

目录

第二章 最小作用量原理	2
2.1 物理学的局部视角与全局视角	3
2.2 从费马原理到变分法	4
2.2.1 费马原理	4
2.2.2 变分和泛函导数	6
2.2.3 欧拉-拉格朗日方程	8
2.3 相空间的最小作用量原理	11
2.4 坐标空间的最小作用量原理	14
2.4.1 最小作用量原理与拉格朗日方程	14
2.4.2 发现洛伦兹力	17
2.4.3 最小作用量原理能够导出任何方程吗?	20
2.5 广义坐标和广义动量	20

第二章 最小作用量原理

陈童

本章是理论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。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

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。

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。

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)式中除以的一个整体常数，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

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

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?

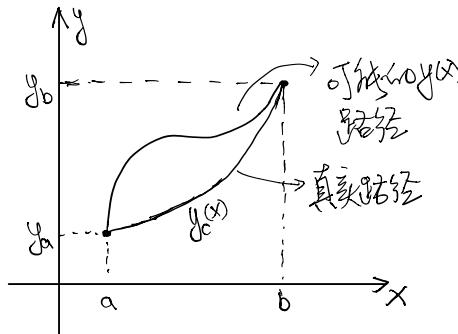


图 2.1: 光的路径

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 $L(y, y') = n(y)\sqrt{1 + (y')^2}$, 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在所有两端由(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，变分和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对指标 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

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

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

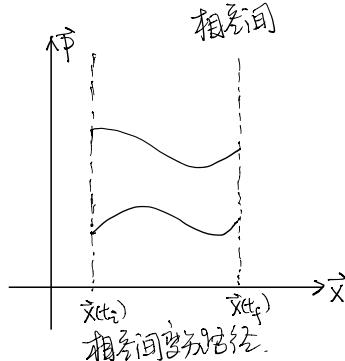
$$S[\mathbf{x}(t), \mathbf{p}(t)] = \int_{t_i}^{t_f} [\mathbf{p} \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{p})] dt. \quad (2.32)$$

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来两端的 \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。

图 2.2: 可能的相空间路径。注意时间 t 我们没有画出。

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，对一个上指标的量求导，其结果相当于一个下指标的量，而对一个下指标的量求导，其结果则相当于一个上指标的量。

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 $p\dot{x} - H$ 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 $L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}m\dot{\mathbf{x}}^2 - V(\mathbf{x})$ ，拉格朗日方程给出来的是

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

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最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加的这一项上，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 $F_{ij} = \epsilon_{ijk} B^k$ 。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 $\epsilon_{jik} = -\epsilon_{ijk}$, 并且 $\epsilon_{123} = 1$ 。

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

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

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地是 $\mathbf{p} = m\dot{\mathbf{x}}$, 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。

2.5 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, ..., 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

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。

为

$$\begin{aligned} S &= \int dt [p_\mu \dot{x}^\mu - H(x, p)] \\ &= \int p_\mu dx^\mu - \int H dt. \end{aligned} \quad (2.65)$$

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 ，

$$\Theta = p_\mu dx^\mu. \quad (2.66)$$

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

$$\Theta = p_\mu dx^\mu = p_a dq^a. \quad (2.67)$$

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

$$p_a = p_\mu \frac{\partial x^\mu}{\partial q^a}. \quad (2.68)$$

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

$$S = \int dt [p_a \dot{q}^a - H(q, p)]. \quad (2.69)$$

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

$$\frac{dq^a}{dt} = \frac{\partial H}{\partial p_a}, \quad \frac{dp_a}{dt} = -\frac{\partial H}{\partial q^a}. \quad (2.70)$$

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

$$x = R \cos(\theta), \quad y = R \sin(\theta). \quad (2.71)$$

将辛势 Θ 分别在两种坐标中表达，即有

$$\begin{aligned} \Theta &= p_x dx + p_y dy = p_\theta d\theta \\ &\Rightarrow p_\theta = [-p_x \sin(\theta) + p_y \cos(\theta)]R \\ &\Rightarrow \mathbf{p}^2 = p_\theta^2 / R^2. \end{aligned}$$

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，

$$H = \frac{\mathbf{p}^2}{2m} + V = \frac{p_\theta^2}{2mR^2} + V(\theta). \quad (2.72)$$

进而可以写出这个例子的哈密顿正则方程

$$\begin{aligned} \dot{\theta} &= \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{mR^2} \\ \dot{p}_\theta &= -\frac{\partial H}{\partial \theta} = -\frac{\partial V}{\partial \theta}. \end{aligned} \quad (2.73)$$

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)，然后再坐标转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

$$S[q(t)] = \int dt L(q, \dot{q}). \quad (2.74)$$

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

$$\frac{\partial L}{\partial q^a} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^a} \right) = 0. \quad (2.75)$$

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，

$$H(q, p) = \text{extrem}_{\{\dot{q}^a\}} [p_b \dot{q}^b - L(q, \dot{q})]. \quad (2.76)$$

显然，为了进行这个勒让德变换，需要求解

$$\frac{\partial}{\partial \dot{q}^a} [p_b \dot{q}^b - L(q, \dot{q})] = 0 \Rightarrow p_a = \frac{\partial L}{\partial \dot{q}^a}. \quad (2.77)$$

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

$$T_1 = \frac{1}{2} m_1 l_1^2 \dot{\theta}_1^2, \quad V_1 = -m_1 g l_1 \cos(\theta_1). \quad (2.78)$$

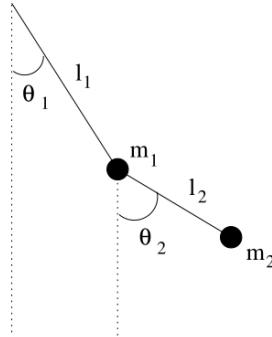


图 2.3: 双摆

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

$$x_2 = l_1 \sin(\theta_1) + l_2 \sin(\theta_2), \quad y_2 = l_1 \cos(\theta_1) + l_2 \cos(\theta_2). \quad (2.79)$$

则容易求出质点2的动能 T_2 为

$$\begin{aligned} T_2 &= \frac{1}{2} m_2 (\dot{x}_2^2 + \dot{y}_2^2) \\ &= \frac{1}{2} m_2 (l_1^2 \dot{\theta}_1^2 + l_2^2 \dot{\theta}_2^2 + 2l_1 l_2 \cos(\theta_1 - \theta_2) \dot{\theta}_1 \dot{\theta}_2). \end{aligned} \quad (2.80)$$

质点2的重力势能 V_2 为

$$V_2 = -m_2 g y_2 = -m_2 g (l_1 \cos(\theta_1) + l_2 \cos(\theta_2)). \quad (2.81)$$

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)

$$\begin{aligned} L &= \frac{1}{2} (m_1 + m_2) l_1^2 \dot{\theta}_1^2 + \frac{1}{2} m_2 l_2^2 \dot{\theta}_2^2 + m_2 l_1 l_2 \cos(\theta_1 - \theta_2) \dot{\theta}_1 \dot{\theta}_2 \\ &+ (m_1 + m_2) g l_1 \cos(\theta_1) + m_2 g l_2 \cos(\theta_2). \end{aligned} \quad (2.82)$$

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

$$\frac{\partial L}{\partial \theta_1} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}_1} \right) = 0, \quad \frac{\partial L}{\partial \theta_2} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}_2} \right) = 0. \quad (2.83)$$

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！