广义坐标

所谓约束体系是指其状态在运动过程中受到了某种限制而不能自由变化的体系。数学上，这意味着描述体系的状态参量——位置和速度——是满足某种关系的，这种关系就称为是约束方程，一般来说它具有如下的形式

$$f\left(\bar{r}_1,\bar{r}_2,\cdots,\bar{r}_n,\dot{\bar{r}}_1,\dot{\bar{r}}_2,\cdots,\dot{\bar{r}}_n,t\right)=f\left(\bar{r},\dot{\bar{r}},t\right)=0 \qquad (1)$$

这里以及以后，在不引起混淆的情况下，我都将把函数中的一组带有下标的自变量缩记为一个不带下标的量。

譬如刚体就是一个特殊的约束体系，因为其中任何两点的距离在运动过程中都是不变的，即 $r_{ab}=\left|\bar{r}_a-\bar{r}_b\right|=\text{const.}$。上一章最后的那个例子也是一个约束问题，在那里，不仅要求下面那个楔形物体只能在水平方向运动[约束方程 $y_2=\text{const.}$]，而且还要求两个物体在运动过程中是始终保持接触的[$y_1=\left(x_2-x_1\right)\tan\theta$]。再比如一个限制在某个曲面上运动的粒子，约束方程就是该曲面的方程，而如果曲面本身又是在空间按给定方式运动——譬如一个粒子在半径以某个给定速度不断增大的球面上的运动——那么约束方程就将显含时间：

$$f\left(x_1,x_2,x_3,t\right)=x_1^2+x_2^2+x_3^2-\left(a_0+v_0t\right)^2=0 \qquad (2)$$

物体之所以不能自由运动，究其原因是由于对体系施加约束的物体（约束物）提供了一个力，这个力与其他的力的一起作用恰好使得物体只能在约束上运动。这种由约束物提供的、使得运动物体只能按照给定方式运动的力就称为约束力，而其他的力则都被称为是主动力。约束力本质上是物体形变后产生的恢复力。当运动物体挤压、拉伸约束物时，二者都会发生形变，并相互以恢复力作用于对方，这就产生了约束力，如果约束力不够大，则物体的运动将有不遵循约束的趋势，于是就会进一步压迫约束物，约束力也就相应地增大，一直到物体的运动恰好遵循约束为止。总之，约束力的特点是应运而生的——因运动需要而产生的。另一方面，约束物是比较硬的，发生极小的形变即可提供足够大的约束力，而在考虑约束物的形状时这极小的形变又可忽略掉，即约束物的形状可以认为是不变的。

因此，在运动物体的挤压下，约束物体发生微小的形变，提供了恰当的约束力，使物体刚好能遵循约束物运动，而约束物的形状又可以近似认为是不变的，这就是约束的含义所在。通常所说的刚性杆、不可伸长的绳子、斜面等等就是约束的具体例子。事实上，当单摆摆动时，一方面，刚性杆是有微小的伸长，因而杆中有了张力，杆以这种张力作用与摆；另一方面，杆的长度可以认为几乎没有改变，因而是不可伸长的刚性杆。这就是不可伸长的刚性杆会提供适当大小的张力的实质。

值得指出的是，当考虑约束问题时，人们通常将粒子受到的作用分为约束力和主动力。而在上一章讨论质点组问题时，我们曾经把作用在粒子上的力分为内力和外力。那么它们之间受否有什么必然的联系呢？实际上，约束力可以是内力，也可以是外力，这取决于施加约束的物体是否包含在我们所感兴趣的那个体系里面；同样的，主动力也可能是内力或者是外力，这取决于这个力是来自体系内部其他粒子的作用还是由外部其他物体所提供的。

下面我想就约束体系的运动学作一些讨论。对于自由的粒子，其在空间中的位置是用位矢$\vec{r}$定义的，选定某个坐标标系后，位矢$\vec{r}$可以用三个坐标分量来表示，例如$\left(x_1,x_2,x_3\right)$；为了确定由n个粒子组成的体系在空间中的位形，对每一个粒子我们就需要一位矢$\vec{r}_a$ $\left(a=1,2,\cdots,n\right)$或者说三个坐标$\left(x_{a,1},x_{a,2},x_{a,3}\right)$，因此就需要有$n$个位矢，也就是$3n$个坐标。通常把为了唯一确定体系的位置所必需的独立变量的数目称为自由度，通常记为s。这里，自由度是$s=3n$。如果体系的运动受到某种限制，也就是说，如果这$3n$个坐标之间满足一些关系（约束方程），那么其中一些坐标就不再是独立的了。事实上，如果有m个约束方程，那么独立坐标的数目或者说体系的自由度就变为了$s=3n-m$。需要注意的是，对于一个自由度为s的体系，描述它的s个变量不必具有长度的量纲，它们也可以是角度或者其他物理量。例如，在前面二维谐振子的例子中，粒子的位置是用一个长度和一个角度描述的。实际上，我们会发现在有些问题中选择面积、体积、能量、动量、电极化强度、磁化强度甚至是某些无量纲的变量描述体系的位形是更加方便的。力学上把这些独立的、并且可以

完全定义体系位置的任意 s 个变量称为体系的广义坐标，后面我们将经常用 $q_1, q_2, \cdots, q_s$ 来表示广义坐标，而其导数 $\dot{q}_i$ 则相应地称为广义速度。

前面我们所提到的例子都不涉及对物体速度的限制，相应的约束方程也仅仅是位置以及时间的函数

$$f\left(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_n, t\right) = f\left(\vec{r}, t\right) = 0 \tag{3}$$

我们将凡是能化为这种形式的约束称为完整约束。而其他的约束则称为非完整约束。与此相关的两个概念是：如果一个体系所受的约束全部都是完整约束，则将其称为完整约束体系，否则就称为非完整体系。

完整约束的重要性在于其约束方程是坐标之间的一个代数方程，而我们就可以利用这些关系将那些不独立的变量消去。

作为一个例子，考察单个粒子在重力场中的运动。如果粒子的运动完全是自由的，这时候我们说它的自由度为 3，而广义坐标就可以选为粒子的三个迪卡尔坐标 $\left(x_1, x_2, x_3\right)$，这时候粒子在任一时刻的位置由三维空间中一个点表示，而随着时间的推进，这个点在空间画出了一条曲线，也就是粒子运动的轨迹。但是，如果粒子被限制在某个曲面（譬如球面）上运动，那么这三个坐标就不再是独立的了，而是必须要满足这个曲面的方程

$$f\left(\vec{r}\right) = f\left(x_1, x_2, x_3\right) = x_1^2 + x_2^2 + x_3^2 - a^2 = 0 \tag{4}$$

因此它的自由度就变为了 $2 = 3 - 1$。这时候，我们就可以选球面坐标 $\left(\theta, \varphi\right)$ 作为广义坐标。而任一时刻粒子所在的位置，你可以用三个不独立的迪卡尔坐标描述，也可以等价地用这两个独立的广义坐标（曲面参数）来描述，二者的关系是由曲面的参数方程来决定的，即

$$\vec{r} = a\sin\theta\cos\varphi\hat{x}_1 + a\sin\theta\sin\varphi\hat{x}_2 + a\cos\theta\hat{x}_3 = \vec{r}\left(\theta, \varphi\right) \tag{5a}$$

或者

$$\begin{aligned} x_1 &= a\sin\theta\cos\varphi = x_1\left(\theta, \varphi\right) \\ x_2 &= a\sin\theta\sin\varphi = x_2\left(\theta, \varphi\right) \\ x_3 &= a\cos\theta \qquad\quad = x_3\left(\theta, \varphi\right) \end{aligned} \tag{5b}$$

进一步，如果物体不仅限制在球面上，而且它只能某个高度运动，即三个坐标除了满足关系(4)之外，还要满足

$$g\left(\vec{r}\right)=g\left(x_1,x_2,x_3\right)=x_3-h=0 \tag{6}$$

这时候自由度就将变为$1=3-1-1$，实际上，这种情况下粒子只能在一条曲线（纬线圈）上运动，而广义坐标则可以选为这条曲线的参数譬如φ。用这个参数方程就等价于（由于$\cos\theta_0=h/a$）

$$\vec{r}=\sqrt{a^2-h^2}\cos\varphi\hat{x}_1+\sqrt{a^2-h^2}\sin\varphi\hat{x}_2+h\hat{x}_3=\vec{r}\left(\varphi\right) \tag{7a}$$

或者

$$\begin{aligned} x_1&=\sqrt{a^2-h^2}\cos\varphi=x_1\left(\varphi\right)\\ x_2&=\sqrt{a^2-h^2}\sin\varphi=x_2\left(\varphi\right)\\ x_3&=h\qquad\qquad\qquad\ =x_3\left(\varphi\right)\end{aligned} \tag{7b}$$

我们说一个类似(4)这样的方程确定了三维空间中的一个曲面，或者说是一个$2=3-1$维的子空间；而两个这样的方程（或者说两个曲面）则确定了三维空间中的一条曲线，或者说是一个$1=3-2$维的子空间，它是两个曲面相交的部分。

现在考虑一个由n个粒子构成的体系。这时候体系在任一时刻的位置就需要空间中的n个点来确定，而当体系运动时，每一个这样的点都在空间画出一条曲线，整个体系的“轨迹”将是杂乱无章的。为此，我们想象一个$3n$维的空间，它的前三个坐标就是第一个粒子的三个坐标，后面三个坐标则是第二个粒子的三个坐标，以此类推。在这样一个空间，其中一个点就代表了体系（中所有粒子）在某一时刻的位置，而随着时间的推进，这个点所画出的曲线就是在这个抽象的空间中体系整体的运动轨迹。现在，约束方程(3)就可以看作是这个空间中的某个$3n-1$ 维的“超曲面”，而代表体系位置的点就只能在这个超曲面上运动。当然，如果体系受到m个完整约束，即其坐标满足m个类似(3)这样的关系，那么它们将给出这个$3n$空间中的曲面或者说是子空间，曲面的维数等于体系的自

由度$s=3n-m$，而代表体系的点就只能在这个s维曲面上运动。这时候我们可以将描述这个曲面的s个参数譬如说$q_1,q_2,\cdots,q_s$作为广义坐标。而这个曲面，或者说分别以$q_1,q_2,\cdots,q_s$为坐标构成的空间就称为位形空间。位形空间的优点在于粒子在其中的运动是完全自由的。在图 1*a* 中，我画出了限制在球面上的粒子在三维欧几里德空间中的三个可能的轨迹，而在图 1*b* 中，则分别画出了在位形空间(θ,φ)中相应的轨迹。

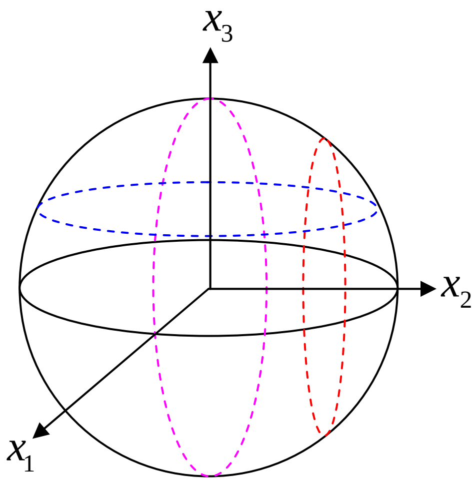

图 **1*a***

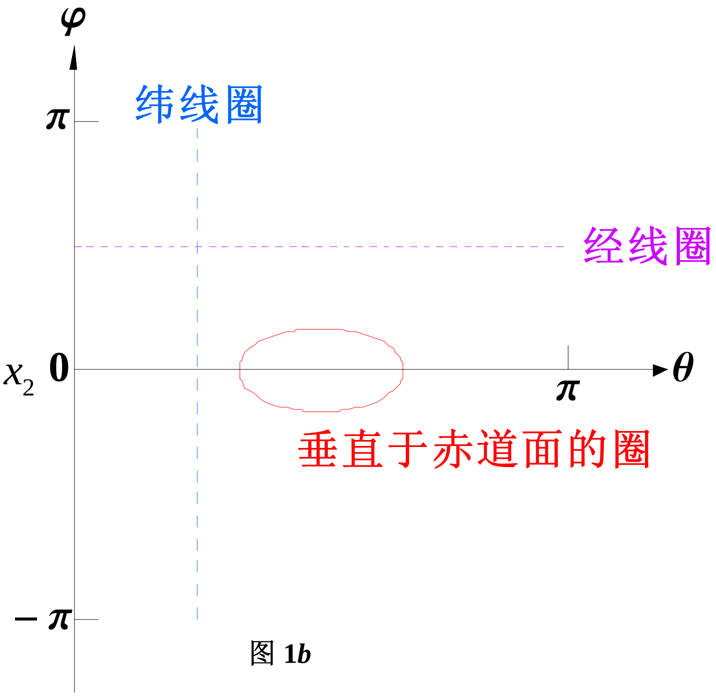

图 **1*b***

与限制在曲面或者曲线上的单粒子情况类似，对于n个粒子受到m个完整约束的体系，其在任一时刻所处的位置，你可以用$3n$个不独立的迪卡尔坐标描述，也可以等价地用s个广义坐标（位形空间或者说那个s维曲面的曲面参数）来描述，二者的关系是由s维曲面的参数方程来决定的，即

$$\vec{r}_a=\vec{r}_a(q,t),\qquad (a=1,2,\cdots,n) \tag{8a}$$

或者写成分量形式

$$x_{a,i}=x_{a,i}(q,t),\qquad (a=1,2,\cdots,n;\ \ i=1,2,3) \tag{8b}$$

我们将关系(8)称为变换方程。它们给出了粒子的迪卡尔坐标与体系的广义坐标之间的联系。

如果将方程(8*a*)两边对时间求微商，就得到

$$\vec{v}_a = \dot{\vec{r}}_a = \frac{\partial \vec{r}_a}{\partial q_i}\dot{q}_i + \frac{\partial \vec{r}_a}{\partial t} = \dot{\vec{r}}_a\left(q, \dot{q}, t\right) \tag{9}$$

这个关系将每一个粒子的速度与体系的广义坐标和广义速度联系了起来。由此可以得到两个很有用的关系式

$$\frac{\partial \dot{\vec{r}}_a}{\partial \dot{q}_i} = \frac{\partial \vec{r}_a}{\partial q_i}, \qquad \frac{\partial \dot{\vec{r}}_a}{\partial q_i} = \frac{d}{dt}\frac{\partial \vec{r}_a}{\partial q_i} \tag{10}$$

第二个式子是说$\vec{r}_a$对广义坐标的偏导数与对时间的（全）微商是可以交换的；而第一个式子则是说粒子速度对广义速度的偏微商中，分子和分母中都出现的对时间微商的这个算子是可以抵消掉的。

从上面的分析我们看到，无论如何，对于完整体系，我们是可以找到一组独立变量来描述它在任意时刻的位置的。那么，对于非完整体系又如何呢？对于包含速度的约束，尽管体系的变量不再是独立的，体系的自由度也会相应的减少，但是一般来说这些约束方程并不能使我们将不独立的坐标消去。为清楚起见，考虑下面的约束

$$\sum_{k=1}^{N} A_k\left(x_1, x_2, \cdots, x_N\right)\dot{x}_k = 0 \tag{11}$$

这是一个线性微分约束（因为约束方程是速度的线性函数），两边乘以dt，得到

$$\sum_{k=1}^{N} A_k dx_k = 0 \tag{12}$$

如果它能写成某个函数的全微分，即

$$\sum_{k=1}^{N} A_k dx_k = d\varphi, \qquad \varphi = \varphi\left(x_1, x_2, \cdots, x_N\right) \tag{13}$$

那么这个约束实际上就是完整约束，因为这种情况下式(12)也就意味着

$$\varphi = \varphi\left(x_1, x_2, \cdots, x_N\right) = \text{const}. \tag{14}$$

但是，由于这时候必然有

$$A_k = \frac{\partial \varphi}{\partial x_k} \tag{15}$$

由此我们可得到

$$\frac{\partial A_k}{\partial x_l}=\frac{\partial}{\partial x_l}\frac{\partial \varphi}{\partial x_k}=\frac{\partial^2 \varphi}{\partial x_l \partial x_k} \tag{16}$$

而由于偏导数是可以交换的，因此这也就意味着

$$\frac{\partial A_k}{\partial x_l}=\frac{\partial^2 \varphi}{\partial x_l \partial x_k}=\frac{\partial^2 \varphi}{\partial x_k \partial x_l}=\frac{\partial}{\partial x_k}\frac{\partial \varphi}{\partial x_l}=\frac{\partial A_l}{\partial x_k} \tag{17}$$

这个条件称为可积条件，也就是方程(11)表示完整约束的条件：

$$\frac{\partial A_k}{\partial x_l}=\frac{\partial A_l}{\partial x_k}, \qquad \left(k,l=1,2,\cdots,N\right) \tag{18}$$

对于一般的线性微分约束，由于A_k可以是任意的，当然它们也就并非一定要满足这样的关系，这时候，我们也就不再能利用约束方程达到消除不独立坐标的目的了。

如果你将$\vec{A}=\left(A_1,A_2,\cdots,A_N\right)$想象为某个$N$维空间中的矢量，那么上面的条件(17)实际上就是在讲$\vec{A}$是无旋的。（回忆一下，如果$\nabla\times\vec{F}=0$或者写成分量形式$\partial F_i \big/ \partial x_j=\partial F_j \big/ \partial x_i$，则力$\vec{F}$是保守的，也就是说功$\vec{F}\cdot d\vec{r}=F_1dx_1+F_2dx_2+F_3dx_3$可以写成某个势能函数的全微分$\vec{F}\cdot d\vec{r}=-dU$，其中$U$只是位置的函数$U=U\left(x_1,x_2,x_3\right)$，负号是由于物体习惯而引入的）。

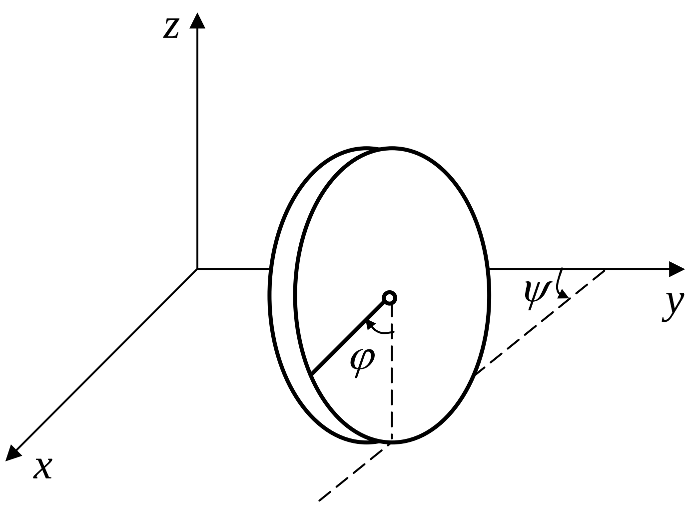

图 2*a*

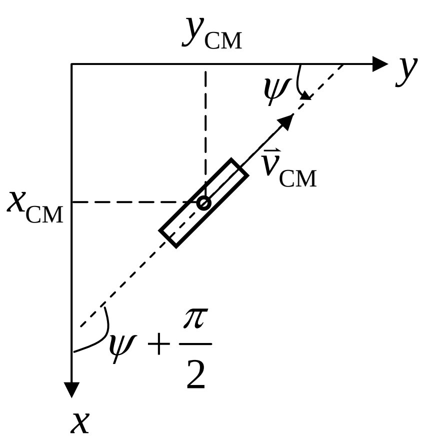

图 2*b*

举一个教科书上常见的不可积约束的例子：在平面上作纯滚动的轮子，假设轮子不会倾斜（即轮子平面始终在铅垂面内），轮子的半径为 a [图 2*a*]。轮子的位置可以用质心坐标 x_{CM}、y_{CM}、描述轮子转动的角度 φ 以及描述轮子平面相对于 y 轴的角度 ψ 来确定。纯滚动意味着质心的速度和自转角速度有下面的关系

$$v_{\text{CM}} = a\dot{\varphi} \tag{19}$$

写成分量形式就是[图 2*b*]

$$\begin{aligned} \dot{x}_{\text{CM}} &= -v_{\text{CM}}\sin\psi = -a\dot{\varphi}\sin\psi \\ \dot{y}_{\text{CM}} &= v_{\text{CM}}\cos\psi = a\dot{\varphi}\cos\psi \end{aligned} \tag{20}$$

或者

$$\begin{aligned} dx_{\text{CM}} + a\sin\psi\, d\varphi = 0 \\ dy_{\text{CM}} - a\cos\psi\, d\varphi = 0 \end{aligned} \tag{21}$$

这是一个不可积的线性微分约束，也就是说由它们无法得到形如

$$f\left(x_{\text{CM}}, y_{\text{CM}}, \varphi, \psi\right) = 0 \tag{22}$$

的方程。这在物理上是显然的：如果让轮子绕任意路线运行一周回到原处，那么 x_{CM}、y_{CM} 和 ψ 就都回到了出发时的数值，而 φ 则未必，这表明这四个变量之间并无一定的函数关系，因此，约束是非完整约束。

但是，如果轮子是沿着某条直线做纯滚动（$\psi = \text{const.}$），这时候约束方程就将是可积的，积分后给出

$$\begin{aligned} x_{\text{CM}} + a\varphi\sin\psi = c_1 \\ y_{\text{CM}} - a\varphi\cos\psi = c_2 \end{aligned} \tag{23}$$

这里 c_1 和 c_2 都是积分常数，此时约束变为完整约束，事实上只需要一个坐标 φ 就可以描述轮子的运动。

我们看到，对于完整体系，独立坐标的数目就等于体系的自由度，而对于非完整约束，约束方程不能用来消除不独立变量，因此，对于非完整体系是找不到合适的独立的广义坐标的，或者说我们需要比体系的自由度更多的广义坐标描述其运动。

除了包含速度的非完整约束，力学中还有一种常见的完整约束，在数学上它是用不等式来表示的。例如，容器中气体分子的运动，这时候由于气体分子不能穿透器壁，所以每一个气体分子都被限制在容器内运动；再比如一个小球在重力场中从一个大球的顶端滚下，约束方程是

$$x^2 + y^2 + z^2 - a^2 \geq 0 \tag{24}$$

这样的约束又称为可解约束（可解除的约束）。

小结一下：凡是约束方程可以写成(3)那种形式的约束称为完整约束；其他的约束则称为非完整约束，包括微分约束（线性微分约束是一个特例）、可解约束。有时我们也根据约束方程是否显含时间t把它分为稳定约束和不稳定约束。

我们知道，约束的效应是提供一个力，使得粒子只能按规定的方式运动。一般来说，主动力的规律是我们所知道的，而约束力由于是应运而生的，因此它本身也是问题的未知因素，只有在你已经知道了运动的情况下才能了解其信息。在 Newton 的框架中，有一个约束就会相应地增加一个未知变量（约束力），因而需要求解的方程也相应增加（约束方程）。下面我将把（完整）约束问题纳入最小作用原理的框架内，并由此到处体系微分运动所满足的微分方程，这些方程相当于 Newton 方程在位形空间的s个独立方向上的投影，因此，约束越多，我们需要求解的方程就越少！这是在求解约束问题时最小作用原理的一个优势。而理解这一点的关键之处在于体系在位形空间中的运动是完全自由的！