 C 所包围的区域。从上面的推导易知， I_C 将在时间演化下保持不变！

很显然，以上的所有结果不仅对时间演化生成的相流成立，而且对任何正则变换单参簇的相流都成立！

庞加莱回归定理

设想你有一盒气体，初始时气体分子处于盒子的左侧，右侧是真空，中间用挡板隔开，如图 (5.3) 所示。然后你撤去中间的挡板，气体当然就扩散到右边。但是，庞加莱说，只要你等待的时间足够长，那总有一个时刻，气体会自动回归到左侧，如图 (5.3) 所示。你

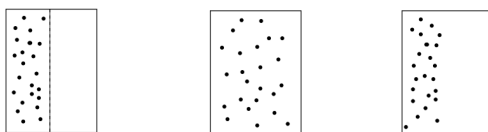


Figure 5.3: 盒中气体的庞加莱回归，如右图。图片来自 David Tong, Classical Mechanics, 以下两幅图也是一样。

说，庞加莱肯定错了，因为这违反了热力学第二定律。但是庞加莱说，热力学第二定律首先要取热力学极限，即取气体分子数目 $N \rightarrow \infty$ ，而庞加莱回归的时间 $t \propto t_c^N$ ($t_c > 1$ 是某个特征量)，因此如果你首先取 $N \rightarrow \infty$ ，那庞加莱回归就看不到了，因为这时候 $t \rightarrow \infty$ 。所以庞加莱回归其实不违法热力学第二定律，对一个宏观系统，它的庞加莱回归时间远比宇宙年龄更长，所以实际上观测不到。

庞加莱回归其实是一个数学定理，是可以证明的！为了证明这个定理，首先要给我们的物理系统加上一个限制条件，即它的实际相空间得是有限的。上面盒子中的气体就满足这个条件，这是因为，首先，盒子限制了气体的空间坐标，使之有限。其次，气体的总能量是守恒的，是一个常数，因此这就限制了气体分子的动量。每一个气体分子的坐标和动量均有限，那所有气体分子在相空间等能量面上的 $6N - 1$ 维相空间当然也有限。

庞加莱回归定理说的是：对于一个相空间有限的哈密顿正则系统，任意取定一个相空间初始点 \mathbf{x}_0 ，则对于它的任意邻域 D_0 ，必定存在一个点 $\mathbf{x}'_0 \in D_0$ ，它将在有限时间内回归 D_0 。简单地说就是，只要等待的时间足够长，那么系统总可以回归到和初态任意接近的状态。

定理的证明如下：取定一个有限的时间步长 τ ，然后演化初始区域 D_0 ，每次演化一个时间步长 τ ，直至无穷。从而就可以得到一系列的区域，记第 k 步演化的区域为 D_k ， $k = 0, 1, 2, \dots, \infty$ 。刘维尔定理告诉我们

$$\text{Vol}(D_k) = \text{Vol}(D_0), \quad \text{for any } k. \quad (5.88)$$

由此可知，必定存在某对整数 k 和 k' (不妨设 $k' > k$)，使得

$$D_k \cap D_{k'} \neq \emptyset, \quad (5.89)$$

这里 \emptyset 表示空集。如图 (5.4) 左图所示。因为否则的话，就说明所有的一系列 D_k 中任何两个都不相交。由于 $\text{Vol}(D_k)$ 有限，那这将意味着 $\text{Vol}(\bigcup_{k=0}^{\infty} D_k) = \infty$ 。但这是不可能的，因为整个相空间有限。

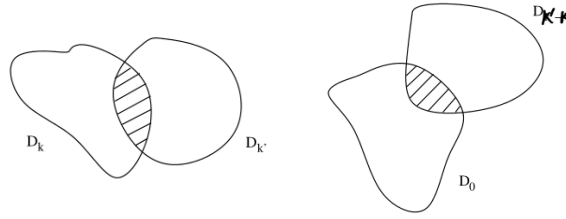


Figure 5.4: 庞加莱回归定理的证明.

记上面的 $D_k \cap D_{k'}$ 为 $D_{k',k}$ ，则

$$D_{k',k} = D_k \cap D_{k'} \neq \emptyset. \quad (5.90)$$

但是时间演化过程所导致的微分同胚映射是可逆的。现在，将上面的结果从第 k' 步开始往回映射 k 步，即有

$$D_{k'-k,0} = D_0 \cap D_{k'-k} \neq \emptyset, \quad (5.91)$$

如图 (5.4) 右图所示。因此， $D_{k'-k,0}$ 中的点在第 $k'-k$ 步回归了！这就完成了庞加莱回归定理的证明。

5.2.5 生成函数与哈密顿正则方程

这一小节的核心目标有两个，一是稍微推广一下前面讨论的正则变换，二是讨论如何构造具体的正则变换。

生成函数

先看如何构造正则变换。前面第 (5.2.2) 小节的 (5.62) 式已经得出，一个变换为正则变换的充分 (常常是充要) 条件是满足下式

$$p_a dq^a - p'_a dq'^a = dF. \quad (5.92)$$

式中 F 为相空间的函数。现在, 假设我们取 F 为某类函数, 记为 $F_1(q, q')$, 即 F 是 q 和 q' 的函数, 这里符号 q 代表所有的 q 坐标, q' 代表所有的 q' 坐标。则 (5.92) 式就给出

$$\begin{aligned} p_a dq^a - p'_a dq'^a &= dF_1 \\ \Rightarrow p_a dq^a - p'_a dq'^a &= \frac{\partial F_1}{\partial q^a} dq^a + \frac{\partial F_1}{\partial q'^a} dq'^a. \end{aligned} \quad (5.93)$$

很显然, 只要取

$$p_a = \frac{\partial F_1}{\partial q^a}(q, q') \quad -p'_a = \frac{\partial F_1}{\partial q'^a}(q, q') \quad (5.94)$$

就能得到一组 (q, p) 和 (q', p') 之间的坐标变换, 由于它自动满足 (5.92) 式, 所以显然是正则变换。即是说, 我们只要选择不同的 $F_1(q, q')$ 函数, 就能根据 (5.94) 式得到各种不同的正则变换, 函数 $F_1(q, q')$ 因此称为正则变换的第一类生成函数。

不妨举一个简单的例子。考虑一个单自由度的系统, 因此相空间为 2 维, 由 q, p 描述。很明显, 下面这个坐标变换是正则变换

$$q' = p, \quad p' = -q. \quad (5.95)$$

不难验证, 这个正则变换可以由第一类生成函数 $F_1 = qq'$ 生成。

下面我们把 (5.92) 式改写成

$$p_a dq^a + q'^a dp'_a = d(F + p'_a q'^a) = dF_2. \quad (5.96)$$

则很容易看出, 只要选 F_2 为 q, p' 的函数, 就能根据下式得到正则变换

$$p_a = \frac{\partial F_2}{\partial q^a}(q, p') \quad q'^a = \frac{\partial F_2}{\partial p'_a}(q, p'). \quad (5.97)$$

$F_2(q, p')$ 因此称为正则变换的第二类生成函数。

很显然, 我们也可以构造第三、第四类生成函数, 如下

$$\begin{aligned} F_3(p, q') : \text{取} \quad q^a &= -\frac{\partial F_3}{\partial p_a}(p, q'), \quad p'_a = -\frac{\partial F_3}{\partial q'^a}(p, q') \\ F_4(p, p') : \text{取} \quad q^a &= -\frac{\partial F_4}{\partial p_a}(p, p'), \quad q'^a = \frac{\partial F_4}{\partial p'_a}(p, p'). \end{aligned} \quad (5.98)$$

值得指出的是, 这四类生成函数并没有穷尽所有的可能性, 比方说, 对于一个 $2 \cdot 2 = 4$ 维相空间, 下面正则变换就不能由以上四类生成函数生成

$$q'^1 = q^1, \quad q'^2 = p_2, \quad p'_1 = p_1, \quad p'_2 = -q^2. \quad (5.99)$$

实际上, 它要由一种混合型 (第一类和第二类混合) 生成函数生成, 具体的生成函数为 $q^1 p'_1 + q^2 q'^2$ 。

正则变换与哈密顿正则方程再讨论

我们知道，正则变换以后，新的正则变量 $\mathbf{x}' = (q', p')$ 也满足哈密顿正则方程。实际上，(5.92) 式可以发展成一种证明这一结论的不同方法。为此，我们将 (5.92) 式凑成如下形式

$$p_a dq^a - H dt = p'_a dq'^a - H dt + dF. \quad (5.100)$$

然后将这个式子两边分别沿着时间积分，从初始的 t_i 积到末尾的 t_f ，即

$$\int_{t_i}^{t_f} [p_a dq^a - H dt] = \int_{t_i}^{t_f} [p'_a dq'^a - H dt] + \int_{t_i}^{t_f} dF, \quad (5.101)$$

可以看到，左边是老正则变量所描述的相空间作用量，右边是新正则变量所描述的相空间作用量加上一个多出来的全微分项。多出来的全微分项当然只在积分的上下限有贡献，也就是只在相空间路径的两个端点上有贡献，可以称之为端点项。如果我们先忽略这个端点项，那很显然，(5.101) 式左边变分等于零就等价于右边变分等于零。左边变分等于零给出的是老正则变量下的哈密顿正则方程，右边变分等于零给出的当然就是新正则变量下的哈密顿正则方程。这样我们的证明就完成了。

但是，端点项得小心处理。因为前面第二章讲相空间最小作用量原理时我们说过，相空间路径端点的 q 坐标是固定的，但是 p 坐标不能固定。这样看来，变分时端点项的变分就不一定是零。这个问题是这样处理的（稍微有些微妙）：首先，注意到任何变分都是对相空间中一条试探路径的无穷小变动。因此，我们可以不限定这些试探路径在端点处的 p 坐标，但与此同时要求对任何试探路径的无穷小变动在路径两端取零，即不仅要求 $\delta q(t_i) = \delta q(t_f) = 0$ ，也要求 $\delta p(t_i) = \delta p(t_f) = 0$ （虽然 $p(t_i), p(t_f)$ 本身不固定）。总之，因为我们用最小作用量原理是这么用的：先选一条试探路径，然后在这条试探路径附近作无穷小变分，看作用量的改变量是否为零，如果是，那开始的试探路径就是真实路径。所以我当然可以要求每一条试探路径端点处的 $\delta p = 0$ ，同时不同的试探路径又可以取不同的 p 端点，也就是 p 端点并不固定。因此，这样处理的话，端点项的变分就是零，从而确实可以忽略，上一段的论证也就没有漏洞了。

到此为止，我们所讨论的主要是不显含时间的正则变换，即正则变量的变换函数关系 $\mathbf{x}'(\mathbf{x})$ 中不显含时间。如果想将我们的讨论应用到显含时间的情形，即应用到 $\mathbf{x}'(\mathbf{x}, t)$ 的情形，那就应该将前面所有的微分表达式理解为瞬时进行的，即**不对时间变量 t 微分**。比如 (5.92) 式中的微分就应该作此理解。不过，对于这种显含 t 的情形就无法证明变换后的系统依然满足同样的哈密顿正则方程了（因为在这种情况下，考虑时间演化时， $\mathbf{x}'(\mathbf{x}, t)$ 要对 t 求偏导）。不过，可以证明，只要不取变换后的哈密顿量为原来的 H ，而是重新选取一个合适的哈密顿量（记为 K ）（它不仅函数表达式与 H 不同，而且是不同的物理量），那变换后的系统也将满足哈密顿正则方程。

证明如下：首先我们引入一个 K ，对 K 的要求是使得下式成立

$$p_a dq^a - H dt = p'_a dq'^a - K dt + dF, \quad (5.102)$$

注意，这里的微分是对所有的变量进行，**包括对 t 进行微分**，而式中的 F 一般来说也显含 t 。换言之，为了使得 (5.102) 式依然满足正则变换的条件 (5.92)，我们应该这样选择 K ，

即使得它刚好抵消 (5.102) 式中关于 t 的偏微分⁴。容易看出, 如果将式中的 F 取成第一类生成函数 $F_1(q, q', t)$, 那 K 即是

$$K = H + \frac{\partial F_1}{\partial t}. \quad (5.103)$$

下面, 从 (5.102) 式出发, 完全类似于上面利用相空间最小作用量原理的证明过程, 可以得到, 变换以后的系统满足如下哈密顿正则方程

$$\dot{q}'^a = \frac{\partial K}{\partial p'_a}, \quad \dot{p}'_a = -\frac{\partial K}{\partial q'^a}. \quad (5.104)$$

即是说, 在这种情况下, K 就是变换以后的哈密顿量。

5.3 哈密顿-雅可比方程

5.3.1 哈密顿-雅可比方程

我们知道, 给定 q 坐标在相空间的起点和终点, 给定连接两点的一条相空间路径 $(q(t'), p(t'))$, 我们就可以计算相空间作用量

$$S[q(t'), p(t')] = \int_0^t [p_a(t') dq^a(t') - H(q(t'), p(t')) dt']. \quad (5.105)$$

这里已经假设 0 时刻为起始时刻, t 为末尾时刻。不妨进一步假设 0 时刻和末尾 t 时刻的 q 坐标分别为 q_0 和 q , 即

$$q^a(0) = q_0^a, \quad q^a(t) = q^a. \quad (5.106)$$

注意, 我们并没有预先给定起末两端的 p 坐标。

现在, 假设我们考察的这条路径是一条满足哈密顿正则方程的路径, 即使作用量取极值的路径。我们计算这条路径的作用量, 很显然, 这个作用量依赖于预先给定的起末坐标 q_0 和 q (起末两端的 p 坐标是自动被哈密顿正则方程决定的)。也即是说, $\text{extrem}_{[q(t'), p(t')]} S[q(t'), p(t')]$ 定义了一个依赖于 t, q_0, q 的函数, 记为 $S(q_0, q, t)$,

$$S(q_0, q, t) = \text{extrem}_{[q(t'), p(t')]} S[q(t'), p(t')]. \quad (5.107)$$

为了研究函数 $S(q_0, q, t)$, 我们假设末端的 q 发生了一个无穷小变动 δq , 极值路径当然会跟着发生一个无穷小改变, 但很显然,

$$\begin{aligned} \delta S(q_0, q, t) &= \text{extrem} \delta S[q(t'), p(t')] \\ &= \int_0^t d(p_a \delta q^a) + \text{extrem} \int_0^t dt' \left[\left(\dot{q}^a - \frac{\partial H}{\partial p_a} \right) \delta p_a - \left(\dot{p}_a + \frac{\partial H}{\partial q^a} \right) \delta q^a \right]. \\ &= p_a(t) \delta q^a(t) = p_a(t) \delta q^a. \end{aligned} \quad (5.108)$$

由此就可以得到,

$$p_a(t) = \frac{\partial S}{\partial q^a}(q_0, q, t). \quad (5.109)$$

⁴注意, 由于 $q' = q'(q, p, t)$, 所以 dq' 中其实也包含对 t 的偏微分。

有了这个关系以后再回顾作用量泛函的表达式 (5.105), 就可以发现,

$$\frac{\partial S}{\partial t}(q_0, q, t) = -H(q, p). \quad (5.110)$$

结合 (5.109) 式和 (5.110) 式, 就可以得到函数 $S(q_0, q, t)$ 所满足的偏微分方程

$$\frac{\partial S}{\partial t} + H(q, \frac{\partial S}{\partial q}, t) = 0. \quad (5.111)$$

这就是**哈密顿-雅可比方程**, 式中我们已经允许了哈密顿量显含时间 t (这对上面的推导没有任何影响)。很显然, 哈密顿-雅可比方程不能完全决定函数 S , 至少我们可以给 S 加上一个任意的常数, 这个任意常数对于讨论并不重要, 所以下面我们都忽略它。

下面将推理逻辑反过来, 假设我们从哈密顿-雅可比方程出发, 我们想求解这个偏微分方程。那上面的推理过程就告诉我们, 如何通过考虑一条起于 q_0 , t 时刻到达 q 的极值路径, 并计算这条路径的作用量泛函, 来构造出哈密顿-雅可比方程的解。在这种构造方式中, 起始点 q_0 可以看成是积分常数。这一构造过程也告诉我们, 如果极值路径 (也称作物理路径, 或经典路径) 可以解析求解 (允许用积分来表达), 那就能得到哈密顿-雅可比方程的形如 $S(q_0, q, t)$ 的解析解。

但是, $S(q_0, q, t)$ 只是哈密顿-雅可比方程的一类解。而且实际上, 我们通常无法求出其解析形式。进一步, 对一个 n 自由度的系统, 如果能找到哈密顿-雅可比方程的一类形如 $S(\alpha, q, t)$ 的解析解 (允许用积分来表达), 其中 α 代表 $\alpha^a (a = 1, 2, \dots, n)$, 为积分常数 (类比于前面的 q_0)。那这样的系统就称作**可积系统**。后面我们将证明, 这 n 个积分常数 $\alpha^a (a = 1, 2, \dots, n)$ 其实是系统的 n 个泊松对易的独立守恒量。反过来, 如果一个哈密顿力学系统有 n 个泊松对易的独立守恒量, 那我们就能用这些守恒量构造出哈密顿-雅可比方程的一类形如 $S(\alpha, q, t)$ 的解析解。所以, 存在 n 个泊松对易的独立守恒量是一个哈密顿力学系统成为一个可积系统的充要条件, 下一章我们会进一步讨论这个问题。

一般地, 假设找到了哈密顿-雅可比方程的一个解 $S(q, t)$, 则根据 (5.109) 式, 我们就得到粒子正则动量与正则坐标的函数关系, 将之代入哈密顿正则方程的第一个方程 (如下), 就能解出粒子在位形空间的运动路径,

$$\dot{q}^a = \frac{\partial H}{\partial p_a} \Big|_{p=\frac{\partial S}{\partial q}}. \quad (5.112)$$

显然, 这种求解运动路径的方法最后只需求解 n 个一阶常微分方程, 而不是求解 n 个二阶常微分方程, 这当然大大简化了问题。然而 (5.112) 解出来的路径真是粒子的经典路径吗? 为了证明这一点, 我们需要进一步验证哈密顿正则方程的第二个方程也是满足的。证明如下, 首先

$$\dot{p}_a = \frac{d}{dt} \left(\frac{\partial S}{\partial q^a} \right) = \frac{\partial^2 S}{\partial t \partial q^a} + \frac{\partial^2 S}{\partial q^b \partial q^a} \dot{q}^b. \quad (5.113)$$

其次, 将哈密顿雅可比方程 (5.111) 对 q 求偏导, 有

$$\begin{aligned} & \frac{\partial^2 S}{\partial t \partial q^a} + \frac{\partial H}{\partial q^a} \Big|_{p=\frac{\partial S}{\partial q}} + \frac{\partial H}{\partial p_b} \Big|_{p=\frac{\partial S}{\partial q}} \frac{\partial^2 S}{\partial q^b \partial q^a} = 0 \\ \Rightarrow & \frac{\partial^2 S}{\partial t \partial q^a} + \frac{\partial H}{\partial q^a} \Big|_{p=\frac{\partial S}{\partial q}} + \dot{q}^b \frac{\partial^2 S}{\partial q^b \partial q^a} = 0. \\ \Rightarrow & \frac{\partial^2 S}{\partial t \partial q^a} + \dot{q}^b \frac{\partial^2 S}{\partial q^b \partial q^a} = - \frac{\partial H}{\partial q^a} \Big|_{p=\frac{\partial S}{\partial q}}. \end{aligned} \quad (5.114)$$

将最后的结果与上面那个式子比较, 就可以得到, $\dot{p}_a = -\frac{\partial H}{\partial q^a}$, 证明因此就完成了。

假设哈密顿量不显含时间, 则有能量守恒, 记这个守恒量为 E , 哈密顿-雅可比方程当然有依赖于这个守恒量的解, 定义

$$S(E, q, t) = W(E, q) - Et. \quad (5.115)$$

将这个定义代入哈密顿雅可比方程, 就可以得到

$$H(q, \frac{\partial W}{\partial q}) = E. \quad (5.116)$$

我们只需求解这个偏微分方程, 就能得到原来的哈密顿-雅可比方程的一类解, 这类解的特点是, 利用它反过来求解出的粒子运动路径都有同样的能量 E 。 $W(E, q)$ 也常常简记为 $W(q)$, 人们有时候称之为哈密顿主函数。

5.3.2 可积系统

上一小节我们说过, 如果系统有一个守恒量 α , 相应的哈密顿-雅可比方程就有一类依赖于这个守恒量的解, $S(\alpha, q, t)$, 从偏微分方程求解的角度来看, α 就是这类解的积分常数。特别的, 对于 n 自由度的可积系统, 人们能找到 n 个相互泊松对易 (待会儿就能看到泊松对易的要求怎么来的) 的独立守恒量 $\alpha^a, a = 1, 2, \dots, n$ 。相应的也能找到哈密顿-雅可比方程的具有 n 个积分常数的解析解, 记为 $S(\alpha, q, t)$ (这里 α 代表 $\alpha^a, a = 1, 2, \dots, n$)。

可积系统之所以称作可积系统, 是因为可以进一步得出粒子运动路径的解析解 (用积分和隐函数表达出来)。下面我们来证明这个结论。

假设找到了哈密顿-雅可比方程的一类含 n 个积分常数的解析解, 记为 $S(\alpha, q, t)$, 写得更清楚一点就是 $S(\alpha^1, \dots, \alpha^n, q^1, \dots, q^n, t)$ 。我们可以将 $\alpha^a, a = 1, 2, \dots, n$ 看作是正则变换以后新的 q 坐标, 进而考虑 (q, p) 到 (α, β) 的正则变换, 其中 β 代表 $\beta_a, a = 1, 2, \dots, n$ 为正则变换以后新的正则动量。我们将哈密顿-雅可比方程的解 $S(\alpha, q, t)$ 取作正则变换的第一类生成函数。很明显, 这是一种显含时间的正则变换, 变换以后的哈密顿量 K 为

$$K = H + \frac{\partial S}{\partial t}. \quad (5.117)$$

根据哈密顿雅可比方程, 我们有

$$K = 0. \quad (5.118)$$

进而由变换以后的哈密顿正则方程, 可以知道

$$\dot{\alpha}^a = 0, \quad \dot{\beta}_a = 0. \quad (5.119)$$

即变换以后的正则坐标 α^a 的确是守恒量, 而且因为它们都是正则坐标, 所以泊松对易! 变换以后的正则动量 β_a 也是守恒量, 不过, 由于正则动量和正则坐标的泊松括号不是零, 所以它们与 α^a 并不泊松对易, 实际上, 对它们的更好解释是, 将之看成初始条件。

由第一类生成函数的知识可以知道

$$\beta_a = -\frac{\partial S}{\partial \alpha^a}. \quad (5.120)$$

解这组式子就能得到位形空间的运动路径 $q^a(t)$ (作为 α, β 的函数)。进一步, 无论是由前面的 (5.109) 式, 还是由第一类生成函数的知识, 我们都有

$$p_a = \frac{\partial S}{\partial q^a}. \quad (5.121)$$

这一组方程就进一步决定了函数 $p_a(t)$ 。可见, 可积系统粒子运动路径的解析解是可以得到的。

5.3.3 举例：哈密顿雅可比方程求解中心力场问题

中心力场问题其实就是一个可积系统。因为这个系统的自由度数为 3, 而它也存在 3 个相互泊松对易的守恒量, 即 H, \mathbf{J}^2, J_z , 这里 \mathbf{J} 是系统的角动量, J_z 是它的 z 分量。 \mathbf{J}^2, J_z 这两者与 H 泊松对易是因为系统具有旋转对称性, 而 $[\mathbf{J}^2, J_z] = 0$ 我们前面在讲泊松括号的时候提到过。下面我们来看如何从哈密顿-雅可比方程得到中心力场问题的解析解。

首先, 由于角动量守恒, 粒子必定作平面运动, 进而可以将自由度数降低为 2。我们取粒子的运动平面为 $x-y$ 平面, 因此可以写出系统的哈密顿量

$$H = \frac{1}{2m}(p_x^2 + p_y^2) + V(r). \quad (5.122)$$

其次, 在 $x-y$ 平面上取极坐标 $x = r \cos \theta$, $y = r \sin \theta$ 。利用辛势在两种坐标系中的不同表达, 我们有

$$p_x dx + p_y dy = p_r dr + p_\theta d\theta. \quad (5.123)$$

由此即得 $p_r = p_x \cos \theta + p_y \sin \theta$, $p_\theta / r = -p_x \sin \theta + p_y \cos \theta$ 。进而可以得到

$$p_x^2 + p_y^2 = p_r^2 + \frac{p_\theta^2}{r^2}. \quad (5.124)$$

从而可以将哈密顿量在极坐标中写成

$$H = \frac{1}{2m}\left(p_r^2 + \frac{p_\theta^2}{r^2}\right) + V(r). \quad (5.125)$$

当然, 人们也可以通过先写出极坐标中的拉格朗日量, 再进行勒让德变换的方式得到上面的式子。

由于能量守恒, 我们可以设

$$S(q, t) = W(r, \theta) - Et. \quad (5.126)$$

进而由 (5.116) 式, 即有

$$\begin{aligned} \frac{1}{2m} \left[\left(\frac{\partial W}{\partial r} \right)^2 + \frac{1}{r^2} \left(\frac{\partial W}{\partial \theta} \right)^2 \right] + V(r) &= E. \\ \Leftrightarrow r^2 \left[2m[E - V(r)] - \left(\frac{\partial W}{\partial r} \right)^2 \right] &= \left(\frac{\partial W}{\partial \theta} \right)^2. \end{aligned} \quad (5.127)$$

很明显, 我们可以用分离变量的方式解上面的方程, 即设

$$W(r, \theta) = W_1(r) + W_2(\theta). \quad (5.128)$$

代入 (5.127) 即有

$$r^2 \left[2m[E - V(r)] - \left(\frac{\partial W_1}{\partial r} \right)^2 \right] = \left(\frac{\partial W_2}{\partial \theta} \right)^2, \quad (5.129)$$

方程的左边只依赖于 r , 而方程的右边只依赖于 θ , 左右要相等, 那只能是都等于一个常数 (由于右边是完全平方, 所以是一个大于零的常数), 设这个常数为 J^2 , 从而有

$$\begin{aligned} \frac{dW_1}{dr} &= \sqrt{2m[E - V(r)] - \frac{J^2}{r^2}} \\ \frac{dW_2}{d\theta} &= J. \end{aligned} \quad (5.130)$$

将这个式子积分, 即有

$$\begin{aligned} W_1(r) &= \int dr \sqrt{2m[E - V(r)] - \frac{J^2}{r^2}} \\ W_2(\theta) &= J\theta. \end{aligned} \quad (5.131)$$

因此即有

$$S = \int dr \sqrt{2m[E - V(r)] - \frac{J^2}{r^2}} + J\theta - Et. \quad (5.132)$$

式中我们已经忽略了不重要的相加常数。

从 (5.130) 的第二个式子可以清楚地看到, $p_\theta = \frac{\partial S}{\partial \theta} = \frac{\partial W_2}{\partial \theta} = J$ 为守恒量, 这就是角动量, 另一个守恒量就是能量 E 。它们也是哈密顿-雅可比方程的两个积分常数, 不妨记 $\alpha^1 = E, \alpha^2 = J$ 。则根据上一小节的理论, 我们还有下面两个运动积分

$$\begin{aligned} \beta_1 &= -\frac{\partial S}{\partial E} = t - \int dr \frac{m}{\sqrt{2m[E - V(r)] - \frac{J^2}{r^2}}}, \\ \beta_2 &= -\frac{\partial S}{\partial J} = \int dr \frac{J/r^2}{\sqrt{2m[E - V(r)] - \frac{J^2}{r^2}}} - \theta. \end{aligned} \quad (5.133)$$

实际上, 人们通常记 $\beta_1 = t_0$ 为初始时刻, 记 $-\beta_2 = \theta_0$ 为初始角度。这样, 我们就完全解出了中心力场问题。人们可以将这里的结果和第四章中的相应结果进行比较, 两者是完全一致的。

6. 哈密顿系统中的规则与混沌

前面我们讲过对称性与守恒定律的关系，也说明过对称性如何决定系统的作用量。其实，对称性还强烈地限制着哈密顿力学系统的动力学行为。一个系统的对称性越高，其动力学行为就越呈现出规则性，而对称性的降低就可能導致混沌。相反，一个系统的行为越表现出规则性往往就意味着它有越高的对称性，不过这个对称性可能是隐藏的。

另一方面，连续对称性总是对应着守恒量，也就是与哈密顿量泊松对易的物理量¹。因此，如果一个力学系统有越多的守恒量，其动力学行为就会越规则，而守恒量的不足也是力学系统呈现混沌行为的必要条件。本章我们将从最规则的力学系统（也就是所谓的可积系统）出发，逐步探讨哈密顿力学系统如何从规则过渡到混沌。

6.1 最规则的哈密顿系统-可积系统

动力学行为最规则的哈密顿系统是所谓的可积系统，也称作刘维尔可积系统。这个概念上一章已经引入了，假设我们考察的是一个 n 自由度的哈密顿系统，那么可积的定义是，存在 n 个泊松对易的独立守恒量，不妨记这些守恒量为 $G_a, a = 1, 2, 3, \dots, n$ ，它们满足 $[H, G_a] = 0$ 。所谓泊松对易即满足

$$[G_a, G_b] = 0, \quad (6.1)$$

而所谓独立即满足

$$dG_1 \wedge dG_2 \wedge \dots \wedge dG_n \neq 0. \quad (6.2)$$

¹假设不显含时间。

对于 $2n$ 维相空间, 不可能有超过 n 个相互独立且泊松对易的守恒量。由于哈密顿量 H 本身也是守恒量², 而且它和所有 G_a 均泊松对易, 因此这意味着 H 不是独立的, 而必然是 G_a 的函数。实际上, 后面将看到, H 完全由这些 G_a 决定, 不再依赖于更多的独立变量。

6.1.1 不变环面

由于 G_a 是守恒量, 因此系统在相空间的运动轨道一定会被限制在 $G_a, a = 1, 2, \dots, n$ 取常数值 n 维超曲面上, 不同的取值就将相空间分层成不同的超曲面, 不妨记 $G_a = g_a, a = 1, 2, \dots, n$ 的那一层超曲面为 \mathcal{M}_g 。可以证明, 如果 \mathcal{M}_g 连通且有限 (用数学术语来说即紧致), 那它一定可以参数化为 n 维环面 $\mathbf{T}^n = S^1 \times S^1 \times \dots \times S^1$, 即 $\mathcal{M}_g = \mathbf{T}^n$, 称作相空间的不变环面, 不变的意思是, 系统始终在这个环面上运动。图 (6.1) 中所示的就是一个二维环面 \mathbf{T}^2 。

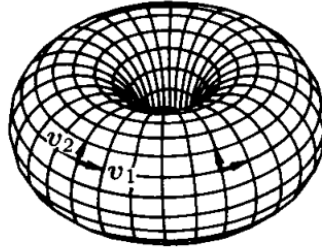


Figure 6.1: 相空间的不变环面。图片来自 Tom W.B. Kibble and Frank H. Berkshire, Classical Mechanics(下图同)。

为了证明上一段的这个结论, 我们引入相空间坐标 $\mathbf{x} = (q, p)$, 并以 $x^i, i = 1, 2, \dots, 2n$ 来表示它的各分量。很显然, 等 G_a 面 \mathcal{M}_g 的法向就是 G_a 的梯度方向, 即有 $\frac{\partial G_a}{\partial \mathbf{x}}, a = 1, 2, \dots, n$ 均垂直于 \mathcal{M}_g 。另一方面, 我们取 $G_a, a = 1, 2, \dots, n$ 为生成元, 让它们在相空间生成相流, 并记相应的相流参数为 λ^a , 即有相流方程

$$\frac{dx^i}{d\lambda^a} = v_{G_a}^i = \omega^{ij} \partial_j G_a, \quad (6.3)$$

式中 $v_{G_a}^i$ 表示相应相流的速度场。很显然, 总共有 n 个这样的相流, 并且由于 $G_a, a = 1, 2, \dots, n$ 相互独立, 所以这些相流的速度场是线性独立的。另外,

$$v_{G_b}^i \partial_i G_a = (\partial_i G_a) \omega^{ij} (\partial_j G_b) = [G_a, G_b] = 0, \quad (6.4)$$

即这些相流的流动方向均与 G_a 的梯度方向相垂直! 从而也即是说, 这些相流均是沿着曲面 \mathcal{M}_g 流动的。当这 n 个线性独立相流流遍 n 维超曲面 \mathcal{M}_g 时, 它们的参数 $\lambda^a, a = 1, 2, \dots, n$ 就给出了 \mathcal{M}_g 的一种坐标。而这些相流是可以一直持续下去的, 它们并没有边界。因此要么这组 n 维参数 $(\lambda^1, \lambda^2, \dots, \lambda^n)$ 的取值范围都是 $(-\infty, +\infty)$, 从而相当于一个 n 维空间 \mathbb{R}^n ,

²我们仅考虑 H 不显含 t 的情形。

要么,如果每个参数的取值都有限的话,那每一个参数的取值范围都只能是在一个圆周 S^1 上,从而相当于一个 n 维环面 \mathbf{T}^n ,而这正是我们想要证明的结论。比如图 (6.1) 中所画的两维环面 \mathbf{T}^2 上的坐标网就是由相流参数组 (λ^1, λ^2) 所参数化的。如此就完成了整个证明。

\mathbf{T}^n 由 n 个圆周作笛卡尔积而成,每一个圆周对应一个拓扑独立的非平凡回路,即不可以通过连续形变相互过渡也不可以连续收缩为一点的回路,因此 \mathbf{T}^n 上有 n 个拓扑独立的非平凡回路,不妨记为 $C_a, a = 1, 2, \dots, n$ 。假设 $\mathcal{M}_g = \mathbf{T}^n$, 则系统就在这个 n 维环面上运动,它在每一个独立回路 C_a 方向都是周期性的,但是整个完整的相轨道却不一定是闭合轨道,因为这里有两种不同的可能性。为了弄清楚问题,我们记系统在 C_a 方向的运动周期为 T_a , 从而它在这个方向的角频率为 $\omega_a = \frac{2\pi}{T_a}$ 。仔细考虑一下不难明白,如果任何两个方向的角频率之比 ω_a/ω_b 都是整数比,则各个方向的运动一定有公共的周期,从而整个系统一定作周期性的运动,从而一定是闭合轨道,图 (6.2) 中左图所画的就是这种情形。但是,如果这些角频率之比都是无理数,那就没有公共的周期,也没有闭合轨道,实际上,这时候系统的相轨道最终将密密布满整个 \mathbf{T}^n ,称作相轨道遍历整个不变环面,图 (6.2) 中右图所画的就是这种情形。一般来说,对于具有角频率 $\vec{\omega} = (\omega_1, \omega_2, \dots, \omega_n)$ 的不变

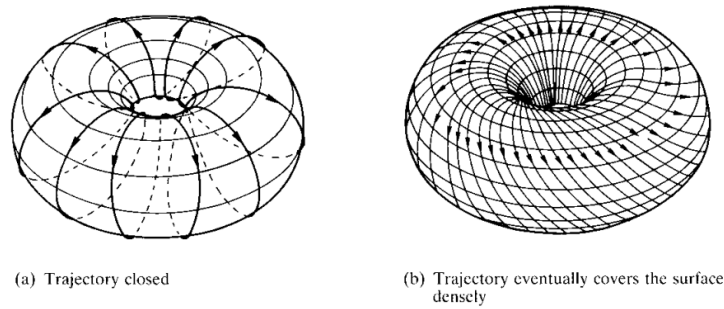


Figure 6.2: 不变环面上的相轨道。

环面,它不能被相应相轨道遍历的充要条件是,存在一组整数 $\vec{m} = (m_1, m_2, \dots, m_n) \neq 0$, 使得 $\vec{\omega} \cdot \vec{m} = m_1 \omega_1 + m_2 \omega_2 + \dots + m_n \omega_n = 0$ 。这个条件称之为共振条件,满足共振条件的不变环面就是所谓的共振环面。

6.1.2 另一个刘维尔定理

刘维尔定理: 可积系统的相空间运动可以通过积分法求解。

类似的结论上一章我们其实证明过,但那个证明假定了能找到哈密顿-雅可比方程的一组解析解。下面要给出的证明完全不依赖于这个假定,相反,我们可以用证明得出的哈密顿正则方程的解析解构造出相应哈密顿-雅可比方程的解析解,如何构造已经在上一章讲过了。

下面我们来证明这个刘维尔定理。证明的关键在于找到一组正则变换,将相空间坐标由原来的 $\mathbf{x} = (q, p)$ 变换为 $\mathbf{x}' = (\psi, G)$, 式中 G 代表所有相互泊松对易的守恒量 G_a , 它们是新的正则动量, ψ 代表新的正则坐标 $\psi^a, a = 1, 2, \dots, n$ 。如果我们成功地做到这一步,那这首先证明了哈密顿量 H 只能依赖于守恒量 G , 不能依赖于 ψ , 这是因为 $0 = [H, G_a]_{\mathbf{x}} =$

$[H, G_a]_{\mathbf{x}'} = \frac{\partial H}{\partial \psi^a}$, 即 $\frac{\partial H}{\partial \psi^a} = 0$ 。其次, 如果我们成功找到这样的正则变换, 则

$$\dot{G}_a = [G_a, H] = 0, \quad \dot{\psi}^a = [\psi^a, H] = \frac{\partial H}{\partial G_a} = \Omega_a, \quad (6.5)$$

式中 Ω_a 定义为 $\frac{\partial H}{\partial G_a}$, 是一些仅仅依赖于 G 的量。从而我们就可以得到系统的运动在这组新正则变量下的解

$$G_a(t) = G_a(0), \quad \psi^a(t) = \psi^a(0) + t\Omega_a. \quad (6.6)$$

将这组解变换回去, 就能得到原来正则变量 (q, p) 下的解。如果这组正则变换可以用某个积分表达式来生成, 那我们就证明了能通过积分法得到系统在原变量 (q, p) 下的解析解。这样就完成了想要的证明。

为了构造这组正则变换, 我们注意到系统的运动总是限制在某层超曲面 \mathcal{M}_g 上, 而 \mathcal{M}_g 由 $G_a(q, p) = g_a$ 确定, 因此我们可以反解出 p_a 作为 g, q 的函数, 记为 $p_a(g, q)$ 。下面在 \mathcal{M}_g 上任意选定一个参考点 q_0 , 并在 \mathcal{M}_g 上考察一条从 q_0 到 q 的路径, 然后将 $p_a(g, q)dq^a$ 沿着这条路径积分³。很明显, 如果积分的结果与具体的路径无关, 而只与起末两点位置有关, 那这个积分就定义了一个函数 $F(G, q)$,

$$F(G, q) = \int_{q_0}^q p_a(g, q)dq^a. \quad (6.7)$$

很显然, 这个函数满足

$$p_a = \frac{\partial F}{\partial q^a}, \quad (6.8)$$

因此我们将之取为正则变换的第二类生成函数, 进而将 G_a 看作新的正则动量, 并定义新的正则坐标为,

$$\psi^a = \frac{\partial F}{\partial G_a}. \quad (6.9)$$

这样, 所需的正则变换就构造出来了。唯一还需要证明的就是 (6.7) 式的积分的确与具体路径无关!

要证明积分和路径无关, 只需证明在 \mathcal{M}_g 的闭合回路⁴上的积分等于零 (请读者自己思考一下为什么是这样?), 即证明 $\oint p_a dq^a|_{\mathcal{M}_g} = 0$ 。但是 $p_a dq^a$ 就是辛势 Θ , 因此就是要证明 $\oint \Theta|_{\mathcal{M}_g} = 0$ 。根据斯托克斯公式, 这就是要证明

$$\oint \Theta|_{\mathcal{M}_g} = \int d\Theta|_{\mathcal{M}_g} = \int \omega|_{\mathcal{M}_g} = 0. \quad (6.10)$$

³本章同样使用求和约定。

⁴假设它可以连续收缩成到 q_0 点

为此只需证明, 辛形式限制在 \mathcal{M}_g 上等于零, 即 $\omega|_{\mathcal{M}_g} = 0$ 。这个证明如下,

$$\begin{aligned}
 \omega|_{\mathcal{M}_g} &= \frac{1}{2} \omega_{ij} \frac{dx^i}{d\lambda^a} \frac{dx^j}{d\lambda^b} d\lambda^a \wedge d\lambda^b \\
 &= \frac{1}{2} \omega_{ij} v_{G_a}^i v_{G_b}^j d\lambda^a \wedge d\lambda^b \\
 &= \frac{1}{2} \omega(v_{G_a}, v_{G_b}) d\lambda^a \wedge d\lambda^b \\
 &= \frac{1}{2} [G_b, G_a] d\lambda^a \wedge d\lambda^b = 0.
 \end{aligned} \tag{6.11}$$

上面的推导过程使用了上一章讲相流时推导的相关公式。到此为止就完成了整个刘维尔定理的证明。

6.1.3 作用-角变量

如果可积系统的运动有限, 从而其 $\mathcal{M}_g = \mathbf{T}^n$, 那就可以选择一组非常有用的正则变量, 称之为作用-角变量。其中作用变量通常记为 $I_a, a = 1, 2, \dots, n$, 是一组互相泊松对易的守恒量, 它和原来的守恒量 G 互为函数关系。 I_a 常常作为正则动量, 与之对应的正则坐标记为 θ^a , 称作角变量。可积系统的哈密顿量只是作用变量 I 的函数, 记为 $H(I)$ 。

作用变量的定义相应于不变环面 \mathbf{T}^n 上的非平凡回路, 对于每一个独立非平凡回路 C_a , 可以定义相应的 I_a 如下

$$I_a = \oint_{C_a} \frac{\Theta}{2\pi} = \oint_{C_a} \frac{p_b dq^b}{2\pi}, \tag{6.12}$$

由于 $\mathcal{M}_g = \mathbf{T}^n$ 由原守恒量 G 定义, 因此这样定义的 I_a 显然是 G 的函数。进而我们也可以将不变环面刻画成 $I_a, a = 1, 2, \dots, n$ 取常数值值的超曲面, 即 $I_a(q, p) = f_a$ 的超曲面, 不同的 f 值就将整个相空间分层成不同的不变环面。比如, 第一章讲势能曲线的时候, 实际上我们已经初步接触过一维情形下的作用变量, 在一维运动的情形中, 相空间是两维的, 而不变环面也是一维的, 就是一条闭合的相轨道, 这时候作用变量就是这条相轨道包围的面积除以 2π , 具体细节读者不妨往回翻看一下。

为了定义角变量 θ^a , 我们仿照对刘维尔定理的证明过程, 首先通过反解 $I_a(q, p) = f_a, a = 1, 2, \dots, n$ 得出 $p_a = p_a(f, q)$, 进而引入正则变换的第二类生成函数 $F(I, q)$, 定义如下

$$F(I, q) = \int_{q_0}^q p_a(I, q) dq^a. \tag{6.13}$$

进而即有

$$p_a = \frac{\partial F}{\partial q^a}, \quad \theta^a = \frac{\partial F}{\partial I_a}. \tag{6.14}$$

θ^a 之所以称作角变量, 是因为它在回路 C_a 上以 2π 为周期, 而在其它 $b \neq a$ 的回路 C_b 上周期是零, 即满足

$$\oint_{C_b} \frac{d\theta^a}{2\pi} = \delta_{ab}^a. \tag{6.15}$$

证明如下,

$$\begin{aligned}
 \oint_{C_b} d\theta^a &= \oint_{C_b} \frac{\partial^2 F}{\partial q^c \partial I_a} dq^c \\
 &= \frac{\partial}{\partial I_a} \oint_{C_b} \frac{\partial F}{\partial q^c} dq^c = \frac{\partial}{\partial I_a} \oint_{C_b} p_c dq^c \\
 &= (2\pi) \frac{\partial I_b}{\partial I_a} = (2\pi) \delta_b^a.
 \end{aligned} \tag{6.16}$$

类似于前面对刘维尔定理的讨论, 作用变量和角变量满足的哈密顿正则方程为

$$\dot{I}_a = 0, \quad \dot{\theta}^a = \frac{\partial H}{\partial I_a} = \omega_a(I), \tag{6.17}$$

式中 $\omega_a(I)$ 定义为 $\frac{\partial H}{\partial I_a}$, 为系统在 C_a 方向上运动的角频率。不仅如此, 而且第一章谈到的玻尔-索末菲量子化条件也可以推广为

$$I_a = n_a \hbar, \tag{6.18}$$

式中 n_a 为非负整数, 是一个标记量子状态的量子数。假设我们以作用变量和角变量为相空间的坐标, 那玻尔-索末菲量子化条件将带来一个直观的相空间图像, 对于一维运动情形, 它意味着相空间沿着 θ 轴可以划分成一个个长为 2π 的区间, 而沿着 I 轴可以划分成一个个长为 \hbar 的区间, 这样就可以将整个相空间分成一些宽为 2π , 高为 \hbar , 从而面积为 $2\pi\hbar$, 的小格子, 称之为相格。玻尔-索末菲量子化条件告诉我们, 每一个相格对应一个量子状态。推广到 n 自由度情形即是, 相空间可以划分成一些体积为 $(2\pi\hbar)^n$ 的相格, 每一个相格对应一个量子状态。

6.1.4 开普勒问题和两维谐振子问题

开普勒问题

前面的章节中求解过粒子在中心力场 $V(r) = -\frac{k}{r} (k > 0)$ 中运动的问题, 也就是所谓的开普勒问题。我们知道, 这个问题可以化简为一个平面运动问题, 相应的哈密顿量为

$$H = \frac{p_r^2}{2m} + \frac{p_\theta^2}{2mr^2} - \frac{k}{r}. \tag{6.19}$$

其中 p_θ 就是角动量, 有时候也记为 J , 它当然是一个守恒量, 因此 p_θ 与 H 泊松对易。所以, 我们有两个相互泊松对易的守恒量, H 和 p_θ , 而这又是一个两自由度的两维运动问题, 所以它构成一个可积系统。我们可以选取 H, p_θ 为正则动量 (两者看成相互独立), 相应的两个正则坐标分别记为 ψ_H, ψ_θ 。很明显

$$\psi_H = \frac{\partial H}{\partial H} = 1, \quad \psi_\theta = 0. \tag{6.20}$$

由此即知道 $\psi_H = t - t_0$, ψ_θ 为常数。

如果我们想进一步求出 r, θ 如何随时间演化, 那根据刘维尔定理的证明过程, 就需要将守恒量 H, p_θ 以及相应的两个正则坐标 ψ_H, ψ_θ 正则变换回去。这个正则变换的生成函

数 (第二类生成函数) $F(H, p_\theta, r, \theta)$ 为

$$\begin{aligned} F(H, p_\theta, r, \theta) &= \int (p_r dr + p_\theta d\theta) \\ &= \int dr \sqrt{2m(H - \frac{p_\theta^2}{2mr^2} + \frac{k}{r})} + p_\theta \theta. \end{aligned} \quad (6.21)$$

这个生成函数按照下式生成正则变换

$$\begin{aligned} t - t_0 = \psi_H &= \frac{\partial F}{\partial H} = \int dr \frac{m}{\sqrt{2m(H - \frac{p_\theta^2}{2mr^2} + \frac{k}{r})}} \\ \psi_\theta &= \frac{\partial F}{\partial p_\theta} = - \int dr \frac{p_\theta / r^2}{\sqrt{2m(H - \frac{p_\theta^2}{2mr^2} + \frac{k}{r})}} + \theta \end{aligned} \quad (6.22)$$

这里的 ψ_θ 就可以理解为初始角度 θ_0 。读者可以将这个结果与第四章第五章的相应结果进行比较, 不同的方法结论是完全一致的。

下面我们看看如何用作用-角变量处理开普勒问题。对于束缚运动 (符合运动区域有限的条件), 开普勒问题的两个作用变量 I_1, I_2 分别为

$$\begin{aligned} I_2 &= \oint \frac{p_\theta d\theta}{2\pi} = p_\theta \\ I_1 &= \oint \frac{p_r dr}{2\pi} = \frac{1}{2\pi} \oint dr \sqrt{2m(H - \frac{p_\theta^2}{2mr^2} + \frac{k}{r})} \\ &= \frac{1}{\pi} \int_{r_{min}}^{r_{max}} dr \sqrt{2m(H - \frac{p_\theta^2}{2mr^2} + \frac{k}{r})}. \end{aligned} \quad (6.23)$$

式中对 r 的回路积分表示从下限 r_{min} 积到上限 r_{max} , 然后再积回下限 r_{min} 。 r_{min}, r_{max} 由 $p_r = 0$ 确定, 从而是方程 $H - \frac{p_\theta^2}{2mr^2} + \frac{k}{r} = 0$ 的两个根, 两根之和以及两根之积分别为

$$r_{min} + r_{max} = -\frac{k}{H}, \quad r_{min} r_{max} = -\frac{p_\theta^2}{2mH}. \quad (6.24)$$

由此也可以知道, 必有总能量 $H < 0$ 。为了进一步算出 I_1 表达式中的积分, 我们可以利用如下公式

$$\int_{r_{min}}^{r_{max}} dr \sqrt{(1 - \frac{r_{min}}{r})(\frac{r_{max}}{r} - 1)} = \frac{\pi}{2}(r_{min} + r_{max}) - \pi \sqrt{r_{min} r_{max}}. \quad (6.25)$$

利用这些结果, 不难算出

$$I_1 = -I_2 + k \sqrt{\frac{m}{-2H}}. \quad (6.26)$$

进而即有

$$H = -\frac{mk^2}{2(I_1 + I_2)^2} = E. \quad (6.27)$$

从上面的结果容易得到 $\omega_1 = \frac{\partial H}{\partial I_1} = \frac{\partial H}{\partial I_2} = \omega_2$, 以及

$$\omega_1 = \omega_2 = \frac{mk^2}{(I_1 + I_2)^3} = \frac{m}{k} \left(-\frac{2E}{m} \right)^{3/2}. \quad (6.28)$$

这说明两件事情: 第一, 由于 $\omega_1/\omega_2 = 1$ 是一个有理数, 这说明粒子的运动是一个周期运动, 特别的, 运动轨道是闭合轨道。第二, 周期 $T = \frac{2\pi}{\omega_1} = (2\pi)km^{1/2}(-2E)^{-3/2}$, 这正是开普勒第三定律, 这是因为半长轴 $a = k(-2E)^{-1}$, 所以 $T = (2\pi)\sqrt{\frac{m}{k}}a^{3/2}$ 。

二维谐振子问题

下面再考察一个例子, 即一个二维谐振子, 其哈密顿量为

$$H = \frac{1}{2}(p_1^2 + \omega_1^2 q_1^2 + p_2^2 + \omega_2^2 q_2^2). \quad (6.29)$$

很显然系统 1,2 两个分量的哈密顿量都是守恒量, 而且相互泊松对易, 分别记为 H_1, H_2 , $H_1 = \frac{1}{2}(p_1^2 + \omega_1^2 q_1^2)$, $H_2 = \frac{1}{2}(p_2^2 + \omega_2^2 q_2^2)$ 。同样很明显的是, $H_1 = E_1$ 给出的是相空间 (q_1, p_1) 平面上的一个椭圆, 椭圆在拓扑上当然等同于圆周 S^1 。同样 $H_2 = E_2$ 也是 (q_2, p_2) 平面上的椭圆, 也拓扑等价于 S^1 , 所以系统的运动正是限制在这两个 S^1 的笛卡尔积上, 这也就是相空间的不变环面 $\mathbf{T}^2 = S^1 \times S^1$ 。所以这个系统当然是一个可积系统, 人们不难按照求解可积分系统的一般办法得出系统的解析解。不过, 我们这里真正想讨论的是如何构造这个系统的作用-角变量。

人们当然可以按照构造作用-角变量的一般方法找到这个系统的作用-角变量, 但这不是最快捷的方法。对于这个系统而言最快捷的方法就是先找到参数化两个独立 S^1 的角变量, 然后通过正则变换进一步找到与角变量对应的作用变量。具体来说, 对于 $H_1 = \frac{1}{2}(p_1^2 + \omega_1^2 q_1^2) = E_1$ 的椭圆, 我们可以定义相空间的坐标变换

$$p_1 = \rho_1 \cos \theta_1, \quad \omega_1 q_1 = \rho_1 \sin \theta_1, \quad (6.30)$$

式中 θ_1 正是参数化这个椭圆的角变量, 但满足 $\rho_1^2 = 2E_1$ 的 ρ_1 却不一定是相应的作用变量, 因为上面的变换只是相空间的坐标变换, 却不一定是正则变换。同样, 对于 $H_2 = \frac{1}{2}(p_2^2 + \omega_2^2 q_2^2) = E_2$ 的椭圆, 我们也可以定义相空间的坐标变换

$$p_2 = \rho_2 \cos \theta_2, \quad \omega_2 q_2 = \rho_2 \sin \theta_2, \quad (6.31)$$

式中 θ_2 也是一个角变量。

找到作用变量的关键是找到以上面两个角变量 θ_1, θ_2 为正则坐标的正则变换, 而找到正则变换的关键则是要求相空间的坐标变换保持辛结构。为此我们进行如下计算

$$\begin{aligned} \omega &= dp_1 \wedge dq_1 + dp_2 \wedge dq_2 \\ &= \frac{1}{\omega_1} \rho_1 d\rho_1 \wedge d\theta_1 + \frac{1}{\omega_2} \rho_2 d\rho_2 \wedge d\theta_2 \\ &= d\left(\frac{1}{2\omega_1} \rho_1^2\right) \wedge d\theta_1 + d\left(\frac{1}{2\omega_2} \rho_2^2\right) \wedge d\theta_2. \end{aligned} \quad (6.32)$$

从这个计算可以知道, 只需令 $I_1 = \frac{1}{2\omega_1} p_1^2$, $I_2 = \frac{1}{2\omega_2} p_2^2$, 就会有

$$\omega = dp_1 \wedge dq_1 + dp_2 \wedge dq_2 = dI_1 \wedge d\theta_1 + dI_2 \wedge d\theta_2. \quad (6.33)$$

这正是正则变换的保辛结构条件。也即是说, 上面定义的 I_1, I_2 正是我们要找的作用变量。

不妨将作用-角变量与原来变量之间的正则变换关系明显地写出来,

$$\begin{aligned} p_1 &= \sqrt{2I_1\omega_1} \cos \theta_1, & q_1 &= \sqrt{\frac{2I_1}{\omega_1}} \sin \theta_1, \\ p_2 &= \sqrt{2I_2\omega_2} \cos \theta_2, & q_2 &= \sqrt{\frac{2I_2}{\omega_2}} \sin \theta_2. \end{aligned} \quad (6.34)$$

很显然, $H_1 = I_1\omega_1$, $H_2 = I_2\omega_2$, 从而

$$H = I_1\omega_1 + I_2\omega_2. \quad (6.35)$$

利用作用-角变量的哈密顿正则方程, 我们容易得到

$$\dot{\theta}_1 = \frac{\partial H}{\partial I_1} = \omega_1, \quad \dot{\theta}_2 = \frac{\partial H}{\partial I_2} = \omega_2. \quad (6.36)$$

从而即有 $\theta_1 = \omega_1 t + \varphi_1$, $\theta_2 = \omega_2 t + \varphi_2$, 代回 (6.34) 式就可以得到原来的变量 q_1, q_2, p_1, p_2 的解。

6.1.5 绝热不变量

作用变量还有一重身份, 就是作为所谓的绝热不变量。为了讲清楚这个问题, 我们考虑一个一维系统, 假设其势能曲线将系统约束在一个有限的区间上作周期运动, 从而其相空间的不变环面就是一条相空间闭合回路 C , 拓扑上等价于 S^1 。

进一步, 再假设势能函数依赖于某个控制参数 λ , 记为 $V(q, \lambda)$, 因此系统的哈密顿量为

$$H = \frac{p^2}{2m} + V(q, \lambda). \quad (6.37)$$

现在, 假设我们以**几乎恒定的速率**缓慢地调节参数 λ , 即让 λ 以几乎恒定的速率随时间缓慢变化, 记为 $\lambda(t)$, 那这就变成了一个显含时间 t 的系统, 能量 E 因此不再守恒, 而是也依赖于时间, 记为 $E(t)$ 。很显然, 如果参数 λ 相对于系统的运动周期来说变化得足够缓慢, 那系统将依然近似作周期运动, 但是它的各个物理量 (比如能量) 都将同时随时间缓慢变化。所谓的绝热不变量, 指的就是这样一个特殊物理量, 它在参数的缓慢调节下保持不变! 下面我们将证明, 作用变量 $I = \frac{1}{2\pi} \oint pdq$ 就是一个绝热不变量。

首先, 注意到在每一个瞬时均有 $E = \frac{p^2}{2m} + V(q, \lambda)$, 从而

$$p = \sqrt{2m[E - V(q, \lambda)]} = p(q, E, \lambda), \quad (6.38)$$

即我们将 E 和 λ 看成两个独立变量, 并用它们表达出动量 p 。从而作用变量 $I = \frac{1}{2\pi} \oint pdq$ 也是 E, λ 的函数, 记为 $I(E, \lambda)$ 。随着参数的缓慢调节, E 和 λ 都将随时间缓慢变化, 但是, 这两者的变化并不相互独立, 而是满足

$$\dot{E} = \frac{\partial H}{\partial t} = \frac{\partial H}{\partial \lambda} \dot{\lambda}. \quad (6.39)$$

这使得这两种变化对作用变量 I 的贡献正好相互抵消, 从而使 I 绝热不变。下面我们就是要证明这一点。

为此, 我们进行如下计算

$$\dot{I} = \frac{\partial I}{\partial E} \dot{E} + \frac{\partial I}{\partial \lambda} \dot{\lambda} = \left(\frac{\partial I}{\partial E} \frac{\partial H}{\partial \lambda} + \frac{\partial I}{\partial \lambda} \right) \dot{\lambda}. \quad (6.40)$$

根据第一章关于势能曲线的相关讨论可以知道,

$$\frac{\partial I}{\partial E} = \frac{T}{2\pi}, \quad (6.41)$$

式中 T 为系统的运动周期。另一方面, 我们也有

$$\begin{aligned} \frac{\partial I}{\partial \lambda} &= \frac{1}{2\pi} \frac{\partial}{\partial \lambda} \oint p dq = \frac{1}{2\pi} \oint \frac{\partial p}{\partial \lambda} dq \\ &= \frac{1}{2\pi} \int_0^T \frac{\partial p}{\partial \lambda} \frac{dq}{dt} dt = \frac{1}{2\pi} \int_0^T \frac{\partial p}{\partial \lambda} \frac{\partial H}{\partial p} dt. \end{aligned} \quad (6.42)$$

为了算出 $\frac{\partial p}{\partial \lambda}$, 我们将 $p = p(q, E, \lambda)$ 的表达式代入哈密顿量 $H(q, p, \lambda)$, 并将等式 $H(q, p, \lambda) = E$ 对独立变量 λ 求导, 即有

$$\frac{\partial H}{\partial \lambda} + \frac{\partial H}{\partial p} \frac{\partial p}{\partial \lambda} = 0. \quad (6.43)$$

将这个结果代入 (6.42) 式即有

$$\frac{\partial I}{\partial \lambda} = -\frac{1}{2\pi} \int_0^T \frac{\partial H}{\partial \lambda} dt. \quad (6.44)$$

将这个结果和 (6.41) 式一起代入 (6.40) 式, 即得

$$\dot{I} = \left(T \frac{\partial H}{\partial \lambda} - \int_0^T \frac{\partial H}{\partial \lambda} dt \right) \frac{\dot{\lambda}}{2\pi}. \quad (6.45)$$

我们看到, 右边的两项已经接近了, 并且呈现出抵消的趋势, 但是, 两者还是有点差别。

下一步, 我们将 (6.45) 式对一个运动周期求平均。任何一个物理量 $A(t, \lambda)$ 的周期平均记为 \bar{A} , 定义如下

$$\bar{A} = \frac{1}{T} \int_0^T dt A(t, \lambda). \quad (6.46)$$

很明显, 如果我们考察 $\bar{\dot{I}}$, 那右边的两项就正好抵消了! 从而即有

$$\bar{\dot{I}} = 0. \quad (6.47)$$

从这个结果也容易得到

$$I(t+T) = I(t). \quad (6.48)$$

也即是说, 对于足够缓慢的参数调节来说, 作用变量最多是以运动周期为周期的函数, 或者说作用变量在一个周期之内的平均变化率为零, 这才是前面我们说作用变量是绝热不变

量的精确含义。注意到相对于参数变化的时间尺度而言,运动周期 T 是非常短的,因此这其实是说,作用变量在参数变化的时间尺度上近似不变。

以一维运动的谐振子为例,其哈密顿量为 $H = I\omega$,因此即使谐振子的频率 ω 随时间缓慢变化,谐振子能量 E 与 ω 的比值 E/ω 也近似保持不变。

实际上,历史上在从经典力学到量子力学的过渡过程中,人们首先注意到了作用变量的绝热不变性,即注意到作用变量在参数的缓慢变化下保持不变的性质,这使得人们自然假设它是一个量子化的量(因为量子化的量不能连续变化,从而自然在参数的连续变化下保持不变),由此人们才提出作用变量的玻尔-索末菲量子化条件。

6.2 近可积系统与混沌系统

近可积系统与 KAM 定理

前面已经看到,对于有限区域内的运动,可积系统的相空间分层成不同的不变环面 \mathbf{T}^n ,并且我们可以将 \mathbf{T}^n 的坐标取作角变量 $\theta = (\theta^1, \theta^2, \dots, \theta^n)$,将不同层的不变环面对应不同的作用变量 $I = (I_1, I_2, \dots, I_n)$ 。即是说,我们可以用作用-角变量来描述可积系统,在这一描述下系统的哈密顿量仅为作用变量的函数。

假设我们有一个可积系统 $H_0(I)$,现在我们给它加上一个微扰,使得整个系统的哈密顿量变成

$$H(I, \theta) = H_0(I) + \varepsilon H_1(I, \theta), \quad (6.49)$$

式中 ε 代表一个小量, $H_1(I, \theta)$ 代表扰动,它不仅依赖于原来的作用变量同时还依赖于原来的角变量。这个扰动以后的系统就是一个近可积系统。

起初人们认为,扰动会破坏可积系统的不变环面,使得原来 n 维不变环面上的相轨道在扰动后弥漫到整个 $2n-1$ 维等能量曲面(注意,哈密顿系统的能量守恒)。但是,1954年 Kolmogorov 在世界数学家大会上指出:非退化的可积系统在上述扰动之后,虽然某些不变环面会被扰动破坏掉(称为共振环面),但所谓满足强非共振条件的不变环面(这种不变环面占大多数)仍然会被保存下来。也就是说近可积系统整个相空间中大部分的相轨道依然是非常规则的!这个发现后来被 Arnold 和 Moser 分别给予了严格的证明,这就是著名的 KAM 定理。

证明 KAM 定理的基本办法是某种改进的牛顿迭代法。粗略地说,就是找一系列的正则变换对扰动后得到的近可积系统进行步步变换,使之越来越靠近一个可积系统(只要对大部分相空间区域能做到这一点就行)。问题是,这个迭代展开在分母上依赖于 $\tilde{\omega} \cdot \vec{m}$, $\tilde{\omega}$ 就是原来不变环面的角频率, \vec{m} 是任何形如 (m_1, m_2, \dots, m_n) 的整数组。因此这个展开对于共振环面(即存在某个 $\vec{m} \neq 0$ 使得 $\tilde{\omega} \cdot \vec{m} = 0$ 的不变环面)是不收敛的。不仅如此,对于任何不变环面,实际上人们总可以选择 \vec{m} ,使得 $\tilde{\omega} \cdot \vec{m}$ 任意接近零,因此看起来这个迭代展开对于任何不变环面都不收敛,这就是所谓的“小分母”困难。Arnold 和 Moser 正是利用改进的牛顿迭代法解决了“小分母”问题,证明对于满足所谓强非共振条件的不变环面,改进后的迭代展开收敛,从而证明了 KAM 定理。对具体证明过程感兴趣的读者,可以阅读 Jurgen Poschel 的文章 A Lecture on the Classical KAM Theorem, <https://arxiv.org/abs/0908.2234v1>。

在那些不变环面被扰动破坏的相空间区域，随着不变环面的破坏可能发生两件事情：第一，系统的相空间轨迹可能会与邻近轨迹发生剧烈的甚至随时间指数增长的偏离，从而导致对初始条件的极端敏感性和混沌行为。第二，它可能使得系统演化缺乏可预测性。即使我们处理的是严格的确定性系统，它也可能产生有效的随机输出，从而使得系统丧失可预测性。

太阳系是否稳定呢？这个大问题曾经引发大量的工作，直到 KAM 定理提供一个也许肯定的回答。太阳系是一个多体问题，具有明显的层次性，其行星与行星之间的相互作用比行星与太阳之间的相互作用要弱得多，可以看成是一个微扰。而行星与太阳的相互作用则是一个可积系统，其解就是开普勒的椭圆轨道。因此太阳系是一个近可积系统，KAM 定理可以适用，由于这个原因，该系统的大体结构是稳健的！但是，我们并不能非常确定目前的太阳系处在相空间的稳定区域。

混沌系统

假设我们考察的既不是可积系统，又不是近可积系统，而是一个一般性的哈密顿系统，其泊松对易的守恒量的数目少于自由度数 n ，甚至当自由度数 n 很大时大大少于 n ，比方说只有一个守恒量，即能量。那这时候系统的动力学行为往往不再规则，而是常常会表现出混沌。

前面的章节中，我们讨论过不少这样的系统，这里不妨简单回顾一下。首先，第二章的最后，我们讨论过双摆系统，那是一个两自由度的系统，但是它只有一个守恒量，即能量。我们说过，当能量足够大时这个系统会表现出混沌行为。其次，第四章我们讨论过受限三体问题，我们说过庞加莱正是在研究受限三体问题时发现混沌的。具体到第四章中对受限三体问题的那个简化模型，它显然是一个两自由度系统，但是只有能量这一个守恒量，因此出现混沌行为也不奇怪。至于一般性的三体问题，显然它有 9 个自由度，但是守恒量只有能量、总动量（三个分量）、还有总角动量（三个分量），共 7 个守恒量，而且相互泊松对易的守恒量实际上只有四个，所以当然一般会展现出混沌行为。最后，第五章中我们考察过两维平面上的涡旋系统，我们只找到 3 个相互泊松对易的守恒量，所以对于 3 个涡旋的情形，那是一个可积系统，但是 4 个和 4 个以上涡旋的运动则通常是混沌的。

7. 振动

本章将研究两类振动问题，其一是力学系统在其稳定平衡位置附近的微小振动，其二是一个振动的力学系统在参数的周期性变化下发生的失稳现象，即所谓的参数共振。我们将主要使用拉格朗日框架分析小振动问题，而主要使用哈密顿力学框架分析参数共振问题。

7.1 小振动

对于一个处于稳定平衡状态的力学系统，给它一个小扰动，它就会在平衡位置附近来回振动，这就是小振动。正如我们即将证明的，小振动是可积的，而且只要合适地选取系统的广义坐标，小振动可以化简成 n 个独立谐振子的简谐振动， n 就是系统的自由度数。

7.1.1 小振动的拉格朗日量

为了研究小振动，我们假设系统的动能 T 具有如下一般形式，

$$T = \frac{1}{2} m_{ab}(q) \dot{q}^a \dot{q}^b, \quad (7.1)$$

这里 q^a 为系统的广义坐标，式中的 $m_{ab}(q)$ 关于两个下标对称，即满足 $m_{ab}(q) = m_{ba}(q)$ ，而且我们使用了求和约定（后文也一样）。从而系统的拉格朗日量 L 可以写成

$$L = \frac{1}{2} m_{ab}(q) \dot{q}^a \dot{q}^b - V(q), \quad (7.2)$$

$V(q)$ 表示势能。

不失一般性的, 我们假设 $q^a = 0$ 为系统的稳定平衡位置, 根据第一章中的相关讨论可以知道, 它也就是势能 $V(q)$ 的极小值位置。要研究的就是 $q^a = 0$ 附近的小振动, 为此可以将上面的拉格朗日量 (7.2) 在 $q = 0$ 处泰勒展开, 将展开量 q^a 看成小量, 进而保留到这个小量的二阶, 即有

$$L = \frac{1}{2} m_{ab}(0) \dot{q}^a \dot{q}^b - V(0) - \frac{\partial V}{\partial q^a} \Big|_{q=0} q^a - \frac{1}{2} \frac{\partial^2 V}{\partial q^a \partial q^b} \Big|_{q=0} q^a q^b + \dots \quad (7.3)$$

在这个式子中, 首先, $V(0)$ 是一个常数, 对于拉格朗日量来说无关紧要, 因此可以忽略。其次, 动能一定大于等于零, 从而 $m_{ab}(0)$ 是一个对称的正定矩阵, 不妨简记作 m_{ab} , 当然这样写出来的是矩阵元形式, 相应的 $n \times n$ 矩阵可以记为 M 。再次, 由于 $q = 0$ 为平衡位置, 也就是势能的极值位置, 从而必有 $\frac{\partial V}{\partial q^a} \Big|_{q=0} = 0$ 。最后, 假设记 $\frac{\partial^2 V}{\partial q^a \partial q^b} \Big|_{q=0} = k_{ab}$, 则由于势能的这个极值是极小值, 所以 k_{ab} 也必定是对称正定矩阵, 同样这写的是矩阵元形式, 相应的矩阵记为 K 。综上所述, 我们就能将小振动的拉格朗日量近似成,

$$L = \frac{1}{2} m_{ab} \dot{q}^a \dot{q}^b - \frac{1}{2} k_{ab} q^a q^b \quad (7.4)$$

假设引入坐标列矢量 $\mathbf{q} = (q^1, q^2, \dots, q^n)^T$, T 代表矩阵转置, 则也可以将上述拉格朗日量写成

$$L = \frac{1}{2} \dot{\mathbf{q}}^T M \dot{\mathbf{q}} - \frac{1}{2} \mathbf{q}^T K \mathbf{q}. \quad (7.5)$$

7.1.2 简正坐标的存在性

这一小节我们将证明, 可以合适地选取广义坐标进而将小振动的拉格朗日量 (7.4) 和 (7.5) 化简成 n 个独立简谐振子的拉格朗日量。

注意到 (7.5) 式中的矩阵 M, K 均是对称正定矩阵, 我们可以引用关于对称矩阵的一个引理。**引理: 任何实对称矩阵都可以正交变换为对角矩阵。**即, 对于任意实对称矩阵 B , 必定存在正交矩阵¹ R , 使得

$$B = R^T D R, \quad (7.6)$$

式中 D 是对角矩阵, 其各对角元就是对称矩阵 B 的各本征值, 如果 B 同时为正定矩阵, 那这些对角元必定都大于零。关于这个引理的证明, 请参阅线性代数教材。

推论: 对于任何正定的实对称矩阵 B , 必定存在可逆矩阵 C , 使得

$$B = C^T C. \quad (7.7)$$

证明非常简单, 只要根据上面的引理取 $C = D^{\frac{1}{2}} R$ 即可。由于假定 B 正定, 所以对角矩阵 D 也正定, 从而可逆, 从而 C 的逆矩阵即是 $C^{-1} = R^T D^{-\frac{1}{2}}$ 。

下面我们将上述引理和推论应用于小振动的研究。首先, 注意到 (7.5) 式中的 M 为正定对称矩阵, 从而根据上面的推论知, 必定存在可逆矩阵 C , 使得 $M = C^T C$, 进而可以将 (7.5) 式写成

$$L = \frac{1}{2} \dot{\mathbf{q}}^T C^T C \dot{\mathbf{q}} - \frac{1}{2} \mathbf{q}^T K \mathbf{q}. \quad (7.8)$$

¹所谓的正交矩阵 R , 即满足 $R^T R = R R^T = 1$ 的矩阵, 这里 1 表示单位矩阵。

下面引入新的广义坐标 \mathbf{q}' ,

$$\mathbf{q} = C^{-1}\mathbf{q}'. \quad (7.9)$$

从而即有

$$L = \frac{1}{2}\dot{\mathbf{q}}'^T\dot{\mathbf{q}}' - \frac{1}{2}\mathbf{q}'^T C^{-1T} K C^{-1} \mathbf{q}'. \quad (7.10)$$

很显然, 式中的 $C^{-1T} K C^{-1}$ 依然是正定对称矩阵, 不妨记 $C^{-1T} K C^{-1} = K'$, 则根据上面的引理, 必定存在正交矩阵 R , 使得 $K' = R^T D R$, 这里 D 为正定的对角矩阵。从而 (7.10) 式可以写成

$$L = \frac{1}{2}\dot{\mathbf{q}}'^T\dot{\mathbf{q}}' - \frac{1}{2}\mathbf{q}'^T R^T D R \mathbf{q}'. \quad (7.11)$$

引入最终的广义坐标 $\mathbf{Q} = (Q_1, Q_2, \dots, Q_n)^T$,

$$\mathbf{Q} = R\mathbf{q}', \quad (7.12)$$

从而即有

$$L = \frac{1}{2}\dot{\mathbf{Q}}^T\dot{\mathbf{Q}} - \frac{1}{2}\mathbf{Q}^T D \mathbf{Q}. \quad (7.13)$$

最后将正定对角矩阵写成 $D = \text{diag}\{\omega_1^2, \dots, \omega_n^2\}$, 即可以得到

$$L = \sum_{\alpha=1}^n \frac{1}{2} (\dot{Q}_\alpha^2 - \omega_\alpha^2 Q_\alpha^2). \quad (7.14)$$

最后的这个表达式我们没有使用求和约定, 很明显, 这正是 n 个独立简谐振子的拉格朗日量, ω_α 就是这些谐振子的角频率。到此为止就完成了本节的证明。

值得说明的是, (7.14) 式中的广义坐标 $Q_\alpha, \alpha = 1, 2, \dots, n$ 就是所谓的简正坐标, 角频率 ω_α 就是所谓的简正频率。

假设引入与简正坐标对应的广义动量 $P_\alpha = \frac{\partial L}{\partial \dot{Q}_\alpha} = \dot{Q}_\alpha$, 则我们很容易通过勒让德变换得到与 (7.14) 式相应的哈密顿量, 为

$$H = \sum_{\alpha} \frac{1}{2} (P_\alpha^2 + \omega_\alpha^2 Q_\alpha^2). \quad (7.15)$$

很显然, 这个系统存在 n 个相互泊松对易的守恒量, 为 $H_\alpha = \frac{1}{2} (P_\alpha^2 + \omega_\alpha^2 Q_\alpha^2), \alpha = 1, 2, \dots, n$, 从而这也就是一个可积系统。但 (7.15) 式不过是小振动问题在合适正则变量下的表述形式, 从而这也就证明小振动是可积的。

7.1.3 如何寻找简正坐标

上一小节只是证明了简正坐标的存在性, 那里的方法并不是我们实际寻找简正坐标的方法。为了实际寻找简正坐标, 我们回到最初的小振动拉格朗日量 (7.4) 和 (7.5)。从 (7.4) 式利用拉格朗日方程, 可以得到一组关于 $q^a(t)$ 的运动微分方程,

$$m_{ab}\ddot{q}^b + k_{ab}q^b = 0. \quad (7.16)$$

写成矩阵形式, 即

$$M\ddot{\mathbf{q}} + K\mathbf{q} = 0. \quad (7.17)$$

我们知道, 简正坐标总是在做简单的简谐振动, 因此为了寻找简正坐标, 我们试着寻求方程 (7.17) 形如下式的解

$$\mathbf{q}(t) = \mathbf{a} \cos(\omega t + \varphi), \quad (7.18)$$

式中常数 \mathbf{a} 为一个 n 维列矢量。

将 (7.18) 式代入 (7.17) 式即可得到

$$(-\omega^2 M + K)\mathbf{a} = 0. \quad (7.19)$$

这是一个关于列矢量 \mathbf{a} 的线性方程组, 这个方程组存在非零解的充要条件是矩阵 $(-\omega^2 M + K)$ 不可逆, 也就是说, 它的行列式要等于零, 从而可得 ω^2 需满足的方程, 即

$$\det(-\omega^2 M + K) = 0. \quad (7.20)$$

这个方程有 n 个解, 不妨记为 $\omega_\alpha^2, \alpha = 1, 2, \dots, n$ 。与 ω_α^2 对应的列矢量 \mathbf{a} 我们记作 $\mathbf{a}_{(\alpha)}$, 即

$$(-\omega_\alpha^2 M + K)\mathbf{a}_{(\alpha)} = 0. \quad (7.21)$$

可以证明 ω_α^2 一定大于零 (从而 ω_α 都是实数), 证明如下: 我们用行矢量 $\mathbf{a}_{(\alpha)}^T$ 从左边乘到方程 (7.21) 上, 进而即可得

$$\omega_\alpha^2 = \frac{\mathbf{a}_{(\alpha)}^T K \mathbf{a}_{(\alpha)}}{\mathbf{a}_{(\alpha)}^T M \mathbf{a}_{(\alpha)}}, \quad (7.22)$$

注意到 K 和 M 均为对称正定矩阵, 即 $\mathbf{a}_{(\alpha)}^T K \mathbf{a}_{(\alpha)} > 0, \mathbf{a}_{(\alpha)}^T M \mathbf{a}_{(\alpha)} > 0$, 从而即可得

$$\omega_\alpha^2 > 0. \quad (7.23)$$

证明完成。

为了简化分析, 下面进一步假设 n 个 ω_α^2 两两不同。由此即可证明

$$\mathbf{a}_{(\beta)}^T M \mathbf{a}_{(\alpha)} = 0, \quad \text{for any } \alpha \neq \beta. \quad (7.24)$$

为了证明这个结论, 我们将方程 (7.21) 转置一下 (注意 M 和 K 均为对称矩阵), 得到

$$-\omega_{(\beta)}^2 \mathbf{a}_{(\beta)}^T M + \mathbf{a}_{(\beta)}^T K = 0. \quad (7.25)$$

下面用 $\mathbf{a}_{(\beta)}^T$ 左乘 (7.21) 式, 用 $\mathbf{a}_{(\alpha)}$ 右乘 (7.25) 式, 然后将这两个结果相减, 即可得

$$(\omega_{(\beta)}^2 - \omega_{(\alpha)}^2) \mathbf{a}_{(\beta)}^T M \mathbf{a}_{(\alpha)} = 0. \quad (7.26)$$

很明显, 若 $\alpha \neq \beta$, 即有 (7.24) 式。进一步若记 $\mathbf{a}_{(\alpha)}^T \mathbf{M} \mathbf{a}_{(\alpha)} = m_\alpha$, 则我们也可以将最终的结果写成如下形式

$$\mathbf{a}_{(\beta)}^T \mathbf{M} \mathbf{a}_{(\alpha)} = m_\alpha \delta_{\alpha\beta}. \quad (7.27)$$

式中的 $\delta_{\alpha\beta}$ 就是所谓的克龙内克符号。利用这个结果, 再将 $\mathbf{a}_{(\beta)}^T$ 左乘 (7.21) 式, 又可以得到

$$\mathbf{a}_{(\beta)}^T \mathbf{K} \mathbf{a}_{(\alpha)} = m_\alpha \omega_\alpha^2 \delta_{\alpha\beta}. \quad (7.28)$$

下面我们引入新的广义坐标 $Q_\alpha, \alpha = 1, 2, \dots, n$, 其定义为

$$\mathbf{q} = \sum_{\alpha=1}^n \mathbf{a}_{(\alpha)} Q_\alpha. \quad (7.29)$$

将这个式子代入小振动拉格朗日量 (7.5) 中, 并利用关系式 (7.27) 和 (7.28), 即可以得到

$$L = \sum_{\alpha} \frac{1}{2} (m_\alpha \dot{Q}_\alpha^2 - m_\alpha \omega_\alpha^2 Q_\alpha^2) \quad (7.30)$$

这正是 n 个独立谐振子的拉格朗日量。这就说明, 新的广义坐标 $Q_\alpha, \alpha = 1, 2, \dots, n$ 其实就是简正坐标。当然, (7.30) 式和上一小节定义简正坐标时标准的 (7.14) 式还略微有些区别, 但这个区别其实无关紧要, 因为人们很容易通过将 $\sqrt{m_\alpha} Q_\alpha$ 重新定义为 Q_α , 进而将 (7.30) 式变为标准的 (7.14) 式。

简正坐标作简谐振动, 从而 $Q_\alpha(t) = c_\alpha \cos(\omega_\alpha t + \varphi_\alpha)$, 这种运动也称为**简正模式**, 相应的振动频率 ω_α 就称为**简正频率**。进而可以由 (7.29) 式得到原坐标 $\mathbf{q}(t)$ 的运动,

$$\mathbf{q}(t) = \sum_{\alpha=1}^n \mathbf{a}_{(\alpha)} c_\alpha \cos(\omega_\alpha t + \varphi_\alpha). \quad (7.31)$$

我们可以将列矢量 $\mathbf{a}_{(\alpha)}$ 的各分量记为 $a_{(\alpha)}^i, i = 1, 2, \dots, n$, 进而将 $a_{(\alpha)}^i$ 看成某个 $n \times n$ 矩阵 A 的第 i 行第 α 列, 即 $A = (a_{(\alpha)}^i)$, A 通常称为模态矩阵 (modal matrix)。如此一来就可以将简正坐标的式子 (7.29) 重写成

$$\mathbf{q} = \mathbf{A} \mathbf{Q}, \quad (7.32)$$

式中 $\mathbf{Q} = (Q_1, \dots, Q_n)^T$ 为简正坐标的列矢量。反过来也有

$$\mathbf{Q} = \mathbf{A}^{-1} \mathbf{q}. \quad (7.33)$$

7.1.4 举例

双摆

第二章最后介绍过双摆, 其拉格朗日量为

$$\begin{aligned} L = & \frac{1}{2} (m_1 + m_2) l_1^2 \dot{\theta}_1^2 + \frac{1}{2} m_2 l_2^2 \dot{\theta}_2^2 + m_2 l_1 l_2 \cos(\theta_1 - \theta_2) \dot{\theta}_1 \dot{\theta}_2 \\ & + (m_1 + m_2) g l_1 \cos(\theta_1) + m_2 g l_2 \cos(\theta_2). \end{aligned} \quad (7.34)$$

下面我们简化一下问题, 设 $m_1 = m_2 = m$, $l_1 = l_2 = l$, 从而双摆的拉格朗日量可以简化为

$$L = ml^2 \dot{\theta}_1^2 + \frac{1}{2} ml^2 \dot{\theta}_2^2 + ml^2 \cos(\theta_1 - \theta_2) \dot{\theta}_1 \dot{\theta}_2 + 2mgl \cos \theta_1 + mgl \cos \theta_2. \quad (7.35)$$

显然, $\theta_1 = \theta_2 = 0$ 为系统的稳定平衡位置。为了研究小振动, 我们在这个平衡位置附近将上述拉格朗日量对小量 θ_1, θ_2 泰勒展开到二阶, 即有

$$L = ml^2 \dot{\theta}_1^2 + \frac{1}{2} ml^2 \dot{\theta}_2^2 + ml^2 \dot{\theta}_1 \dot{\theta}_2 - mgl\theta_1^2 - \frac{1}{2} mgl\theta_2^2. \quad (7.36)$$

利用拉格朗日方程即可以得到小振动的运动微分方程

$$\begin{aligned} 2ml^2 \ddot{\theta}_1 + ml^2 \ddot{\theta}_2 + 2mgl\theta_1 &= 0 \\ ml^2 \ddot{\theta}_2 + ml^2 \ddot{\theta}_1 + mgl\theta_2 &= 0. \end{aligned} \quad (7.37)$$

引入广义坐标列矢量 $\theta = (\theta_1, \theta_2)^T$, 就可以将这组方程写成

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \ddot{\theta} + \frac{g}{l} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \theta = 0. \quad (7.38)$$

为了找到简正坐标和简正模式, 我们需要求解相应的本征方程

$$\left[-\omega^2 \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} + \frac{g}{l} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \right] \mathbf{a} = 0. \quad (7.39)$$

解得一组解为 $\omega_1^2 = (2 - \sqrt{2})(\frac{g}{l})$, 相应的 $\mathbf{a}_{(1)} = (1, \sqrt{2})^T$, 另一组解为 $\omega_2^2 = (2 + \sqrt{2})(\frac{g}{l})$, 相应的 $\mathbf{a}_{(2)} = (1, -\sqrt{2})^T$ 。从而可以得到模态矩阵 A , 为

$$A = \begin{pmatrix} 1 & 1 \\ \sqrt{2} & -\sqrt{2} \end{pmatrix} \quad (7.40)$$

从而简正坐标为

$$\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & \frac{\sqrt{2}}{2} \\ 1 & -\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}. \quad (7.41)$$

当然那个整体的 $\frac{1}{2}$ 因子并不重要。

7.2 参数共振

7.2.1 现象与问题

参数共振就是通过周期性地改变系统的参数而产生的共振现象。日常生活中人们在荡秋千时可以观察到这种现象。如图 (7.1) 所示, 荡秋千的人可以通过适时地作出下蹲和起立的动作来把秋千荡得越来越高。具体来说, 每次秋千摆到最高位置时, 人要迅速蹲下, 使重心降低, 如图 (7.1) 中的第 1, 3 幅图, 而秋千通过平衡位置时, 人又要迅速起立, 使重心升高, 如图 (7.1) 中的第 2 幅图。这样, 秋千来回荡一次, 整个系统的重心就呈周期性地上升和下降两次, 系统重心的运动轨迹可以绘成图 (7.1) 中的最后一幅图。

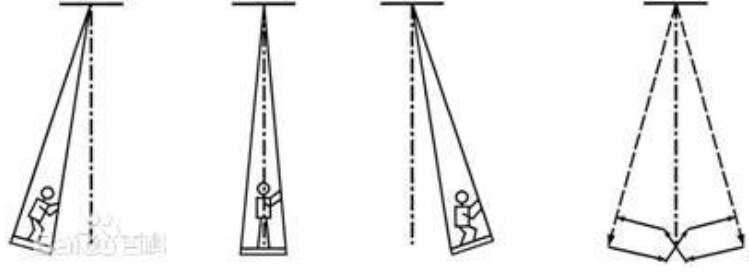


Figure 7.1: 荡秋千时的参数共振。图片来自网络。

很显然，重心的位置决定的是秋千的有效摆长，假设记秋千的固有频率 $\omega_0 = \sqrt{\frac{g}{l_0}}$ ，则上述过程相当于以 $2\omega_0$ 的频率周期性改变秋千的有效摆长 l 。问题是，这种共振是如何发生的呢？特别是，为什么人的蹲起频率得是秋千频率的两倍呢？另外，秋千基本上是单自由度的力学系统，多自由度力学系统有没有参数共振现象呢？

7.2.2 参数共振的一般性分析

为了一般性地分析参数共振现象，我们考察一个 n 自由度系统，以列矢量 \mathbf{x} 表示其相空间坐标， $\mathbf{x} = (q^1, \dots, q^n, p_1, \dots, p_n)^T$ ， T 表示矩阵转置，并笼统地以 $x^i, i = 1, 2, \dots, 2n$ 表示 \mathbf{x} 的各个分量。假设系统的哈密顿量为

$$H = \frac{1}{2} B_{ij} x^i x^j = \frac{1}{2} \mathbf{x}^T \mathbf{B} \mathbf{x}, \quad (7.42)$$

式中系统的参数 B_{ij} 构成一个 $2n \times 2n$ 的对称矩阵，记为 \mathbf{B} 。由于能量总大于零，即 $H > 0$ ，因此一般来说我们要求 \mathbf{B} 是一个正定矩阵。进一步假设这些参数依赖于时间并按照周期 T 变化，即

$$B_{ij}(t+T) = B_{ij}(t), \quad (7.43)$$

或者也可以写作 $\mathbf{B}(t+T) = \mathbf{B}(t)$ 。这类系统的最简单例子就是一个频率周期性变化的谐振子，这时候 $H = \frac{1}{2}(p^2 + \omega^2 q^2)$ ， $\omega^2(t+T) = \omega^2(t)$ 。

上述系统的哈密顿正则方程为

$$\dot{x}^i = \omega^{ij} \partial_j H = \omega^{ij} B_{jk}(t) x^k. \quad (7.44)$$

数学家弗洛凯最早研究了这类参数随时间周期性变化的一阶微分方程组，这就是所谓的弗洛凯理论 (Floquet theory)。我们注意到，对于这类方程，若 $x^i(t)$ 为其解，则 $x^i(t) = x^i(t+T)$ 同样为其解。由于这是一组一阶线性微分方程，它的解空间必定为一个线性空间，因此这就意味着， $x^i(t+T)$ 必定为 $x^i(t)$ 的线性变换，即必定存在矩阵 S^i_j 使得

$$x^i(t+T) = S^i_j x^j(t). \quad (7.45)$$

进一步，由于方程 (7.44) 是一个哈密顿正则方程，从而根据第五章的知识可以知道， $x^i(t)$ 的时间演化本身是正则变换，从而 $x^i(t+T)$ 与 $x^i(t)$ 之间的线性变换必定同时是正则变换。

记矩阵 S 的第 i 行第 j 列为 S^i_j , 那么怎么算出变换矩阵 S 呢? 方法之一是在 $t = 0$ 时刻取方程 (7.44) $2n$ 个线性无关的初始条件 $x^i_{(j)}(0), j = 1, 2, \dots, 2n$ 为

$$x^i_{(j)}(0) = \delta^i_j, \quad (7.46)$$

然后解出它们演化到 T 时刻的值 $x^i_{(j)}(T)$, 则由于

$$x^i_{(j)}(T) = S^i_k x^k_{(j)}(0) = S^i_j, \quad (7.47)$$

因此这组 T 时刻的值正好给出矩阵 S^i_j 。

矩阵 S 的重要性在于, 它决定了方程 (7.44) 解的长时间行为。这是因为, 根据 (7.45) 我们有

$$\mathbf{x}(t+T) = S\mathbf{x}(t) \Rightarrow \mathbf{x}(NT) = S^N\mathbf{x}(0), \quad (7.48)$$

很显然, 只要取足够大的 N 就能从上式得到解的长时间行为。实际上, 对于一个一般性的解, 其长时行为取决于矩阵 S 模长最大的本征值²。为了看清楚这个结论, 我们记 S 的 $2n$ 个本征值为 $\lambda_\alpha, \alpha = 1, 2, \dots, 2n$, 相应的本征矢量为 a^i_α , 不妨假设 $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_{2n}|$ 。现在将解 $\mathbf{x}(t)$ 在 $t = 0$ 时刻的初始值 $\mathbf{x}(0)$ 展开为 $2n$ 个本征矢量 a^i_α 的线性组合, 即 $x^i(0) = \sum_\alpha c_\alpha a^i_\alpha$, 则有

$$\mathbf{x}(NT) = S^N\mathbf{x}(0) \Rightarrow x^i(NT) = \sum_\alpha c_\alpha \lambda_\alpha^N a^i_\alpha. \quad (7.49)$$

很显然, 当 $N \rightarrow \infty$ 时, 必有

$$x^i(NT) = \sum_\alpha c_\alpha \lambda_\alpha^N a^i_\alpha \rightarrow c_1 \lambda_1^N a^i_1 + \dots, \quad (7.50)$$

从而矩阵 S 模长最大的本征值 λ_1 和相应的本征矢量决定了方程 (7.44) 解的长时间行为。

特别的, 如果模长最大的本征值 λ_1 满足 $|\lambda_1| > 1$, 则 $|\lambda_1^N|$ 在 $N \rightarrow \infty$ 时将随着 N 指数增长, 从而方程 (7.44) 的一般解将在长时间上随时间指数增长, 我们称这样的解为不稳定解, 称相应的系统为不稳定系统。反之, 如果 $|\lambda_1| \leq 1$, 则方程 (7.44) 的任何解都不会随着时间指数增长, 我们称这些解为稳定解, 称这样的系统为稳定系统。**所谓的参数共振, 就是指那些在参数的周期性变化下不稳定的系统。尤其我们要研究什么情况下一个没有发生参数共振的稳定系统, 在参数扰动之下产生参数共振, 变成不稳定系统的现象。**具体来说, 就是要研究在什么情形下, 一个由哈密顿量 (7.42) 描述的稳定系统, 在给参数 $B_{ij}(t)$ 加上某种保持周期 T 不变的小扰动后, 会产生参数共振变成不稳定系统。由于参数的保周期小扰动仅仅只能对矩阵 S 产生微扰, 不会使得 $|\lambda_1| < 1$ 突变到 $|\lambda_1| > 1$, 因此很显然这种使得稳定系统失稳的情形, 只可能发生在扰动最大模长本征值 $|\lambda_1| = 1$ 的系统时。

另一方面, 由于线性变换 (7.45) 同时是正则变换, 而正则变换必须保持辛结构, 即满足

$$\omega_{ij} \frac{\partial x'^i}{\partial x'^m} \frac{\partial x'^j}{\partial x'^n} = \omega_{mn}, \quad (7.51)$$

²为简单起见, 我们假设矩阵 S 可以对角化。

从而即可以得到

$$\omega_{ij} S_m^i S_n^j = \omega_{mn} \Leftrightarrow S^T J S = J, \quad (7.52)$$

满足这个式子的矩阵 S 就称作辛矩阵。由于辛矩阵 S 是 $x^i(t+T)$ 与 $x^i(t)$ 之间的正则变换的雅可比矩阵，根据第五章证明刘维尔定理时得到的附带结果可以知道，它的行列式必为 1，即

$$\det(S) = 1. \quad (7.53)$$

另外，根据第五章中矩阵 J 的定义式也容易直接算得

$$\det(J) = 1. \quad (7.54)$$

为了研究何时会发生参数共振，有必要研究一下辛矩阵的本征值集合（称为本征谱）有什么特点。为此我们考察辛矩阵 S 的特征多项式 $p(\lambda) = \det(S - \lambda \cdot 1)$ ，式中 1 表示单位矩阵。由 (7.52) 式以及 $J^2 = -1$ (根据定义直接验证) 可得， $S = -JS^{T-1}J$ ，从而

$$\begin{aligned} p(\lambda) &= \det(S - \lambda \cdot 1) = \det(-JS^{T-1}J + J\lambda J) \\ &= \det(-S^{T-1} + \lambda \cdot 1) = \det(-1 + \lambda S) \\ &= \lambda^{2n} \det(S - \frac{1}{\lambda} \cdot 1) = \lambda^{2n} p(1/\lambda). \end{aligned} \quad (7.55)$$

上述推导中应用了 $\det(S) = \det(J) = 1$ ，以及 $\det(A^T) = \det(A)$ 这些事实。由推导的结果可知，若 $p(\lambda) = 0$ 则必有 $p(1/\lambda) = 0$ ，反之亦然。这就说明， S 的本征值成对出现，若 λ 为本征值，则 $1/\lambda$ 也必定为本征值。另外， S 为一实矩阵，因此若 λ 为本征值，则 $\bar{\lambda}$ 也必为其本征值，这里 $\bar{\lambda}$ 表示 λ 的复数共轭。

因此，矩阵 S 的本征值必定四个一组地出现，分别为

$$\lambda, \bar{\lambda}, 1/\lambda, 1/\bar{\lambda}. \quad (7.56)$$

在复平面上，这四个本征值关于实轴和关于单位圆周均对称³，如图 (7.2) 所示。但是本征

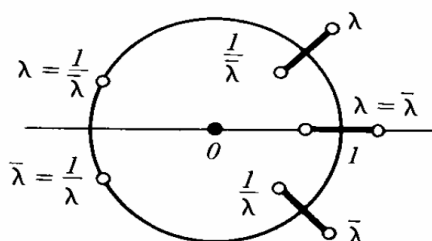


Figure 7.2: 辛矩阵的本征值分布。图片来自阿诺德《经典力学的数学方法》

³所谓关于单位圆周对称是指 λ 和 $1/\bar{\lambda}$ 成对出现。

值分布在单位圆周上的情形例外，这时候由于 $\lambda = 1/\bar{\lambda}, \bar{\lambda} = 1/\lambda$ ，所以本征值实际上是两个一对出现，而没有四个一组，这一对为 $\lambda, \bar{\lambda}$ 。同样，本征值分布在实轴上的情形也例外，这时候本征值也只成对，为 $\lambda, 1/\lambda$ 。具体分布情形如图 (7.2) 所示。

但是根据前面关于系统稳定性的讨论可以知道， S 的本征值中只要有一个不在单位圆周上，则系统必定不稳定。为了看清楚这一点，不妨设本征值 λ 不在单位圆周上，则很显然，本征值 λ 和 $1/\lambda$ 中必定有一个落在圆外，另一个落在圆内，落在圆外的那个本征值模长当然大于 1，从而这就说明，最大模长本征值的模长必定大于 1，从而系统必定不稳定。这也说明，对于稳定系统，其 S 的本征值必定全都分布在单位圆周之上。

定理：若矩阵 S 的 $2n$ 个本征值各不相同，而且都分布在单位圆周上，则系统必定稳定，而且在对参数的保持周期 T 不变的小扰动下也依然保持稳定。

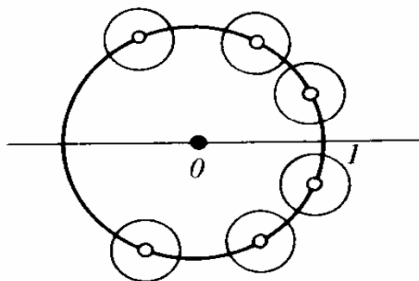


Figure 7.3: 稳定的本征值分布。图片来自阿诺德《经典力学的数学方法》

证明：首先，单位圆周上的本征值当然满足模长等于 1，从而系统必定稳定，真正需要证明的是它在对参数的扰动下依然保持稳定。具体证明如下，如图 (7.3) 所示，由于 $2n$ 个本征值各不相同，所以我们可以把它们分别包含在 $2n$ 个不相交的邻域内，且让这些邻域关于单位圆周和实轴均对称。现在，设想对系统的参数进行一个保周期 T 的微扰，从而相当于将矩阵 S 作一个微扰，变为 S' 。很显然，只要扰动足够小，则 S' 的本征值必定依然分布在那 $2n$ 个不相交邻域内。实际上 S' 的所有本征值必定依然分布在单位圆周上。因为否则，假设有一个本征值 λ' 不在单位圆周上，例如位于圆外，则根据上面关于辛矩阵本征值分布的讨论可以知道，必定还有一个成对的本征值 $1/\bar{\lambda}'$ 位于同一邻域的圆内部分。这样，这个邻域内扰动之前的一个本征值在扰动之后就变成了两个，这就意味着 S' 的不同本征值数大于 $2n$ 。但这是不可能的，因为 S' 依然为 $2n \times 2n$ 矩阵，它最多只能有 $2n$ 个不同本征值。这就证明了 S' 的所有本征值必定分布在单位圆周之上，从而对参数进行扰动之后系统依然保持稳定！这就完成了证明。

上面的定理说明，当我们扰动稳定系统的参数时，相应矩阵 S 的本征值只有在撞上另一个本征值时才可能成对地离开单位圆周 (如图 (7.4))，当然，这时与这一对本征值复共轭的本征值对也同时相撞并成对离开单位圆周。这时候就会使得原来稳定的系统产生参数共振，变成不稳定系统。

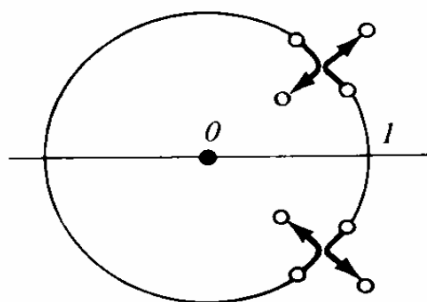


Figure 7.4: 本征值相撞离开单位圆周。图片来自阿诺德《经典力学的数学方法》

7.2.3 单自由度系统的参数共振

这一小节我们将对 $n = 1$ 的单自由度系统的参数共振进行更细致的分析。待分析的系统为一个频率参数周期性变化的谐振子，其哈密顿量为

$$H = \frac{1}{2}[p^2 + \omega^2(t)q^2], \quad (7.57)$$

其中角频率参数 $\omega(t)$ 满足 $\omega^2(t+T) = \omega^2(t)$, T 为参数变化的周期，我们记 $\omega_p = \frac{2\pi}{T}$ 为参数变化的角频率。很显然，这个系统的相空间为 2 维，相空间坐标 $\mathbf{x} = (q, p)^T$ ，因此矩阵 S 是一个 2×2 矩阵。系统的哈密顿正则方程为

$$\dot{q} = p, \quad \dot{p} = -\omega(t)q. \quad (7.58)$$

这个方程就是一个变频谐振子的微分方程。为了求出矩阵 S ，我们取两个线性无关的初始值 $\mathbf{x}_{(1)}(0) = (1, 0)^T$, $\mathbf{x}_{(2)}(0) = (0, 1)^T$ ，进而根据 (7.47) 式，矩阵 S 由下式给出

$$S = \begin{pmatrix} x_{(1)}^1(T) & x_{(2)}^1(T) \\ x_{(1)}^2(T) & x_{(2)}^2(T) \end{pmatrix}. \quad (7.59)$$

由于这里 S 为 2×2 矩阵，所以它只有两个本征值 λ_1, λ_2 ，根据上一小节的分析，这两个本征值必定构成一对。由于不能够凑成四个一组，所以这一对本征值的分布只有两种可能性，要么关于实轴对称地（即互为复共轭）分布在单位圆周上，要么关于单位圆周对称地分布在实轴上，如图 (7.5) 所示。当这一对本征值分布在实轴上时，我们有

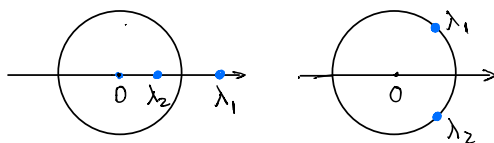


Figure 7.5: 单自由度系统 S 矩阵本征值分布。

$$|\text{Tr}(S)| = |\lambda_1 + \lambda_2| = \left| \lambda_1 + \frac{1}{\lambda_1} \right| > 2, \quad (7.60)$$

式中 $\text{Tr}(S)$ 表示矩阵 S 的迹。根据上一小节的一般性分析, 这时候系统不稳定, 有参数共振。当这一对本征值分布在单位圆周上时, 我们可以记 $\lambda_1 = e^{i\alpha}$, 则 $\lambda_2 = \bar{\lambda}_1 = e^{-i\alpha}$, 则

$$|\text{Tr}(S)| = |\lambda_1 + \lambda_2| = |e^{i\alpha} + e^{-i\alpha}| = 2|\cos \alpha| \leq 2, \quad (7.61)$$

根据上一小节的分析, 这时候系统是稳定的。而且, 当 $e^{i\alpha} \neq e^{-i\alpha}$ 时, 即 $|\text{Tr}(S)| < 2$ 时, 由于两个本征值不同, 所以根据上一小节的分析, 系统在参数的保持周期 T 不变的扰动下将依然保持稳定, 从而即使在扰动之下也不会产生参数共振。稳定系统在扰动之下失稳的情形发生在 $e^{i\alpha} = e^{-i\alpha}$ 时, 即发生在 $\lambda_1 = \lambda_2 = \pm 1$ 的两本征值碰撞时, 很显然这等价于

$$|\text{Tr}(S)| = 2. \quad (7.62)$$

失稳就发生在将这种临界系统的本征值扰动到脱离单位圆周时。

为了分析得更具体一些, 我们不妨假设 (7.57) 式给出的哈密顿量中的 $\omega^2(t)$ 为

$$\omega^2(t) = \omega_0^2 [1 + \varepsilon \cos(\omega_p t)], \quad (7.63)$$

式中 $\omega_p = \frac{2\pi}{T}$ 为 $\omega^2(t)$ 变化的角频率。并且假设我们仅仅只考察通过变动 ω_0 和 ε 的方式扰动 $\omega^2(t)$ 。我们可以在 $\omega_0 - \varepsilon$ 平面上示意性地画出 $|\text{Tr}(S)| < 2$ 和 $|\text{Tr}(S)| > 2$ 的区域, 也就是系统的稳定区和不稳定区 (也就是参数共振的区域), 如图 (7.6) 所示。两个区域的交界线就表示 $|\text{Tr}(S)| = 2$ 的临界系统。

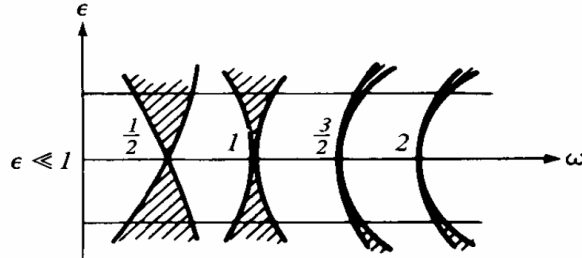


Figure 7.6: $\omega_0 - \varepsilon$ 平面上的稳定区和参数共振区。阴影部分表示参数共振区。图中横坐标实际上应该是 ω_0/ω_p , ω_p 是固定的。图片来自阿诺德《经典力学的数学方法》。

从图 (7.6) 中我们看到, 整个水平轴都是一些被临界点隔开的稳定区域。这实际上是怎么确定的: 首先, 我们在 (7.63) 式中取 $\varepsilon = 0$ (对应水平轴)。这时候要求解的哈密顿正则方程就是

$$\dot{q} = p, \quad \dot{p} = -\omega_0 q. \quad (7.64)$$

其通解为

$$\begin{aligned} q(t) &= c_1 \cos(\omega_0 t) + c_2 \sin(\omega_0 t) \\ p(t) &= -c_1 \omega_0 \sin(\omega_0 t) + c_2 \omega_0 \cos(\omega_0 t) \end{aligned} \quad (7.65)$$

对于初值 $\mathbf{x}_{(1)}(0) = (1, 0)^T$ ，容易解出

$$\mathbf{x}_{(1)}(T) = (\cos(2\pi\omega_0/\omega_p), -\omega_0 \sin(2\pi\omega_0/\omega_p))^T. \quad (7.66)$$

对于初值 $\mathbf{x}_{(2)}(0) = (0, 1)^T$ ，容易解出

$$\mathbf{x}_{(2)}(T) = (\sin(2\pi\omega_0/\omega_p)/\omega_0, \cos(2\pi\omega_0/\omega_p))^T. \quad (7.67)$$

从而矩阵 S 为

$$S = \begin{pmatrix} \cos(2\pi\omega_0/\omega_p) & \frac{1}{\omega_0} \sin(2\pi\omega_0/\omega_p) \\ -\omega_0 \sin(2\pi\omega_0/\omega_p) & \cos(2\pi\omega_0/\omega_p) \end{pmatrix}. \quad (7.68)$$

进而可以得到


$$|\mathrm{Tr}(S)| = 2|\cos(2\pi\omega_0/\omega_p)| \leq 2. \quad (7.69)$$

可见， $\omega_0 - \varepsilon$ 平面的水平轴上只有稳定区域，而 $|\mathrm{Tr}(S)| = 2$ 的临界点满足

$$\omega_0/\omega_p = \frac{k}{2}. \quad (7.70)$$

式中 k 为正整数。特别的， $\omega_p = 2\omega_0$ 为最简单的临界点，这时候只要稍微扰动 ε 就会产生参数共振。这正好解释了本节一开始提出的问题，即，为什么荡秋千时要以 $2\omega_0$ 的频率变化摆长才能产生参数共振。

关于参数共振，我们就讲到这里，更多的计算例子请参阅阿诺德《经典力学的数学方法》。



8. 刚体的运动

刚体是非相对论力学中的一种理想模型，它是一个质点系，但是其中任何两个质点之间的距离始终恒定不变，即

$$|\mathbf{x}_i - \mathbf{x}_j| = \text{constant}, \quad (8.1)$$

式中 \mathbf{x}_i 表示组成刚体的第 i 个质点的位置矢量。在实际应用中，我们常常认为一些刚体的质量是连续分布的，从而可以引入单位体积的质量分布，即质量密度 $\rho(\mathbf{x})$ ，并进而将对离散的各质点的求和改成连续的积分，比如将 $\sum_i m_i$ 替换成 $\int d\tau \rho(\mathbf{x})$ (这里 $d\tau$ 代表三维体积元 $dx dy dz$)。

但是，刚体只是非相对论力学中的理想模型，利用相对论的简单知识可以证明，绝对符合定义的刚体是不存在的！这是因为，对于刚体来说，当我们在它的一端踢一脚，由于距离绝对不变，它的另一端立即就得有反应，但是，根据狭义相对论，任何信号的传递速度都有限，因此一端的影响要传递到另一端肯定需要时间，即是说，另一端不可能立即有反应。从而在一端被踢动而另一端还没来得及反应的这段时间内，两端的距离就变了，因此根据狭义相对论，刚体是不存在的！但是，在非相对论力学的近似下，我们可以近似认为信号的传播速度为无穷大，从而可以近似将很多物体看成刚体。当然，大家以后学广义相对论就能发现，在广义相对论中可以有流体模型，但刚体模型是没有的，道理就是刚才讲过的道理。

我们可以将刚体的运动看成是由 3 个独立分量的平动叠加上绕某个点 O (比如质心) 的转动。对于平动的分析来说，完全可以把整个刚体看成单个质点而不需要引入任何新的知识，需要仔细讨论的是刚体绕某个点 O 的转动。忽略平动以后我们可以认为 O 是固定

的,从而这就是刚体的定点转动。我们将会看到,描述刚体定点转动也需要 3 个变量,因此加上 3 个平动自由度,刚体的总自由度数目为 6。

(说明一下,本章不再区分上下指标,但依然使用求和约定,即是说,只要表达式中出现了两个重复指标我们就默认对它求和,不再要求这两个指标分别为上下指标。)

8.1 运动学

8.1.1 刚体转动的位形空间

为了考察刚体绕某个固定点 O 的转动,我们以 O 为坐标原点,引入两种坐标系,其一是固定在空间的空间坐标系,记它的三个坐标分量的单位矢量为 $\tilde{\mathbf{e}}_a, a = 1, 2, 3$ 。另一种坐标系是固定在刚体上随着刚体一起转动的刚体坐标系,其三个坐标分量的单位矢量记为 $\mathbf{e}_a, a = 1, 2, 3$, 如图 (8.1) 所示。很显然,这些基矢量满足如下关系

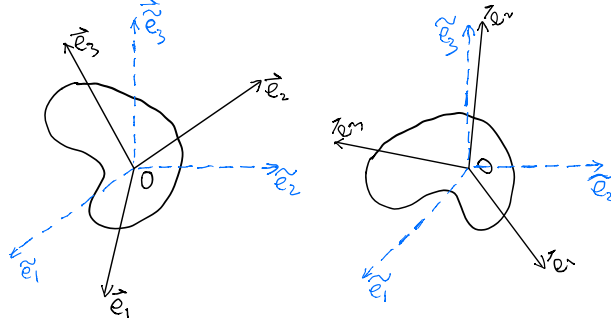


Figure 8.1: 空间坐标系和刚体坐标系。

$$\tilde{\mathbf{e}}_a \cdot \tilde{\mathbf{e}}_b = \mathbf{e}_a \cdot \mathbf{e}_b = \delta_{ab}. \quad (8.2)$$

由于刚体坐标系是活动坐标系,所以其矢量基 $\{\mathbf{e}_a\}$ 实际上依赖于时间,有时候记为 $\{\mathbf{e}_a(t)\}$ 。

很明显,刚体的取向状态(称作刚体的位形)完全由刚体坐标系矢量基的取向 $\{\mathbf{e}_a\}$ 确定。从而刚体的**位形空间**(记为 \mathcal{C})就是刚体坐标系所有可能取向所构成的空间。

给定刚体坐标系的一个取向 $\{\mathbf{e}_a\}$,我们可以把它在空间坐标系 $\{\tilde{\mathbf{e}}_a\}$ 中作矢量展开,即(注意,式中有默认的求和)

$$\mathbf{e}_b = \tilde{\mathbf{e}}_a R_{ab}. \quad (8.3)$$

很显然,展开系数 R_{ab} 构成一个 3×3 的矩阵,相应的矩阵形式可以记为 R 。容易证明 R 必定为正交矩阵,这是因为

$$\delta_{ab} = \mathbf{e}_a \cdot \mathbf{e}_b = \tilde{\mathbf{e}}_c R_{ca} \cdot \tilde{\mathbf{e}}_d R_{db} = R_{ca} R_{cb} = (R^T R)_{ab}, \quad (8.4)$$

这就说明 $R^T R$ 为单位矩阵,从而 R 为 3×3 的正交矩阵。而且,由于刚体坐标系和空间坐标系均为右手系,因此其变换矩阵的雅可比行列式必定为 1,这就说明 $\det(R) = 1$,从而排除了 $\det(R) = -1$ 的情形。综上可知,刚体坐标系的所有可能取向和行列式等于 1 的正交

矩阵一一对应。从而刚体的位形空间就等价于所有行列式等于 1 的正交矩阵的集合。通常称这种正交矩阵为特殊正交矩阵 (Special Orthogonal Matrix) 并记它们的集合为 $SO(3)$ 。从而即有, 刚体的位形空间 \mathcal{C} 等于 $SO(3)$, 即

$$\mathcal{C} = SO(3). \quad (8.5)$$

3×3 的矩阵有 9 个矩阵元, 但是考虑到正交矩阵的条件 $R^T R = 1$, 因此这 9 个矩阵元要满足 6 个独立的限制方程, 从而真正独立的参数只有 3 个。即是说, 集合 $SO(3)$ 可以用 3 个参数来参数化, 从而刚体的位形空间 \mathcal{C} 也可以用 3 个广义坐标来参数化, 怎么选这 3 个参数是后面会进一步讨论的内容。正因为 \mathcal{C} 是 3 维的, 所以前面我们说刚体有 3 个转动自由度!

8.1.2 角速度矢量

由于刚体坐标系 $\{\mathbf{e}_a\}$ 随着刚体一起运动, 这就在位形空间 \mathcal{C} 中画出一条路径 $\{\mathbf{e}_a(t)\}$, 因此当然可以计算 $\dot{\mathbf{e}}_a(t)$ 。我们将 $\dot{\mathbf{e}}_a$ 在刚体系 $\{\mathbf{e}_a\}$ 中展开, 即设

$$\dot{\mathbf{e}}_a = \omega_{ac} \mathbf{e}_c, \quad (8.6)$$

式中 ω_{ac} 为展开系数。可以证明, 它必然满足

$$\omega_{ab} + \omega_{ba} = 0, \quad (8.7)$$

即 ω_{ab} 关于两个指标反对称。证明如下: 首先, 我们将方程 (8.6) 点乘 \mathbf{e}_b , 得到

$$\mathbf{e}_b \cdot \dot{\mathbf{e}}_a = \omega_{ab}. \quad (8.8)$$

其次, 我们将方程 $\mathbf{e}_a \cdot \mathbf{e}_b = \delta_{ab}$ 对时间求导, 注意到右边 δ_{ab} 是常数, 对时间求导等于零, 从而即有

$$\dot{\mathbf{e}}_a \cdot \mathbf{e}_b + \mathbf{e}_a \cdot \dot{\mathbf{e}}_b = 0 \Rightarrow \omega_{ab} + \omega_{ba} = 0. \quad (8.9)$$

进一步利用 3 维空间特殊的列维-西维塔符号 ${}^1\epsilon_{abc}$, 即可以定义角速度矢量, 其在刚体坐标系中的分量 ω_a 为

$$\omega_a = \frac{1}{2} \epsilon_{abc} \omega_{bc} \Leftrightarrow \omega_{bc} = \epsilon_{abc} \omega_a, \quad (8.10)$$

上面我们利用了列维-西维塔符号的恒等式 $\epsilon_{abc} \epsilon_{ade} = \delta_{bd} \delta_{ce} - \delta_{be} \delta_{cd}$ 。进而就可以将方程 (8.6) 重写为

$$\dot{\mathbf{e}}_a = \omega_{ac} \mathbf{e}_c = \epsilon_{acb} \omega_b \mathbf{e}_c = -\epsilon_{abc} \omega_b \mathbf{e}_c. \quad (8.11)$$

或者也可以将这个式子写成

$$\dot{\mathbf{e}}_a = \vec{\omega} \times \mathbf{e}_a, \quad (8.12)$$

¹参见第 2 章的相关部分

式中 $\vec{\omega}$ 就是角速度矢量，其定义为 $\vec{\omega} = \omega_a \mathbf{e}_a$ 。在 (8.12) 式的推导过程中，我们也利用了关系式 $\mathbf{e}_a \times \mathbf{e}_b = \varepsilon_{abc} \mathbf{e}_c$ 。由 (8.12) 式可以知道，刚体上一个位置矢量为 $\mathbf{x} = x_a \mathbf{e}_a$ 的质点的速度为，

$$\dot{\mathbf{x}} = \vec{\omega} \times \mathbf{x}, \quad (8.13)$$

为了得出这个结果我们只需注意这个质点在刚体坐标系中的坐标分量 x_a 是不变的。

以上结果也可以用更直观的方式推导出来。为此我们假设刚体在 δt 的无穷小时间内绕着某个瞬时转轴 \vec{n} 转过一个无穷小角度 $\delta\phi$ ，记 $\vec{n}\delta\phi = \vec{\delta\phi}$ 。如图 (8.2) 所示，设刚体上的矢量 \mathbf{x} 转到了 \mathbf{x}' ，改变量为 $\delta\mathbf{x} = \mathbf{x}' - \mathbf{x}$ 。从图中很容易看出 $|\delta\mathbf{x}| = \delta\phi|\mathbf{x}|\sin\theta$ ，也容易看

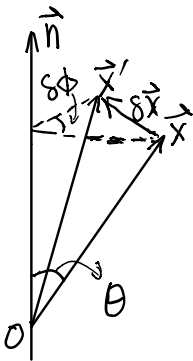


Figure 8.2: 绕瞬时转轴旋转 $\delta\phi$ 角。

出 $\delta\mathbf{x}$ 的方向正好是 \vec{n} 叉乘 \mathbf{x} 的方向，因此即有

$$\delta\mathbf{x} = \vec{\delta\phi} \times \mathbf{x}. \quad (8.14)$$

将这个式子除以 δt 即得 (8.13) 式，其中

$$\vec{\omega} = \frac{\vec{\delta\phi}}{\delta t} = \frac{\delta\phi}{\delta t} \vec{n}. \quad (8.15)$$

这正给出了角速度矢量的直观含义。值得强调的是，转轴 \vec{n} 通常不固定，而是会随着时间改变。

8.2 惯量张量

8.2.1 惯量张量

这一节我们要介绍刚体转动最重要的概念之一，叫做惯量张量。它其实是大家在力学中学过的转动惯量概念的推广，不过在力学中我们只分析转轴固定的定轴转动，所谓转动惯量描述的就是刚体绕这个固定转轴的转动惯性。但现在我们研究的转动其转轴并不固定，从而描写转动惯性的量就要更复杂，要推广成所谓的惯量张量。

为了引入惯量张量，让我们计算一下刚体转动时的动能，

$$\begin{aligned}
 T &= \sum_i \frac{1}{2} m_i \dot{\mathbf{x}}_i^2 = \sum_i \frac{1}{2} m_i (\vec{\omega} \times \mathbf{x}_i)^2 \\
 &= \sum_i \frac{1}{2} m_i [(\mathbf{x}_i^2)(\vec{\omega}^2) - (\mathbf{x}_i \cdot \vec{\omega})^2] \\
 &= \frac{1}{2} I_{ab} \omega_a \omega_b.
 \end{aligned} \tag{8.16}$$

式中 I_{ab} 为

$$I_{ab} = \sum_i m_i [(\mathbf{x}_i^2) \delta_{ab} - (\mathbf{x}_i)_a (\mathbf{x}_i)_b], \tag{8.17}$$

其中 $(\mathbf{x})_a$ 表示矢量 \mathbf{x} 在刚体坐标中的第 a 分量，即 $(\mathbf{x})_a = x_a$ 。很显然， I_{ab} 关于两个指标是对称的，它可以看成是一个对称矩阵 \mathbf{I} 的矩阵元。人们通常把 I_{ab} 称作**惯量张量的分量形式**，简称为惯量张量。对于刚体质量连续分布的情况，惯量张量 \mathbf{I} 为，

$$\mathbf{I} = \int d\tau \rho(\mathbf{x}) \begin{pmatrix} y^2 + z^2 & -xy & -xz \\ -xy & x^2 + z^2 & -yz \\ -xz & -yz & x^2 + y^2 \end{pmatrix}. \tag{8.18}$$

值得强调的是，惯量张量是在刚体坐标系中定义的，并且从上面的表达式很容易看出， $I_{xx} = I_{11}$ 就是绕 \mathbf{e}_1 轴（即 x 轴）定轴转动的转动惯量， $I_{yy} = I_{22}$ 和 $I_{zz} = I_{33}$ 分别是绕 \mathbf{e}_2 轴和 \mathbf{e}_3 轴的转动惯量，可见惯量张量的确是转动惯量的推广。

但是刚体坐标系的选取有任意性，我们当然可以将刚体坐标系重新选取为 $\mathbf{e}'_a = O_{ab} \mathbf{e}_b$ ，这里 O 为正交矩阵。注意到角速度矢量 $\vec{\omega}$ 应该与刚体系的选取无关，即 $\vec{\omega} = \mathbf{e}'_a \omega'_a = \mathbf{e}_a \omega_a$ ，从而新刚体坐标下的角速度分量必为 $\omega'_a = O_{ab} \omega_b$ 。进一步，刚体转动动能 T 当然也与刚体坐标的选取无关，从而即有 $I'_{ab} \omega'_a \omega'_b = I_{ab} \omega_a \omega_b$ ，由此可知新坐标下惯量张量的分量 I'_{ab} 必定为 $I'_{ab} = O_{ac} O_{bd} I_{cd}$ 。即是说，在坐标系的正交变换下，惯量张量按照如下方式变换

$$I_{ab} \rightarrow I'_{ab} = O_{ac} O_{bd} I_{cd}, \tag{8.19}$$

或者我们也可以用矩阵形式将这个变换关系写成

$$\mathbf{I} \rightarrow \mathbf{I}' = O \mathbf{I} O^T. \tag{8.20}$$

在线性代数中这种变换也称作矩阵 \mathbf{I} 的正交变换。

而 \mathbf{I} 是一个对称矩阵，根据线性代数的知识可以知道，它必定可以正交变换为一个对角矩阵，其对角元就是 \mathbf{I} 的本征值。换言之，我们总可以选择一个合适的刚体坐标系，在这个坐标系中，惯量张量表现为对角的形式，即

$$\mathbf{I} = \begin{pmatrix} I_1 & & \\ & I_2 & \\ & & I_3 \end{pmatrix}. \tag{8.21}$$

这样的刚体坐标系就称作惯量主轴坐标系，它的三个正交的坐标轴就是惯量主轴，惯量张量在惯量主轴坐标系中的对角元 I_1, I_2, I_3 就称之为主转动惯量。对于一个有对称性的刚体，其对称轴通常是惯量主轴。另外，由于转动动能 $\frac{1}{2}I_{ab}\omega_a\omega_b$ 总大于等于零，所以矩阵 \mathbf{I} 是半正定的，从而主转动惯量必定大于等于零，即 $I_1 \geq 0, I_2 \geq 0, I_3 \geq 0$ 。

举例：薄圆盘

考察一个半径为 R 厚度可以忽略的薄圆盘，假设它绕其中心转动。根据对称性我们可以找到圆盘的 3 个惯量主轴，分别为沿着盘面的 $\mathbf{e}_1, \mathbf{e}_2$ 轴，以及垂直盘面的 \mathbf{e}_3 轴，如图 (8.3) 所示。根据对称性很明显有 $I_1 = I_2$ ，即

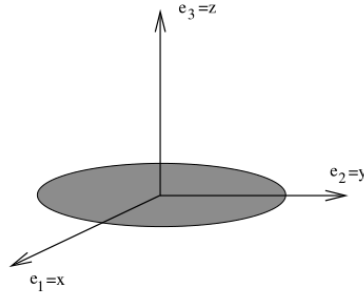


Figure 8.3: 薄圆盘的惯量主轴。图片来自 David Tong, 《classical dynamics》

$$I_1 = I_2 = \int d\tau \rho x^2 = \int d\tau \rho y^2 = \frac{1}{2} \int d\tau \rho [x^2 + y^2] = \frac{1}{2} I_3. \quad (8.22)$$

式中 $\rho = m/(\pi R^2)$ 为质量面密度， $d\tau$ 为面积元。很容易算出上面的面积分，进而得

$$I_3 = \frac{1}{2} m R^2 = 2I_1 = 2I_2. \quad (8.23)$$

刚体的角动量

下面来讨论一下如何用惯量张量和角速度分量来表达出刚体转动的角动量 \mathbf{J} 。为此进行如下推导

$$\begin{aligned} \mathbf{J} &= \sum_i m_i \mathbf{x}_i \times \dot{\mathbf{x}}_i = \sum_i m_i \mathbf{x}_i \times (\vec{\omega} \times \mathbf{x}_i) \\ &= \sum_i m_i [\mathbf{x}_i^2 \vec{\omega} - \mathbf{x}_i (\mathbf{x}_i \cdot \vec{\omega})]. \end{aligned} \quad (8.24)$$

我们可以在刚体坐标系中考察 \mathbf{J} 的分量形式 $\mathbf{J} = J_a \mathbf{e}_a$ ，则从上面的表达式容易看出

$$J_a = I_{ab} \omega_b. \quad (8.25)$$

从这个表达式可知，对于刚体定点转动而言，一般来说，其角速度矢量的方向和角动量的方向并不相同，而是相差一个线性变换。特别的，在惯量主轴坐标系中，我们有

$$J_1 = I_1 \omega_1, \quad J_2 = I_2 \omega_2, \quad J_3 = I_3 \omega_3. \quad (8.26)$$

8.2.2 平行轴定理

惯量张量的定义依赖于坐标原点，也就是固定点 O 。 O 点原则上可以任意，但有一个点比较特殊，那就是质心 C 。假设我们先在质心 C 处建立质心刚体坐标系，然后再将这个刚体坐标系平移到 O 点，我们想问，刚体绕 C 点转动和绕 O 点转动的惯量张量的关系是什么？不妨设 O 相对于质心的位置矢量为 \mathbf{a} ，设质点在质心系中的位置矢量为 \mathbf{x}' ，在平移后的 O 点坐标系中的位置矢量为 \mathbf{x} ，很显然 $\mathbf{x} = \mathbf{x}' - \mathbf{a}$ 。另外，不妨记绕质心 C 转动的惯量张量为 $I_{ab}^{(c)}$ ，记绕 O 点转动的惯量张量为 I_{ab} ，则有

$$\begin{aligned} I_{ab} &= \sum_i m_i [(\mathbf{x}_i)^2 \delta_{ab} - (\mathbf{x}_i)_a (\mathbf{x}_i)_b] = \sum_i m_i [(\mathbf{x}'_i - \mathbf{a})^2 \delta_{ab} - (\mathbf{x}'_i - \mathbf{a})_a (\mathbf{x}'_i - \mathbf{a})_b] \\ &= \sum_i m_i [(\mathbf{x}'_i)^2 \delta_{ab} - (\mathbf{x}'_i)_a (\mathbf{x}'_i)_b] + \sum_i m_i [(\mathbf{a})^2 \delta_{ab} - (\mathbf{a})_a (\mathbf{a})_b] \\ &\quad - 2\mathbf{a} \cdot \sum_i m_i \mathbf{x}'_i \delta_{ab} + \sum_i m_i [(\mathbf{x}'_i)_a (\mathbf{a})_b + (\mathbf{x}'_i)_b (\mathbf{a})_a] \\ &= I_{ab}^{(c)} + m [(\mathbf{a})^2 \delta_{ab} - (\mathbf{a})_a (\mathbf{a})_b]. \end{aligned}$$

上面的推导过程中利用了质心系的定义 $\sum_i m_i \mathbf{x}'_i = 0$ ，式中的 $m = \sum_i m_i$ 代表刚体的总质量。不妨将最后的结果重写一遍

$$I_{ab} = I_{ab}^{(c)} + m [(\mathbf{a})^2 \delta_{ab} - (\mathbf{a})_a (\mathbf{a})_b]. \quad (8.27)$$

这个结果就是所谓的平行轴定理。值得注意的是，平行轴定理中多出来的项 $m [(\mathbf{a})^2 \delta_{ab} - (\mathbf{a})_a (\mathbf{a})_b]$ 正好是刚体的质量全部集中在质心时绕 O 点的惯量张量。

8.2.3 同时考虑平动和转动

前面说过，刚体的一般运动可以分解成质心的平动叠加上绕质心的转动。但是前面仅仅只考虑了转动，为了同时考虑转动和平动，我们假定刚体坐标系是一个质心系，它的坐标原点（即质心）相对于空间坐标系在运动。记质心在空间坐标系中的位置矢量为 \mathbf{x}_c ，记刚体上的质点在刚体坐标系中的位置矢量为 \mathbf{x}'_i ，而在固定的空间坐标系中的位置矢量为 \mathbf{x}_i ，很显然

$$\mathbf{x}_i = \mathbf{x}'_i + \mathbf{x}_c. \quad (8.28)$$

由于刚体绕质心的转动，我们有

$$\dot{\mathbf{x}}'_i = \vec{\omega} \times \mathbf{x}'_i, \quad (8.29)$$

因此刚体坐标系中的一切分析都和前面一样。

下面我们来计算一下同时考虑质心平动和刚体绕质心转动时刚体的动能，

$$\begin{aligned} T &= \sum_i \frac{1}{2} m_i \dot{\mathbf{x}}_i^2 = \sum_i \frac{1}{2} m_i (\dot{\mathbf{x}}'_i + \dot{\mathbf{x}}_c)^2 \\ &= \sum_i \frac{1}{2} m_i (\dot{\mathbf{x}}'_i)^2 + \sum_i m_i \dot{\mathbf{x}}'_i \cdot \dot{\mathbf{x}}_c + \frac{1}{2} m \dot{\mathbf{x}}_c^2 \\ &= \frac{1}{2} I_{ab} \omega_a \omega_b + \frac{1}{2} m \dot{\mathbf{x}}_c^2. \end{aligned} \quad (8.30)$$

式中 $m = \sum_i m_i$, 上面的推导过程中利用了 $\sum_i m_i \mathbf{x}'_i = d(\sum_i m_i \mathbf{x}'_i)/dt = 0$ (因为对于质心系 $\sum_i m_i \mathbf{x}'_i = 0$)。由此可见, 刚体的总动能等于质心平动动能加上刚体绕质心转动的动能。因此分析刚体的一般运动时, 我们可以将质心的平动与绕质心的转动独立分析。

8.3 欧拉陀螺

现在我们来研究一个不受外力矩作用的刚体的转动, 它可以是一个绕着固定点自由转动的刚体, 甚至也可以是一个悬挂在太空中的星星。在后面这种情况中, 假设我们不关心星星质心的运动, 仅仅关心星星绕质心的自由转动。欧拉最早研究了这样的自由转动, 所以人们也常常称这种自由转动的刚体为欧拉陀螺。由于不受外力矩, 所以欧拉陀螺的空间角动量是守恒的, 即满足

$$\frac{d\mathbf{J}}{dt} = 0. \quad (8.31)$$

现在, 将 \mathbf{J} 在刚体坐标系中表达为 $\mathbf{J} = J_a \mathbf{e}_a$, 则有

$$\begin{aligned} \frac{dJ_a}{dt} \mathbf{e}_a + J_a \frac{d\mathbf{e}_a}{dt} &= 0 \\ \Rightarrow \frac{dJ_a}{dt} \mathbf{e}_a + J_a \vec{\omega} \times \mathbf{e}_a &= 0. \end{aligned} \quad (8.32)$$

现在, 我们取惯量主轴坐标系, 从而 $J_1 = I_1 \omega_1, J_2 = I_2 \omega_2, J_3 = I_3 \omega_3$, 则可以算得上面方程的分量形式

$$\begin{aligned} I_1 \dot{\omega}_1 + \omega_2 \omega_3 (I_3 - I_2) &= 0 \\ I_2 \dot{\omega}_2 + \omega_3 \omega_1 (I_1 - I_3) &= 0 \\ I_3 \dot{\omega}_3 + \omega_1 \omega_2 (I_2 - I_1) &= 0. \end{aligned} \quad (8.33)$$

这一组方程就是所谓的欧拉方程。

顺带说一下, 欧拉方程虽然常常用来研究自由转动的欧拉陀螺, 但它也可以推广到受力矩作用的刚体, 这时候方程 (8.33) 的右边就不再为零, 而分别是力矩的三个分量。

欧拉方程也可以在哈密顿力学框架下推导出来, 为此只需要注意两点: 其一, 角动量各分量满足如下代数关系

$$[J_a, J_b] = -\varepsilon_{abc} J_c. \quad (8.34)$$

注意这个代数关系和前面第五章讲泊松括号时作为例子推导出来的角动量代数关系差一个负号。这是因为, 第五章的那个代数关系各分量是相对于空间固定坐标系而言的, 而现在的各分量是相对于转动的刚体坐标系的, 即现在 $J_a = \mathbf{e}_a \cdot \mathbf{J}$, 因此也需要考虑到矢量 \mathbf{e}_a 本身是变量。最后再利用第五章中的标准角动量代数, 推导出来的就是 (8.34) 式, 推导虽不算难, 但过程稍微有些长, 我们留给感兴趣的读者自行探索²。其二, 需要注意到, 在惯

²值得说明的是, 角动量代数完全是由空间旋转的数学性质决定的, 数学上只允许这两种形式, 就是第五章的形式以及这里多出一个负号的形式。

量主轴坐标系中，欧拉陀螺的动能（从而也就是哈密顿量）为，

$$H = \sum_{a=1}^3 \frac{J_a^2}{2I_a}. \quad (8.35)$$

利用这两点，并代入物理量 J_a 的哈密顿运动方程 $\dot{J}_a = [J_a, H]$ ，就可以算得比如 $\dot{J}_1 = -J_2 J_3 \left(\frac{1}{I_2} - \frac{1}{I_3} \right)$ ，代入 $J_1 = I_1 \omega_1, J_2 = I_2 \omega_2, J_3 = I_3 \omega_3$ 就可以得到前面的欧拉方程。以上的推导说明，**欧拉陀螺是一个哈密顿系统**。

但是，欧拉陀螺作为哈密顿系统是不同寻常的，因为整个系统的基本力学变量只有 $J_a, a = 1, 2, 3$ ，这 3 个基本力学变量当然要参数化这个哈密顿系统的相空间，但这就说明，欧拉陀螺的相空间是 3 维的！我们知道，通常的相空间由于可以分解成正则坐标和正则动量，所以一定是 $2n$ 维的，因此欧拉陀螺不是通常的哈密顿系统。为了把它变成一个通常的哈密顿系统，我们注意到 $\mathbf{J}^2 = J_1^2 + J_2^2 + J_3^2$ 与哈密顿量 H 泊松对易，即

$$[\mathbf{J}^2, H] = 0. \quad (8.36)$$

从而 \mathbf{J}^2 是这个哈密顿系统的守恒量，如此一来我们就可以把系统限制在 \mathbf{J}^2 为常数的相空间球面 S^2 上来考察问题，

$$\mathbf{J}^2 = J_1^2 + J_2^2 + J_3^2 = C \geq 0. \quad (8.37)$$

我们可以把这个相空间球面 S^2 看成限制以后的相空间，很显然它是 2 维的，因此在 S^2 上就能构成一个通常的哈密顿系统！不仅如此，这个限制以后的哈密顿系统还是一个可积系统，因为作为一个相空间为 2 维的哈密顿系统，它有一个守恒量，即 H ，所以当然是可积系统。

上述可积系统的相空间不变环面必定是一些闭合的相空间轨道，实际上，这些相空间轨道可以由 $\sum_{a=1}^3 \frac{J_a^2}{2I_a} = E$ 的椭球面和上述 S^2 的相交线决定，即为 (J_1, J_2, J_3) 空间中下面两个曲面的交线，

$$\begin{aligned} J_1^2 + J_2^2 + J_3^2 &= C \geq 0. \\ \frac{J_1^2}{2EI_1} + \frac{J_2^2}{2EI_2} + \frac{J_3^2}{2EI_3} &= 1. \end{aligned} \quad (8.38)$$

假设 $I_1 < I_2 < I_3$ ，那么这些相空间轨道如图 (8.4) 所示。从图中很容易看出来，在 J_1 和 J_3 轴附近，相空间轨道为环绕坐标轴的圆周。这是因为， J_1, J_3 轴分别为椭球面的短轴和长轴，所以它和球面的交线当然为简单的圆周（请想象一下）。也即是说，如果欧拉陀螺绕着 \mathbf{e}_1 或者 \mathbf{e}_3 轴转动，那这种转动将是稳定的，因为扰动它的话只相当于将相轨道扰动到 J_1 或 J_3 轴附近的小圆周上，陀螺的转动大体上依然绕着 \mathbf{e}_1 或 \mathbf{e}_3 轴。但是，从图 (8.4) 中很容易看出， J_2 轴附近的相轨道完全不同，绕着 \mathbf{e}_2 轴的转动完全不稳定！ \mathbf{e}_2 轴就是转动惯量 I_2 大小居中的轴，所以结论是，欧拉陀螺绕着转动惯量大小居中的主轴转动是不稳定的！这个结论通常称作**网球拍定理** (tennis racket theorem)。网球拍定理的直观演示非常有趣，请读者在知乎上搜索“网球拍定理”，找一些相关的演示视频来看。

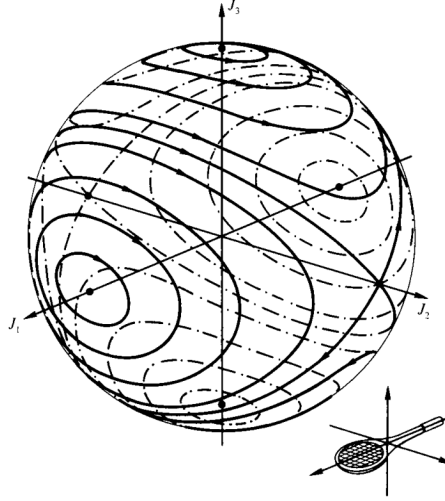


Figure 8.4: 欧拉陀螺的相空间轨道。图片来自 Tom W.B. Kibble Frank H. Berkshire, Classical Mechanics

特别的，如果所考察的欧拉陀螺是一个对称陀螺，有绕 \mathbf{e}_3 轴的旋转不变性，从而 $I_1 = I_2 \neq I_3$ 。那么相应的相空间轨道也将绕着 J_3 轴旋转不变，如图 (8.5) 所示。这个结论也很容易通过直接求解欧拉方程来证明，求解并不难，我们留给读者作练习。最后，**请注意**，以上所有相轨道描述的运动都是相对于刚体坐标系的，而不是相对于空间坐标系的。

另一种证明网球拍定理的方法

另一种证明网球拍定理的方法是直接分析欧拉方程。同样，我们假设 $I_1 < I_2 < I_3$ ，假设初始时刚体绕着 \mathbf{e}_2 轴转动，即

$$\omega_2 = \Omega, \quad \omega_1 = \omega_3 = 0. \quad (8.39)$$

下面假设加上一个小扰动，扰动后

$$\omega_2 = \Omega + \varepsilon_2, \quad \omega_1 = \varepsilon_1, \quad \omega_3 = \varepsilon_3, \quad (8.40)$$

式中 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 为随时间演化的小量。将上面的 ω_a 代入欧拉方程 (8.33)，并保留到一阶小量（即忽略 ε^2 阶和更高阶小量），即有

$$I_1 \dot{\varepsilon}_1 + \Omega \varepsilon_3 (I_3 - I_2) = 0$$

$$I_2 \dot{\varepsilon}_2 = 0$$

$$I_3 \dot{\varepsilon}_3 + \varepsilon_1 \Omega (I_2 - I_1) = 0.$$

从上面方程中容易得到

$$\begin{aligned} \ddot{\varepsilon}_3 + \frac{\Omega^2}{I_1 I_3} (I_2 - I_3) (I_2 - I_1) \varepsilon_3 &= 0 \\ \Rightarrow \ddot{\varepsilon}_3 + A \varepsilon_3 &= 0, \end{aligned} \quad (8.41)$$

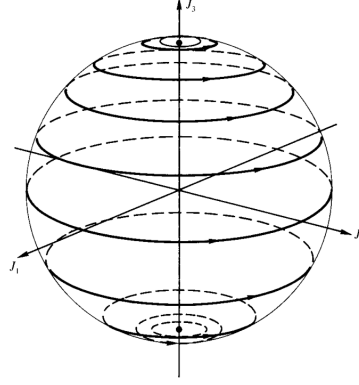


Figure 8.5: 对称的欧拉陀螺的相空间轨道。图片来自 Tom W.B. Kibble Frank H. Berkshire, Classical Mechanics

式中 $A = \frac{\Omega^2}{I_1 I_3} (I_2 - I_3)(I_2 - I_1)$ 。很显然，当 $A > 0$ 时，这是一个小振动方程，描写的是一个小小振动，从而系统稳定。但是，当 $A < 0$ 时，这个方程的通解是指数增长的，从而不稳定。而按照我们的假设， $I_1 < I_2 < I_3$ ，很显然 $A < 0$ ，这就证明了绕 \mathbf{e}_2 轴的转动是不稳定的！类似的证明也说明绕 \mathbf{e}_1 轴和 \mathbf{e}_3 轴的转动是稳定的。这就证明了网球拍定理。

8.4 欧拉角

8.4.1 欧拉角

前面说过，刚体转动的位形空间 \mathcal{C} 是 3 维的，但是到现在为止我们一直没有给出过这个位形空间的广义坐标。现在我们来介绍一种最常用的广义坐标选取，即所谓的欧拉角。

我们要做的就是参数化刚体坐标 $\{\mathbf{e}_a\}$ 的所有取向，或者等价地参数化所有 3×3 的特殊正交矩阵。以固定的空间坐标系为基准，刚体坐标系的一个一般取向如图 (8.6) 所示。为了参数化刚体坐标系的取向，我们假设初始时刚体坐标系与空间坐标系完全重合，然后按照下面 3 步旋转到达一般取向，

$$\{\tilde{\mathbf{e}}_a\} \xrightarrow{R_3(\phi)} \{\mathbf{e}'_a\} \xrightarrow{R_1(\theta)} \{\mathbf{e}''_a\} \xrightarrow{R_3(\psi)} \{\mathbf{e}_a\}. \quad (8.42)$$

首先第一步，将与空间坐标系重合的刚体坐标系绕 $\tilde{\mathbf{e}}_3$ 轴旋转 ϕ 角，得到 $\mathbf{e}'_a = \tilde{\mathbf{e}}_b R_3(\phi)_{ba}$ ，其中矩阵 $R_3(\phi)$ 为

$$R_3(\phi) = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (8.43)$$

如图 (8.7) 所示。

第二步，将上面得到的 $\{\mathbf{e}'_a\}$ 绕 \mathbf{e}'_1 轴旋转 θ 角，得到 $\mathbf{e}''_a = \mathbf{e}'_b R_1(\theta)_{ba}$ ，其中矩阵 $R_1(\theta)$

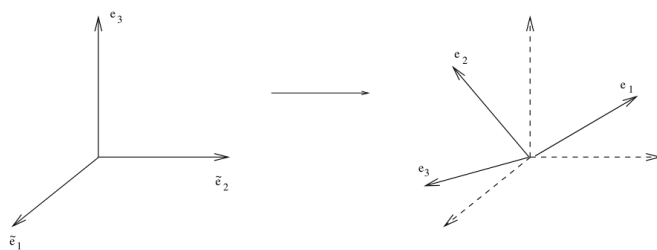


Figure 8.6: 刚体坐标系相对于空间坐标系的一般取向。图片来自 David Tong, Classical dynamics, 下面 3 幅图同样。

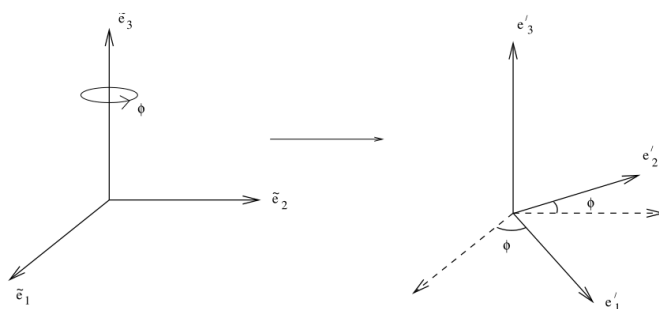


Figure 8.7: 绕 $\tilde{\mathbf{e}}_3$ 轴旋转 ϕ 角。

为,

$$R_1(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix}. \quad (8.44)$$

如图 (8.8) 所示。

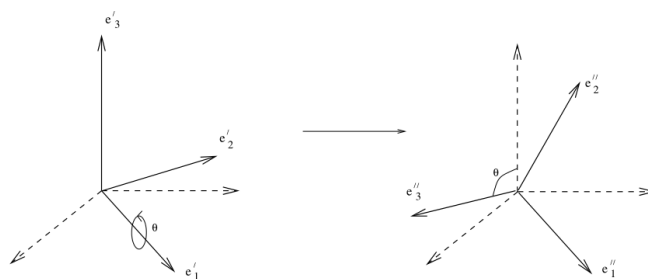


Figure 8.8: 绕 \mathbf{e}'_1 轴旋转 θ 角。

第三步, 将上一步得到的 $\{\mathbf{e}''_a\}$ 绕 \mathbf{e}''_3 轴旋转 ψ 角, 最终达到 $\mathbf{e}_a = \mathbf{e}''_b R_3(\psi)_{ba}$, 其中

$R_3(\psi)$ 为

$$R_3(\psi) = \begin{pmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (8.45)$$

如图 (8.9) 所示。

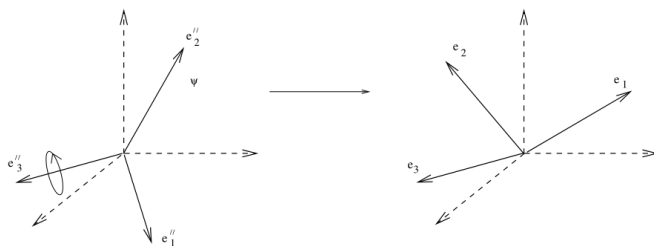


Figure 8.9: 绕 \mathbf{e}_3'' 轴旋转 ψ 角。

将三步联合起来, 即有 $\mathbf{e}_a = \tilde{\mathbf{e}}_b R_{ba}(\phi, \theta, \psi)$, 式中

$$R(\phi, \theta, \psi) = R_3(\phi) R_1(\theta) R_3(\psi). \quad (8.46)$$

这三个角度 ϕ, θ, ψ 就是**欧拉角**, 也就是我们要选取的转动位形空间广义坐标。矩阵 $R(\phi, \theta, \psi)$ 可以直接算出来, 为

$$\begin{pmatrix} \cos \psi \cos \phi - \cos \theta \sin \phi \sin \psi & -\cos \phi \sin \psi - \cos \theta \cos \psi \sin \phi & \sin \theta \sin \phi \\ \sin \phi \cos \psi + \cos \theta \sin \psi \cos \phi & -\sin \psi \sin \phi + \cos \theta \cos \psi \cos \phi & -\sin \theta \cos \phi \\ \sin \theta \sin \psi & \sin \theta \cos \psi & \cos \theta \end{pmatrix}.$$

任何一个特殊正交矩阵都可以表示成这种形式。

注意到 3 个欧拉角中, θ 是 \mathbf{e}_3 轴与 $\tilde{\mathbf{e}}_3$ 轴的夹角, ϕ 和 ψ 分别是绕 $\tilde{\mathbf{e}}_3$ 轴和绕 \mathbf{e}_3 轴转动的角度, 所以通常约定 3 个欧拉角的取值范围分别为

$$\theta \in [0, \pi], \quad \phi \in [0, 2\pi], \quad \psi \in [0, 2\pi]. \quad (8.47)$$

8.4.2 角速度的欧拉角表示法

我们可以将刚体的旋转角速度用欧拉角表示出来。为此首先注意到, 根据上面的 3 步旋转, 有

$$\vec{\omega} = \dot{\phi} \tilde{\mathbf{e}}_3 + \dot{\theta} \mathbf{e}'_1 + \dot{\psi} \mathbf{e}_3. \quad (8.48)$$

下面只需将 $\tilde{\mathbf{e}}_3$ 和 \mathbf{e}'_1 用最终的刚体坐标 $\{\mathbf{e}_a\}$ 表示出来。这并不难, 首先根据上面的 3 步, 并利用矩阵 $R(\phi, \theta, \psi)$ 的正交性, 我们有 $\tilde{\mathbf{e}}_3 = R_{3a}(\phi, \theta, \psi) \mathbf{e}_a$, 即

$$\tilde{\mathbf{e}}_3 = \sin \theta \sin \psi \mathbf{e}_1 + \sin \theta \cos \psi \mathbf{e}_2 + \cos \theta \mathbf{e}_3. \quad (8.49)$$

类似的, 根据上面 3 步的第 3 步, 也容易得到

$$\mathbf{e}'_1 = \cos \psi \mathbf{e}_1 - \sin \psi \mathbf{e}_2. \quad (8.50)$$

进而即可以得到角速度在刚体坐标中的表达式,

$$\vec{\omega} = [\dot{\phi} \sin \theta \sin \psi + \dot{\theta} \cos \psi] \mathbf{e}_1 + [\dot{\phi} \sin \theta \cos \psi - \dot{\theta} \sin \psi] \mathbf{e}_2 + [\dot{\psi} + \dot{\phi} \cos \theta] \mathbf{e}_3.$$

或者也可以写成

$$\begin{aligned} \omega_1 &= \dot{\phi} \sin \theta \sin \psi + \dot{\theta} \cos \psi \\ \omega_2 &= \dot{\phi} \sin \theta \cos \psi - \dot{\theta} \sin \psi \\ \omega_3 &= \dot{\psi} + \dot{\phi} \cos \theta. \end{aligned} \quad (8.51)$$

利用上面这个角速度的欧拉角表达式, 我们就可以在解出 $\omega_1, \omega_2, \omega_3$ 以后, 进一步把欧拉陀螺相对于空间坐标的转动情况用 $\dot{\phi}, \dot{\theta}, \dot{\psi}$ 表达出来, 因为三个欧拉角就是相对于空间坐标系定义的。

8.5 拉格朗日陀螺

欧拉陀螺通常可以指任何自由转动的刚体, 而拉格朗日陀螺则通常指的就是小朋友们玩的陀螺。如图 (8.10) 所示, 假设这个陀螺绕固定点 P 转动, 其质心到 P 点的距离为 l , 并且陀螺当然受重力的作用。很显然, 这样的陀螺有一个对称轴, 它绕这个对称轴旋转不变, 这个对称轴当然是惯量主轴, 记为 \mathbf{e}_3 , 由于旋转不变性, 与 \mathbf{e}_3 垂直的另两个主轴可以任意选, 只要两者相互垂直就行。显然, $I_1 = I_2 \neq I_3$, 即是说拉格朗日陀螺是一个对称陀螺。三个欧拉角如图 (8.10) 所示, 其中 ϕ 的运动称为进动, θ 的运动称为章动, ψ 的运动则称为自转。

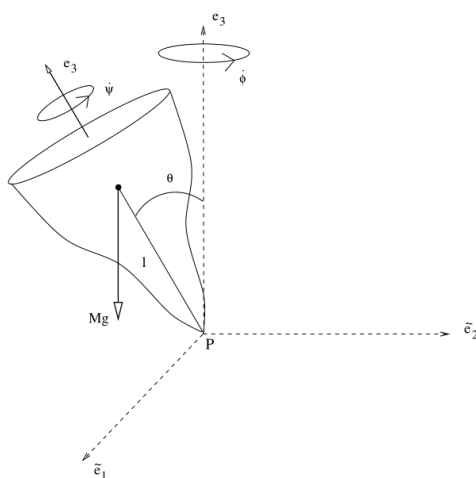


Figure 8.10: 绕固定点 P 转动的陀螺。图片来自 David Tong, classical dynamics, 下图同。

很容易写出欧拉陀螺的拉格朗日量, 为

$$\begin{aligned} L &= \frac{1}{2}I_1(\omega_1^2 + \omega_2^2) + \frac{1}{2}I_3\omega_3^2 - mgl \cos \theta \\ &= \frac{1}{2}I_1(\dot{\theta}^2 + \sin^2 \theta \dot{\phi}^2) + \frac{1}{2}I_3(\dot{\psi} + \cos \theta \dot{\phi})^2 - mgl \cos \theta. \end{aligned} \quad (8.52)$$

很容易求出与 3 个欧拉角对应的广义动量, 为

$$\begin{aligned} p_\theta &= \frac{\partial L}{\partial \dot{\theta}} = I_1 \dot{\theta} \\ p_\psi &= \frac{\partial L}{\partial \dot{\psi}} = I_3(\dot{\psi} + \cos \theta \dot{\phi}) = I_3 \omega_3 = J_3 \\ p_\phi &= \frac{\partial L}{\partial \dot{\phi}} = I_1 \sin^2 \theta \dot{\phi} + I_3 \cos \theta (\dot{\psi} + \dot{\phi} \cos \theta) \\ &= I_1 \sin^2 \theta \dot{\phi} + \cos \theta p_\psi. \end{aligned} \quad (8.53)$$

利用这些广义动量, 就可以将系统的能量 E 表达成哈密顿量

$$\begin{aligned} E &= \frac{1}{2}I_1(\dot{\theta}^2 + \sin^2 \theta \dot{\phi}^2) + \frac{1}{2}I_3(\dot{\psi} + \cos \theta \dot{\phi})^2 + mgl \cos \theta \\ \Rightarrow H &= \frac{p_\theta^2}{2I_1} + \frac{(p_\phi - p_\psi \cos \theta)^2}{2I_1 \sin^2 \theta} + \frac{p_\psi^2}{2I_3} + mgl \cos \theta. \end{aligned} \quad (8.54)$$

由于这个哈密顿量的表达式与 ϕ, ψ 无关, 所以显然有

$$[p_\phi, H] = [p_\psi, H] = 0. \quad (8.55)$$

从而系统有三个相互泊松对易的守恒量 p_ϕ, p_ψ, H 。而这又是一个自由度数为 3 的哈密顿系统, 从而拉格朗日陀螺是一个可积系统。

由于 p_ϕ, p_ψ 守恒, 而且哈密顿量不显含 ϕ, ψ , 所以我们可以先忽略 ϕ, ψ 的运动, 进而将上述系统约化为单个自由度 θ 的系统, 约化系统的哈密顿量为

$$H = \frac{p_\theta^2}{2I_1} + U(\theta), \quad (8.56)$$

式中约化系统的势能 $U(\theta)$ 由下式给出

$$U(\theta) = \frac{(p_\phi - p_\psi \cos \theta)^2}{2I_1 \sin^2 \theta} + \frac{p_\psi^2}{2I_3} + mgl \cos \theta. \quad (8.57)$$

约化系统当然也是可积系统, 其守恒量就是能量,

$$\frac{1}{2}I_1 \dot{\theta}^2 + U(\theta) = E. \quad (8.58)$$

这就是一个标准的一维问题, 我们在第一章中就已经学会如何对这样的问题进行积分了。根据第一章的分析可以知道, θ 的运动范围必定在 $U(\theta) \leq E$ 的区间上。

我们可以画出 $U(\theta)$ 的势能曲线。注意到当 $\theta \rightarrow 0, \pi$ 时, $U(\theta) \rightarrow \infty$ (除非 $p_\phi = \pm p_\psi$, 这种例外我们后面再分析), 并且根据微积分知识可以证明 $U(\theta)$ 只有一个极小值, 记为 θ_0 , 所以势能曲线必定如图 (8.11) 所示。给定能量 E , 很显然, θ 必定在 $[\theta_1, \theta_2]$ 的区间

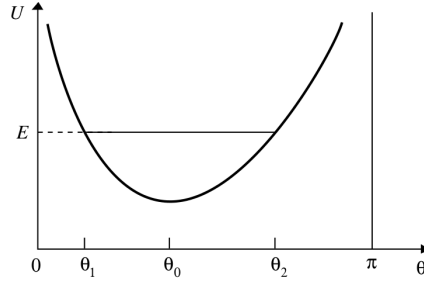


Figure 8.11: $U(\theta)$ 的势能曲线。图片来自 Kibble, classical mechanics, 下面 3 幅图同样。

上来回运动, 如图 (8.11) 所示。当然, 如果能量水平线 E 和势能曲线在最低点相切, 那 θ 就只能等于 θ_0 , 这种情况描述的就是一个以固定倾斜角度 θ_0 绕着竖直轴稳定进动的陀螺。

按照一维问题求解完 $\theta(t)$ 的运动以后, 就可以进一步求 ϕ, ψ 的运动, 方程如下

$$\begin{aligned}\dot{\phi} &= \frac{\partial H}{\partial p_\phi} = \frac{p_\phi - p_\psi \cos \theta}{I_1 \sin^2 \theta} \\ \dot{\psi} &= \frac{\partial H}{\partial p_\psi} = \frac{p_\psi}{I_3} - \cos \theta \frac{p_\phi - p_\psi \cos \theta}{I_1 \sin^2 \theta}.\end{aligned}\quad (8.59)$$

习惯上常常将 θ, ϕ 的运动画在两维球面上, 以 θ 为球面的纬度, 以 ϕ 为球面的经度。 ϕ 的运动就是陀螺的进动, 而 θ 在 $[\theta_1, \theta_2]$ 区间上的来回运动就是陀螺的章动。很显然, 这个球面上的运动轨迹也就是陀螺转轴 \mathbf{e}_3 的运动轨迹。

从公式 (8.59) 可以看出, 进动角速度 $\dot{\phi} = 0$ 当且仅当 $\theta = \arccos(p_\phi/p_\psi)$, 如果这个角度落在 $[\theta_1, \theta_2]$ 的区间之外, 或者 $|p_\phi/p_\psi| > 1$, 那在整个 $[\theta_1, \theta_2]$ 区间上进动角速度 $\dot{\phi}$ 将不会改变正负号, 即陀螺永远按照一个方向进动。这时候在 (θ, ϕ) 球面上, \mathbf{e}_3 轴的运动轨迹必定如图 (8.12) 所示。

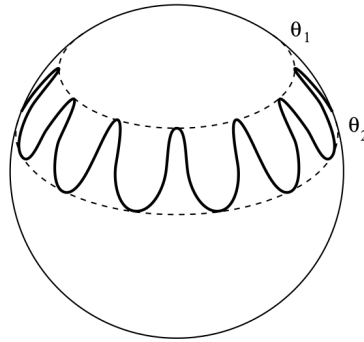


Figure 8.12: 陀螺永远按照一个方向进动的情形。

反过来, 如果 $\theta_1 < \arccos(p_\phi/p_\psi) < \theta_2$, 那随着陀螺的章动, 其进动角速度将会变号, 因此陀螺在纬度靠近 θ_1 那一端的进动方向就和纬度靠近 θ_2 那一端的进动方向相反, 从而 \mathbf{e}_3 轴的运动轨迹必定如图 (8.13) 所示 (注意 \mathbf{e}_3 轴总是顺着这条轨迹运动)。

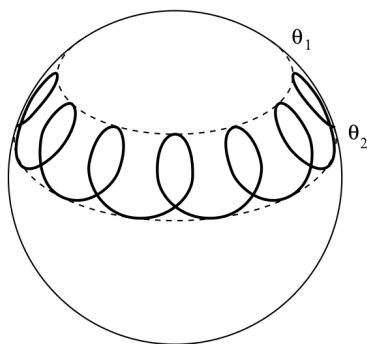


Figure 8.13: 陀螺的进动不断改变方向的情形。

特别的，如果 $\arccos(p_\phi/p_\psi) = \theta_1$ ，那 \mathbf{e}_3 轴的轨迹线在 θ_1 的那一端就会退化为尖点，如图 (8.14) 所示。在那些尖点上，陀螺的章动速度 $\dot{\theta} = 0$ ，进动角速度 $\dot{\phi}$ 也等于零，从而陀螺的动能达到最小。这种情形当我们让陀螺转动但是初始时自转轴 \mathbf{e}_3 瞬时静止就会出现。值得注意的是， \mathbf{e}_3 轴的轨迹曲线不会在 θ_2 处变为尖点，原因在于，一个初始时转轴

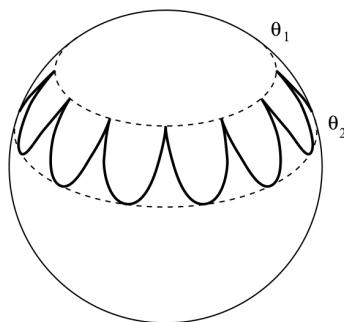


Figure 8.14: 初始时自转轴瞬时静止的陀螺运动。

瞬时静止的陀螺无法自动抬高其自转轴。

转轴垂直的陀螺

下面来分析一种重要的特殊情况，即陀螺的自转轴 \mathbf{e}_3 垂直且保持不变的转动，也即 θ 恒等于零的情形。我们要分析什么时候这种转动是稳定的，什么时候会转化为上面分析的有章动的一般情形。首先，从 $U(\theta)$ 的表达式容易看出， $\theta = 0$ 能出现就意味着必定有 $p_\phi = p_\psi = J_3$ ，因为否则 $\theta = 0$ 时必有 $U(0) = \infty$ 从而不可能。

下面在 $U(\theta)$ 的表达式中取 $p_\phi = p_\psi = J_3$ ，从而有

$$U(\theta) = \frac{J_3^2}{2I_1} \tan^2\left(\frac{1}{2}\theta\right) + \frac{J_3^2}{2I_3} + mgl \cos \theta. \quad (8.60)$$

为了研究 $\theta = 0$ 的稳定性，我们将这个势能表达式在 $\theta = 0$ 处泰勒展开到二阶小量，即有


$$U(\theta) = \left(\frac{J_3^2}{2I_3} + mgl\right) + \frac{1}{2}\left(\frac{J_3^2}{4I_1} - mgl\right)\theta^2 + \dots \quad (8.61)$$

很显然，如果

$$\frac{J_3^2}{4I_1} - mgl > 0 \Rightarrow \omega_3^2 > \frac{4I_1 mgl}{I_3^2} = \omega_0^2, \quad (8.62)$$

那 $\theta = 0$ 就是势能的极小值，从而系统就是稳定的。即是说，如果角速度 $\omega_3 > \omega_0$ ，那转轴垂直就是稳定的！但是反过来，如果 $|\omega_3| < \omega_0$ ，那 $\theta = 0$ 就是势能的极大值，从而转动就是不稳定的。

对于一个真实的陀螺来说，我们可以给它一个很大的角速度 ω_3 ，让它一直绕着垂直的转轴转动，但是摩擦力会逐渐将角速度降下来，当 $\omega_3 < \omega_0$ 时，陀螺就开始章动和进动了。



参考文献

除了 Polyakov 的视频 (B 站上有), 基本上我只参考了下面几份文献:

- 朗道《力学》;
- Tom W.B. Kibble and Frank H. Berkshire, Classical Mechanics;
- 阿诺德,《经典力学的数学方法》;
- David Tong, Classical Dynamics (可以在 David Tong 的页面上找到);
- 刘川,《理论力学》;
- 李德明、陈昌民,《经典力学》

这 6 份文献中, 刘川老师的《理论力学》和朗道《力学》有一些相似之处, 两本相比较更推荐朗道《力学》。李德明和陈昌民老师的《经典力学》是一本很好的书, 虽然用来初学会有些难, 但是学过分析力学后回头再看一本书, 却是很能提高的。

但入门的话, 最推荐 Tom W.B. Kibble and Frank H. Berkshire, Classical Mechanics, 虽然我读过的部分也不太多, 但都很不错。其次, David Tong 的讲义很多地方也很不错。

阿诺德《经典力学的数学方法》有不少精妙的地方, 但对数学兴趣不大的物理人我不推荐读, 把物理人都带偏到数学上去不见得好。

读我的讲义在“微分形式”那一节遇到困难的同学, 推荐大家参考龚升老师很经典的《简明微积分》, 对于 3 维情形, 那里讲得更详细更清楚, 因此更容易入门。

另外, 我的讲义介绍了最初步的可积系统知识。但可积系统实际上是一个数学上很坚深的领域, 而且现在更多的研究可能在量子可积系统。但如果你对经典可积系统很感兴趣, 那我推荐下面这本书: Olivier Babelon, Denis Bernard and Michel Talon, Introduction to Classical Integrable Systems

其它书我都没有参考过。比如著名的 Goldstein, 太厚了, 而且我也不太喜欢, 因此没有参考。

