

Chapter 1

引力的序曲

1.1 Poisson 方程

引力，是人尽皆知的吸引作用。每个人自蹒跚学步起便与之“抗衡”。我们身边随时随地都存在着这样一种力量，没有支撑的物体会因为它而下坠，有支撑的物体会因为它而稳定。即便如此，生活中还是有许多驾驭引力的高手，比如那些轻松掷中篮筐的职业球员；或者挥拍自如的羽毛球手，当然，他们大抵是不需要计算什么物理学的；再如鸟儿，它们和羽毛球手都得借助气流，但鸟儿更为厉害的是，它自身就天生有一对契合空气动力学的机翼，从而高耸入云、自由翱翔。但我们知道，气流在大范围上的无扰压强分布仍需借助引力而稳定，因此结论或许稍显意外：引力虽可能使鸟儿丧命，但也正是引力使得鸟儿脱身。

这皆拜引力所赐，它是如此地普遍，人们的文明自古以来便在思索其缘由，世界各族人民都给出了不同色彩的描述。面对落体运动，中国上古时期的人们将其解释为“物”趋向于回归到它“原本的地方”，这或许是对的，现在我们知道，引力大概也只有吸引作用，但不知若他们瞧见后来的孔明灯¹或者西洋人更大的热气球会作何解释。其实，正是引力造就了压强差。

受哲学及逻辑学的启蒙，西方世界对引力的描述要更为精准一些。面对自由落体，古希腊时期的 Aristotle 尝云，“置两个物体于同一高度释放，重的那一个将率先落地”。公元 5~6 世纪，一位希腊哲学家 Ioannes Philoponus 记载了这样一个发现：在同一点的各种物体都以完全相同的方式运动着，无论大小，无论重量。虽说同一“高度”不一定代表着“同一点”，但同一点很显然属于同一高度，这即对 Aristotle 学派的物理思想做出了挑战。时间来到 16 世纪，彼时的意大利弥漫着复兴的气息。一位研究自然的学者 Galileo 得知 Aristotle 学派所秉持的思想有难以弥补的悖论，即

¹ 不一定真为诸葛亮的“发明”，但文献显示孔明灯的使用年代确实很早。

所谓关于两个相互牵制的物体，下落加速应该归因于谁的问题。随后利用斜坡和单摆实验发现了下落距离正比于时间平方，且与其质量无关的规律。根据连续性思想以及忽略不必要阻力的思想，其可合理外推至自由落体。引入微积分学，Galileo 发现物体受引力作用的加速度之模应为一个常数 a_g 。因此在同一点的各种物体都以完全相同的方式被引力加速。他在西欧致力于传播这一事实，得到了广泛回应与讨论。后来人们知道，在高度变化过大时，其对加速度的影响亦即愈发明显了起来，故加速度只能在局部近似不变。经过数十年的思索与竞争，Newton 找到了统一地上“重力”学问和天上“引力”规律的方法，于 17 世纪末在他的《自然哲学的数学原理》[?] 中阐明了一个更为精确的规律：

$$\text{力} \quad \text{场源质量} \quad \text{质点起于场源的位矢}$$

$$\mathbf{F} = -\frac{GMm}{|\mathbf{r}|^3} \mathbf{r},$$

其中假定场源位于正交坐标系原点。结合 Newton 第二定律 $\mathbf{F} = \frac{d}{dt}(m\dot{\mathbf{r}}) = m\ddot{\mathbf{r}}$ ，两式中所带相同的 m 可约去，从而与之无关：

$$\ddot{\mathbf{r}} = -GM \frac{\mathbf{r}}{|\mathbf{r}|^3}. \quad (1.1.1)$$

平方反比律使我们联想到 Coulomb 静电力。在电磁学中 Faraday 提出了场 (*field*) 的概念。而 18~19 世纪，经由 Gauss、Maxwell 等人对场论数学的发展，场逐渐作为一种物理实质而在物理学里占据一席。我们可将场论借至力学里来。狭义上用第二定律定义的“力”只是一种“须要接触”的相互作用，而在宏观上不“接触”也具有相互作用的力适合用势能 ϕ 描述为²

$$\mathbf{F} = -m\nabla\phi.$$

因此，假设试验质点放置在恒稳的引力势 ϕ 中，这时经典的“引力”由上式给出。根据 Newton 第二定律，此即是说加速度是引力势的负梯度。类比到电磁学的电场强度 $\mathbf{E} = \mathbf{F}/q$ 上，加速度就亦可称作引力场强。引力势梯度与引力场强只差负号这一点，使得我们有

$$\nabla\phi = GM \frac{\mathbf{r}}{|\mathbf{r}|^3}. \quad (1.1.2)$$

类比电磁学中的电场 Gauss 定律，我们同样可研究曲面通量，这要用到其上的单位法矢 (*unit normal vector*) \mathbf{n} ，以下简称法矢，相应面元为 dS 。3 维区域一般可分为“一整块”的连通区域和“明显分开”的非连通区域两类。考虑到后者总可视作若干连通分支之和，不妨只考虑单个连通区域，其表面正是闭曲面。闭曲面将空间分为有界部分（即所包围区域）和无界部分，因此可区分空间的内外。一般偏好将法矢取

² 负号纯粹是习惯，但确实是个好习惯。

定为朝外, 称为外指 (*outward-pointing*) 法矢, 简称外法矢. 物理上最喜欢的区域还应是单连通的, 就是说, 其中任意形似“橡皮圈”的闭合曲线 (称简单闭线) 所围成的曲面都含于此区域中. 从直观上, 单连通区域不能像甜甜圈那样有“洞”, 但可以像稍厚的球壳状区域那样有“腔”. 若排除所有“腔”, 我们的单连通区域就形如球体, 其表面就称简单闭曲面. 微积分学的 Gauss 定理指出, 对任意使表面 ∂D 分段光滑³的单连通 D 及其上连续的场 $\nabla \cdot \mathbf{A}$ 有

$$\int_{\partial D} \mathbf{A} \cdot \mathbf{n} dS = \int_D \nabla \cdot \mathbf{A} dV, \quad (1.1.3)$$

我们借助它计算任意分段光滑的简单闭曲面通量. 以下默认 D 具此性质. 回到 (1.1.2) 式, 计算可知

$$\nabla \cdot (\mathbf{r}/|\mathbf{r}|^3) = 0, \quad \mathbf{r} \neq \mathbf{0}. \quad (1.1.4)$$

若 D 不含质点, 则 D 内被积函数连续, 可应用 Gauss 定理得

$$\int_{\partial D} \nabla \phi \cdot \mathbf{n} dS = 0.$$

若 D 包含质点, 由于 $\mathbf{r}/|\mathbf{r}|^3$ 有且仅有未定义点 $\mathbf{r} = \mathbf{0}$, 不能直接使用 Gauss 定理, 我们只得间接计算通量. 最简单地, 取一以质点为球心、半径大于零的任意球面 ∂B , 其上 \mathbf{n} 满足 $\theta = 0$ 且 $|\mathbf{r}|$ 为定值, 积分出

$$GM \int_{\partial B} \frac{\mathbf{r}}{|\mathbf{r}|^3} \cdot \mathbf{n} dS = 4\pi GM. \quad (1.1.5)$$

然而可以证明, 任意包含质点且分段光滑的简单闭曲面 ∂D 都应有

$$\int_{\partial D} \nabla \phi \cdot \mathbf{n} dS = 4\pi GM.$$

这是因为, 总可取一个囊括 ∂D 的球面 ∂B , 二者可围出一个球壳状区域 R . 记 ∂R 上的外法矢为 \mathbf{n} , 其通量仅由内表面 ∂D 和外表面 ∂B 贡献. 被积函数在单连通 R 上连续且为零, 故可应用 Gauss 定理得表面通量为零, 又因 ∂D 作为 D 表面时的外法矢为 $\mathbf{n}_D = -\mathbf{n}$, 而 ∂B 仍满足 $\mathbf{n}_B = \mathbf{n}$, 则

$$0 = \int_{\partial R} \nabla \phi \cdot \mathbf{n} dS = \int_{\partial B} \nabla \phi \cdot \mathbf{n}_B dS - \int_{\partial D} \nabla \phi \cdot \mathbf{n}_D dS,$$

故相等. 对如上过程有一更为直观的理解. 在微积分学里, 单位球中以圆心为顶点延伸的一个锥体 (一般选圆锥) 所截球面面积称为立体角 (*solid angle*) Ω , 其范围显然是 $[0, 4\pi]$. 记 \mathbf{r} 同 \mathbf{n} 间夹角为 θ , 注意 $\nabla \phi$ 与 \mathbf{r} 同向共线, 则该点贡献的通量为

$$\nabla \phi \cdot \mathbf{n} dS = GM \cos \theta dS / |\mathbf{r}|^2. \quad (1.1.6)$$

³简要回顾一下分析学概念. 某处连续指该处极限等于函数值, 记作 C^0 ; 一阶导函数在某处连续记作 C^1 (以此类推); 光滑指 C^∞ . C^1 一定说明可导 (该处导数存在), 反之不然. 以上概念对函数而言时表示定义域上恒成立.

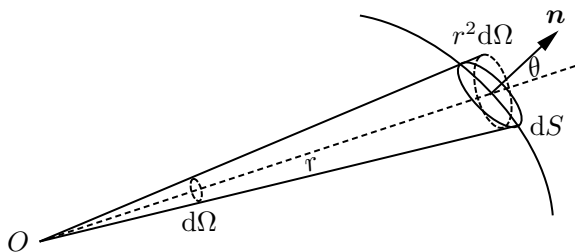


图 1.1: 立体角

现在我们取 dS 关于质点所张成的锥体, 根据如图 1.1 所示几何关系, $|\cos \theta| dS$ 就表示 \mathbf{r} 处球面方向上的投影. 立体角定义在单位球上, 则此球面投影就是按比例缩放的 $|\mathbf{r}|^2 d\Omega$, 故

$$\nabla \phi \cdot \mathbf{n} dS = GM \operatorname{sgn}(\cos \theta) d\Omega,$$

这里 sgn 表示取符, 即

$$\operatorname{sgn} x = \begin{cases} x/|x|, & x \neq 0; \\ 0, & x = 0. \end{cases} \quad (1.1.7)$$

若质点在 D 内, 其上任意外法矢相对质点朝外, 则 $\cos \theta$ 恒为正, 因此

$$\int_{\partial D} \nabla \phi \cdot \mathbf{n} dS = GM \int_{\partial D} d\Omega = 4\pi GM.$$

一旦质点在 D 外, 则可作 \mathbf{r} 切于曲面, 切点满足 $\cos \theta = 0$. 以 D 内某点位矢为转轴, 可对 D 截取不同剖面. 在截面所在平面上建立极坐标, 截面轮廓作为简单闭线应有幅角最值, 这便是满足 $\cos \theta = 0$ 的点. 可以预料, 从任意截面出发, 截取方向旋转半周后会截得相同截面, 故所有切点在空间中一定划过闭合图形, 至少是 ∂D 上的一条简单闭线. 记 $\cos \theta > 0$ 的点形成 ∂D_+ , 而 $\cos \theta < 0$ 的点形成 ∂D_- . 在任意截面上, $\cos \theta > 0$ 的部分和 $\cos \theta < 0$ 的部分所张幅角的边重合, 故 $\partial D_+, \partial D_-$ 所张椎体一致, 进而

$$\int_{\partial D} \nabla \phi \cdot \mathbf{n} dS = GM \left(\int_{\partial D_+} d\Omega - \int_{\partial D_-} d\Omega \right) = 0.$$

总之, 引入如下定义

$$\delta_r(D) = \begin{cases} 1, & \mathbf{r} \in D; \\ 0, & \mathbf{r} \notin D. \end{cases} \quad (1.1.8)$$

我们可整合地写为

$$\int_{\partial D} \nabla \phi \cdot \mathbf{n} dS = 4\pi GM \delta_r(D). \quad (1.1.9)$$

这里 $\delta_{\mathbf{r}}(D)$ 表示一个函数, 给任意 $D \subset \mathbf{R}^3$ 赋予一个实数. 不同于通常的关于点 $\mathbf{r} \in \mathbf{R}^3$ 的函数, 这里自变量 D 不是点而是集合. 注意, 上式表明, 我们并不奢求场源位于原点, 其位置可以任意, 而只关心是否在闭曲面里.

当引力源若干时, 令 M_1, \dots, M_n 含于曲面且总和为 M , 而 M_{n+1}, \dots, M_{n+m} 不含, 则

$$\begin{aligned} \int_{\partial D} \nabla \phi \cdot \mathbf{n} \, dS &= \int_{\partial D} \left(\sum_{i=1}^{n+m} \nabla \phi_i \right) \cdot \mathbf{n} \, dS \\ &= \sum_{i=1}^n \int_{\partial D} \nabla \phi_i \cdot \mathbf{n} \, dS + \sum_{i=n+1}^{n+m} \int_{\partial D} \nabla \phi_i \cdot \mathbf{n} \, dS \\ &= 4\pi G \sum_{i=1}^n M_i = 4\pi G M, \end{aligned}$$

说明闭曲面通量正比于其所包含的总质量. 记

$$\int_{\partial D} \nabla \phi \cdot \mathbf{n} \, dS = 4\pi G M(D). \quad (1.1.10)$$

这样对含 n 个质点的系统就是令

$$M(D) = \sum_{i=1}^{n+m} M_i \delta_{\mathbf{r}_i}(D) = \sum_{i=1}^n M_i. \quad (1.1.11)$$

像这种集合的函数其实称为测度 (*measure*), 当然, 也需满足一些条件: 空集的测度规定为零⁴; 一般要求测度非负; 可加性, 比如两非交图形之并的面积等于二者面积相加, 若有交则按集合论的容斥原理处理即可.

对非质点的物质模型而言, 通常可用密度 (*density*) 的积分衡量测度. 一点处的质量密度就是

$$\rho = \lim_{\Delta V \rightarrow 0} \frac{\Delta M}{\Delta V},$$

其中 ΔV 是该处的体元, 而 ΔM 是质元. 密度 $\rho(\mathbf{r})$ 分布连续的模型称为连续介质 (*continuum*). 我们将连续质量 M 分为若干质元 $\Delta M_i = \rho(\mathbf{r}_i) \Delta V_i$. 这样连续介质就相当于质点系统的连续极限, 即

$$M(D) = \lim_{\Delta V_i \rightarrow 0} \sum_{i=1}^n \rho(\mathbf{r}_i) \Delta V_i = \int_D \rho(\mathbf{r}) \, dV. \quad (1.1.12)$$

密度连续性使我们可用 Gauss 定理得

$$\nabla^2 \phi = 4\pi G \rho, \quad (1.1.13)$$

⁴ 还应根据实际意义具体规定. 比如对面积测度来说, 一条线段的面积仍为零, 因为面积这个概念相对于线段是超维的.

此即引力场的 *Poisson* 方程. 然而在引力作用下, 各质元 \mathbf{r}_i 往往与时间 t 有关, 与其关注某一质元的运动, 不如固定 \mathbf{r} 而考察密度场在此处的变化, 因此我们不会写 $\rho(\mathbf{r}_i(t))$, 而是写 $\rho(\mathbf{r}, t)$.

然而, 流体、固体等连续介质模型都仅是近似表述, 实际物体都是由粒子 (*particle*) 构成的, 这说明 ΔV 的任意小必将导致其中不含有任何粒子, 这样密度似乎处处为零了. 问题出在哪儿? 其实, 上述 ΔV 是一种平均概念. 物理学的连续介质方法研究的是其平均性质, 例如湖水所产生的引力场并不取决于单个水分子的状态, 而只依赖于大量分子的平均性质. 但平均性质在各点可以不同, 例如湖水底部的压强高于上层, 温度也随着深度变化, 再如大气密度亦随位置改变而改变. 那么多少粒子的平均才合适? 首先, 粒子必须足够多, 使得单个粒子的行为没有影响, 太少了就会出现类似密度为零的佯谬, 比如流体元包含大量粒子, 其中密度、平均速度等取值可视作相同. 然而粒子亦必须足够少, 以保证相对地均一, 即其中各处粒子速度、动能、粒子间距等不能相差过大, 进而就能描述点与点间性质的差异. 这种在某点合适地取平均的粒子邻域称为该处一个单元 (*element*). 密度定义中 ΔM 应视作遍及 ΔV 内众多粒子所对应的质量. 这在数学上不太清晰, 但对物理直觉很有用. 若这种单元不存在, 例如极度稀薄的气体或某些非平衡态系统, 则连续介质方法的根基就失效了, 则需另辟蹊径. 尽管如此, 数学上仍认为物理所用的微小单元其实无异于一点, 即所谓的连续介质方法就是给每一点直接赋予相应密度. 因此严格来说, 连续介质的质量是定义出来的: 我们先用密度场 $\rho(\mathbf{r}, t)$ 确定一个连续介质模型, 再定义其质量测度为密度积分.

此外, 对非连续的“密度”, 比如质点模型, 我们可借助 *Dirac* 函数以弥补 Gauss 定理使用的局限性. 回到前文的 $\delta_{\mathbf{r}}(D)$, 其实它称为 *Dirac* 测度. 尝试将其表为积分, 进而就可使质量测度表为密度积分. 定义这样一种仅在原点处不连续的函数 $\delta(\mathbf{r})$ 满足

$$\int_D f(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}_i) dV = f(\mathbf{r}_i) \delta_{\mathbf{r}_i}(D), \quad (1.1.14)$$

其中 f 是光滑的. 这就是 Dirac δ 函数, 这一表述称为挑选性. 这样质点系满足

$$M(D) = \int_D \sum_{i=1}^n M_i \delta(\mathbf{r} - \mathbf{r}_i) dV,$$

因此总密度可表为

$$\rho(\mathbf{r}, t) = \sum_{i=1}^n \rho_i(\mathbf{r}, t) = \sum_{i=1}^n M_i \delta(\mathbf{r} - \mathbf{r}_i(t)). \quad (1.1.15)$$

对任意不包含 \mathbf{r}_i 的 D , 取 $f = 1$ 可知 $\delta(\mathbf{r} - \mathbf{r}_i)$ 的体积分都为零, 由于 $\delta(\mathbf{r})$ 在这样的 D 中连续, 只能意味着物理上的密度在非质点处为零, 即

$$\mathbf{r} \neq \mathbf{r}_i \Leftrightarrow \delta(\mathbf{r} - \mathbf{r}_i) = 0, \quad (1.1.16)$$

因此 $\delta(\mathbf{r})$ 是偶函数. 根据此前对 (1.1.5) 式的讨论可知

$$\int_{\partial D} \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} \cdot \mathbf{n} \, dS = 4\pi\delta_{\mathbf{r}_i}(D) = \int_D 4\pi\delta(\mathbf{r} - \mathbf{r}_i) \, dV. \quad (1.1.17)$$

因此为从形式上满足 Gauss 定理

$$\int_{\partial D} \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} \cdot \mathbf{n} \, dS = \int_D \nabla \cdot \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} \, dV,$$

应有

$$\nabla \cdot \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3} = 4\pi\delta(\mathbf{r} - \mathbf{r}_i). \quad (1.1.18)$$

至此, 凡涉及质点密度时就可借助 Dirac 函数应用 Gauss 定理.

Dirac 最先提出 δ 函数, 并用其解决了许多物理问题, 但须注意其定义显然不严格. 令 $D = \mathbf{R}^3$ 可知 δ 函数的全空间积分为 1. 但 (1.1.16) 表明 δ 几乎处处⁵为零, 其全空间积分只能为零, 矛盾. 此外, $\delta(\mathbf{0})$ 其实没有定义, 即密度在点模型中发散. 可任意构造一个无穷的函数序列 $\{\delta_i\}$ 逼近至 δ : 对每一 δ_i 可要求上述全空间积分归一, 譬如选择统计学中的正态分布, 令分布的平均值为零, 当其方差趋于零时 (对应编号 $i \rightarrow \infty$) 就可满足 (1.1.16). 这样 $\delta(\mathbf{0})$ 只能为 “正无穷”. 可见 δ 并非一个普通意义的函数. 数学界开始时拒绝接受它, 但这些漏洞却似乎并不影响演算, δ 函数屡战屡捷, 总是能得正确的物理结果. 这引起了某些数学家的兴趣, 于 1940s 提出了广义函数 (*generalized function*) 理论 (具体见附录 ??), 从而把 δ 函数置于严格基础之上. 这样 δ 函数的积分性质在广义函数意义下可以成立, 演算的严格性得以保证. 简单来说, 把象征性的记号 $\delta(\mathbf{r})$ 看作函数虽然不对, 但物理学家并非孤立地看待它, 而是连同其积分来使用, 最后都要把它置于积分号内. Dirac 在研究量子力学时往往取 $D = \mathbf{R}$, 因此

$$\int_{\mathbf{R}} f(x)\delta(x - x_0) \, dx = f(x_0),$$

这样 δ 函数积分就失去了测度意义, 它只被 Dirac 用来将函数 $f(x)$ 映射为实数, 即其中的

$$\int_{\mathbf{R}} (\cdot)\delta(x - x_0) \, dx = \delta_{x_0}(\cdot)$$

看作一个映射, 输入任一函数 f 就输出一个实数 $f(x_0)$. 描述一个函数所需的, 就是给定义域内每一点赋予实数. 一点由 1 个独立实数描述, 称自由度为 1, 然而定义域内的点无限多, 因此一个函数具有无穷自由度. 数学家们给这种自变量并非分立实数的映射取名为泛函 (*functional*), 即所谓函数的函数. 泛函的自变量称为宗量. 以后凡用独立的 δ 函数, 只需注意, 实质是先讨论测度意义 (是否在区域内), 再利用其泛函意义 (得到积分值), 这样 δ 函数的代数演算就是严格的.

⁵ 几乎处处指例外情况在同一标准下比较时, 测度为零. 这个例子里, δ 函数为零的范围明显有发散体积, 而例外点 $\mathbf{r} = \mathbf{0}$ 的 “体积” 为零. 微积分学告诉我们, 函数值只在有限点改变甚至无意义时, 并不影响可积性和积分值.

1.2 Newton 引力的缺陷

Newton 引力取得了巨大成功，但其局限性亦十分显著。继续扩大高度，把视线放到更远的距离去，一个明显的问题是，经典引力将成为一种能传播无限远的超距作用。Newton 本人拒绝这种看法，但也不能提供合理的物理实质，而始终认为其使用平方反比律只是在“数学地描述”吸引现象。那么在很远的地方，引力还将是平方反比吗？想要解决超距作用，一种思想便是想象整个空间弥漫着一种媒介性质的物质，也就是上节所说的场，它可以将作用传播到很远很远的地方。然而，读者容易验证 Poisson 方程有如下一种解：

$$\phi(\mathbf{r}, t) = -G \int_D \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} dV',$$

只需借助 δ 函数挑选性即可，并注意

$$\nabla(1/|\mathbf{r}|) = -\mathbf{r}/|\mathbf{r}|^3, \quad \mathbf{r} \neq \mathbf{0}. \quad (1.2.1)$$

此式表明 \mathbf{r} 处 t 时的 ϕ 同时由空间各点在 t 时的 ρ 决定，依然说明引力场传播无限快。由于默认任意距离满足平方反比，场的引入不能根治其超距作用的矛盾。即使如此，想要修改经典引力理论，借助场这个模型依旧有希望。只要先不承认平方反比的正确性，而是企图直接研究引力场的物理性质，这样就有望研究出一些比较整体的东西出来，而在较弱情况还原到平方反比去。场中相互作用的传播速度总该是一个有限数（但可能是一个没有上限的有限数）。人们自然想到另外一种场：19 世纪，人们已经发现了其中波以有限速度传播的场——电磁场。我们得考虑从电磁场上抽象出一点性质出来。在 ?? 节将会知道，所选择出来的抽象性质是相对性原理 (*principle of relativity*)。

在很近的距离下，虽然 Newton 引力传播确实可视作瞬间，但却出现了另一与事实误差很大的问题：19 世纪末，水星近日点进动现象被首次观测。毫无干扰的情况下，Newton 理论当然不能解释进动，因此必须要考虑其它因素，如同当年从天王星轨道扰动发现的“笔尖下的行星”海王星那样。由于地球自转轴

的进动，从地球上观测水星的进动大部分都是所谓的岁差，但这纯粹是参考系的视觉效果，因此我们就先扣除掉它。那么在经典引力下，“其它因素”就是其余行星造成的摄动及太阳实际形状了。然而后者相对影响小得多，我们可以只研究前者⁶。以太阳为参考，水星所处的外引力场写作

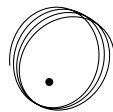


图 1.2: 轨道进动 (离心率和进动角均夸大)

$$\mathbf{F}(\mathbf{r}, t) = \mathbf{F}_{\odot} + \mathbf{F}_p = -\frac{GM_{\odot}m}{r^3}\mathbf{r} - Gm \sum_i \frac{M_i(\mathbf{r} - \mathbf{R}_i)}{|\mathbf{r} - \mathbf{R}_i|^3},$$

⁶关于更精细的考虑，参见 Goldstein 《经典力学》第 5~8 章，以及相关的天体力学书籍。

前一项是太阳的贡献，后面是各行星的作用总和。现在研究后者。设 \mathbf{r}, \mathbf{R}_i 夹角 θ_i ，则

$$d_i = |\mathbf{r} - \mathbf{R}_i| = R_i^2(t) + r^2 - 2R_i(t)r \cos \theta_i(t),$$

忽略其余行星轨道的离心率之影响，即假定做半径 \mathbf{R}_i 的圆周运动。设单个行星角速度为 ω_i ，将 \mathbf{F}_p 在时域上做 Fourier 展开

$$\mathbf{F}_p = -Gm \sum_i \sum_{n \in \mathbb{Z}} a_n(r) e^{in\omega_i t} \frac{\mathbf{r} - \mathbf{R}_i}{|\mathbf{r} - \mathbf{R}_i|},$$

其中系数为

$$a_n = \frac{\omega_i}{2\pi} \int_0^{2\pi/\omega_i} \frac{M_i e^{-in\omega_i t} dt}{R_i^2 + r^2 - 2R_i r \cos \theta_i}.$$

只保留最低阶的 $n = 0$ 并用 $d\theta_i$ 替换 $\omega_i dt$ 就可使引力势同时间无关，得到近似表达

$$\mathbf{F}_p(r) = -Gm \sum_i \int_0^{2\pi} \frac{M_i d\theta/2\pi}{R_i^2 + r^2 - 2R_i r \cos \theta} \frac{\mathbf{r} - \mathbf{R}_i}{|\mathbf{r} - \mathbf{R}_i|}, \quad (1.2.2)$$

引入线密度 $\lambda_i = M_i/(2\pi R_i)$ ，并沿水星初始方向为极轴，在黄道面上建立极坐标系 (r, θ) ，便发现上式表明，其余行星对水星的作用相当于 M_i 均匀分布的一个圆环。这符合物理直觉，毕竟水星的进动极为缓慢，完整进动一周所需时间相较于其它行星周期很长，因此其它行星的作用的确和圆环类似。但对离水星较近的行星（比如金星），二者周期将在同一数量级，因此不便于近似为圆环。可见圆环模型成立条件是 $R_i \gg r$ 。幸好，除金星外其它行星都离得较远，将金星视作圆环的影响并不致命。问题现在相当于，在太阳引力基础上，计算附加这样一个微扰后行星的进动。由圆环对称性，可知作用力与 \mathbf{r} 同向，即

$$\mathbf{F}_p = Gm \frac{\mathbf{r}}{r} \sum_i \lambda_i R_i \int_0^{2\pi} \frac{\cos \alpha d\theta}{d_i^2},$$

可见其他引力的作用可视作朝外的有心力，我们就可直接讨论分量

$$F = F_\odot + F_p.$$

我们希望先计算出 F_p 与 r 的关系，虽无初等表达，但由于该式本身就具条件 $R_i \gg r$ ，故可近似计算。一个聪明的做法并非按 r/R_i 展开，而是重构模型本身。将 (r, θ) 平移以使原点与水星重合，记新坐标 (d, α) ，根据余弦定理

$$R_i^2 = r^2 + d_i^2 + 2rd_i \cos \alpha,$$

由于 $\alpha = 0$ 时 $d_i = R_i - r$ ，取解

$$d_i = -r \cos \alpha + \sqrt{R_i^2 - r^2 \sin^2 \alpha}.$$

由 $R_i \gg r$ 知圆环可近似看作按 α 均匀分布, 故 $R_i d\theta \approx d_i d\alpha$, 于是

$$\begin{aligned} \int_0^{2\pi} \frac{R_i \cos \alpha d\theta}{d_i^2} &\approx \int_0^{2\pi} \frac{\cos \alpha d\alpha}{-r \cos \alpha + \sqrt{R_i^2 - r^2 \sin^2 \alpha}} \\ &= \int_0^\pi \left(\frac{\cos \alpha}{-r \cos \alpha + \sqrt{R_i^2 - r^2 \sin^2 \alpha}} \right. \\ &\quad \left. - \frac{\cos \alpha}{r \cos \alpha + \sqrt{R_i^2 - r^2 \sin^2 \alpha}} \right) d\alpha \\ &= \frac{2r}{R_i^2 - r^2} \int_0^\pi \cos^2 \alpha d\alpha = \frac{\pi r}{R_i^2 - r^2}, \end{aligned}$$

即

$$F_p = Gm\pi r \sum_i \frac{\lambda_i}{R_i^2 - r^2}. \quad (1.2.3)$$

接下来只以初始位置 $(r_0, 0)$ 为极轴方向建立极坐标, 以研究水星运动. 有心力的形式为

$$m(\ddot{r} - \dot{\theta}^2 r) = F(r), \quad 2\dot{r}\dot{\theta} + \ddot{\theta}r = 0,$$

第二式导致角动量守恒, 记 $mr^2\dot{\theta} = L$. 两式消 t 得 *Binet* 式

$$\frac{d^2\mu}{d\theta^2} + \mu = -\frac{mF}{L^2\mu^2} = J(\mu). \quad (1.2.4)$$

如果是圆轨道模型, 则

$$\frac{d\mu}{d\theta} = 0, \quad F_0 = -\frac{L^2}{mr_0^3} \Rightarrow \mu_0 = J(\mu_0).$$

然而圆轨道闭合而不一定稳定. 稳定即是说, 物体运动受到微扰后依然束缚在圆轨道附近. 下面寻找稳定条件. 在 μ_0 附近展开 $J(\mu)$, 记 $x = \mu - \mu_0$, 只保留至一阶可得波动方程

$$\frac{d^2x}{d\theta^2} + \left(1 - \frac{dJ}{dx}\bigg|_0\right)x = 0.$$

如果括号为负, x 将随 θ 指数增长或衰减, 显然不再稳定. 因此稳定要求括号为正, 则 x 将随 θ 作周期振荡. 代入 $J(\mu)$ 定义得

$$\frac{dJ}{dx}\bigg|_0 = \frac{d}{d\mu}\bigg|_{\mu_0} \left(-\frac{mF}{L^2\mu^2}\right) = \frac{2m}{L^2\mu_0^3}F_0 - \frac{m}{L^2\mu_0^2} \frac{dF}{d\mu}\bigg|_{\mu_0} = -2 - \frac{r_0}{F_0}F'(r_0).$$

这里撇号表示对 r 求导. 则稳定轨道要求有心力的形式满足

$$1 - \frac{dJ}{dx}\bigg|_0 = 3 + \frac{r_0}{F_0}F'(r_0) > 0. \quad (1.2.5)$$

在波动方程中, 这个括号的含义是频率平方, 注意其中是对 θ 求导而非 t , 故应是

$$\frac{2\pi}{\Delta\theta} = \sqrt{3 + \frac{r_0}{F_0} F'(r_0)}.$$

其中 $\Delta\theta$ 的含义就是一次微扰振荡周期

$$T_r = \frac{\Delta\theta}{\dot{\theta}} = \frac{\Delta\theta m r_0^2}{L} = \Delta\theta \sqrt{-\frac{F_0}{m r_0}}$$

所扫过的角度. 可见只有 F_\odot 时使 $\Delta\theta = 2\pi$, 相当于又回到原位而无进动. 而一旦错开, 一次周期内就出现了进动. 进动率为

$$\omega = \frac{\Delta\theta - 2\pi}{T},$$

式中 T 为恒星年 (*sidereal year*)⁷. 现在代入 F_\odot, F_p , 则

$$\Delta\theta = 2\pi \sqrt{\frac{F_\odot(r_0) + F_p(r_0)}{F_\odot(r_0) + 3F_p(r_0) + r_0 F'_p(r_0)}}, \quad (1.2.6)$$

其中

$$F'_p(r_0) = Gm\pi \sum_i \lambda_i \frac{R_i^2 + r_0^2}{(R_i^2 - r_0^2)^2}. \quad (1.2.7)$$

为更便于计算, 可注意 $F_p(r_0) \approx r_0 F'_p(r_0) \ll F_\odot(r_0)$, 进而展开至 $F_p(r_0)/F_\odot(r_0)$ 和 $F'_p(r_0)/F_\odot(r_0)$ 的一阶. 首先改写为

$$\Delta\theta = 2\pi \sqrt{\frac{1 + F_p(r_0)/F_\odot(r_0)}{1 + (3F_p(r_0) + r_0 F'_p(r_0))/F_\odot(r_0)}},$$

分子分母都展开为

$$\Delta\theta = 2\pi \left(1 - \frac{3F_p(r_0) + r_0 F'_p(r_0)}{2F_\odot(r_0)} \right) \left(1 + \frac{F_p(r_0)}{2F_\odot(r_0)} \right),$$

化简, 忽略二阶项有

$$\Delta\theta = 2\pi \left(1 - \frac{F_p(r_0)}{F_\odot(r_0)} - \frac{r_0 F'_p(r_0)}{2F_\odot(r_0)} \right), \quad (1.2.8)$$

代入数据可知 ω 约 $532''/\text{世纪}$, 仍然与观测值 $575''/\text{世纪}$ 相差 $43''/\text{世纪}$. 可见, 经过一系列类似尝试后, 很难从 Newton 理论精确地计算出进动角的实际观测结果, 总

⁷ 某行星的恒星年是太阳在黄道上的视位置相对遥远恒星变化一周所经历时间, 也相当于该行星相对于遥远恒星公转一周所经历的时间. 千禧年时地球恒星年的平均长度为 365.25636 天. 回归年 (*tropical year*) 是太阳的平黄经变化一周所经历时间. 由于太阳视运动非均匀, 选取不同的起点会得到不同的长度. 若选取春分点作为基准, 太阳从春分点出发运行一周再次回到春分点的时间为春分点年. 千禧年时回归年的平均长度为 365.24219 天. 恒星年比回归年要长 20 分 24.5 秒, 这个差异正是地球造成的岁差.

是无法用 Newton 引力算出那额外进动。这说明，水星由于离太阳“很近”，以至于其所处的引力场强烈到不符合平方反比律。修改 Newton 引力理论仿佛迫在眉睫。

引力在很大的尺度上有什么效果？前文提到，要想从局部的现象揣测出宏观的性质，似乎得从电磁场本身抽象出一种性质，并将其移用到引力场上。当时的情况是，人们已经普遍接近于抛弃经典力学中的相对性原理：物理规律在所有惯性系中相同（详见 ?? 节），企图寻找一种可以与其它惯性系相区分的特殊惯性系——“以太”，作为包括光在内各种电磁波的传播介质，进而说明 Maxwell 的方程只适用于这种参考系。随着光速测量实验的进展，人们逐渐意识到以太理论难以克服的困难。Einstein 本人在 1946 年的《自述》(*Autobiographisches*) 中说，他不太记得光速测量实验是否影响了他的思索，但他一心想要“调和经典力学和电动力学”，选择保留相对性原理。如果是这样，那么问题就只能出在经典力学的另一根基——Newton 的绝对时空观⁸上。

为理解这件事，我们必须再次回到现代物理学的起源。在西方文化中，有两种理解“空间”的传统方式：作为一个实体 (*entity*) 或作为一种关系 (*relation*)。

- “空间是一个实体”意味着当除了空间之外没有其他东西时，空间仍然存在。它本身就存在，物体在其中运动。这是 Newton 描述空间的方式，称为绝对空间。后面我们将强调，这也是狭义相对论中理解时空（不是空间）的方式。尽管自古以来（在民主传统中）就被认为是这样，但这种理解空间的方式并不是西方文化中的传统主导观点。从 Aristotle 到 Descartes，主流观点是把空间理解为一种关系。
- “空间是一种关系”意味着世界是由物理物体，或物理实体组成的。这些物体具有这样的属性：它们可以彼此接触，亦可不接触。空间就是物体之间的这种“接触”、“毗连”或“紧邻”的关系。例如，Aristotle 将一个物体的空间位置定义为围绕它的物体集合的（内部）边界。这就是关系空间。

与此二理解空间的方式严格相联系的，有对应两种理解运动的方式。

- 绝对运动 (*absolute motion*): 如果空间是一个实体，运动可以被定义为从空间的一个部分到空间的另一个部分。这就是 Newton 对运动的定义。
- 相对运动 (*relative motion*): 如果空间是一种关系，运动只能被定义为从一个物体的紧邻性到另一个物体的紧邻性。这就是 Descartes⁹ 和 Aristotle¹⁰ 对运动的定义。

⁸Newton 本人主张绝对空间观而其它哲学家则不然，但在 Einstein 之前，几乎所有人都支持绝对时间观，其中 Poincaré 和 Lorentz 在观念上的革新仍不够彻底。

⁹“我们可以说，运动是物质的一部分或物体的一部分，从与它紧邻的、被认为是静止的那些物体附近转移到其他一些物体附近。”——Descartes, 《哲学原理》[?], 第 II-25 节, 第 51 页

¹⁰Aristotle 坚持认为，运动是相对的。他用一个人在船上行走的例子来说明这一点。这个人是相对于船运动的，而

物理学家的问题在于，这两种关于空间和运动的思维方式中，哪一种可以更有效地描述世界。对 Newton 来说，空间是绝对的，运动也是绝对的¹¹。这是对 Descartes 主义的第二次违反。当然，Newton 也并不是怀揣轻松的心态迈出这一步的：他在《原理》中用了很长的开头部分来解释他的选择的原因。对 Newton 有利的最有力论据完全是马后炮：他的理论结构当然运行得特别好，而 Descartes 的物理学从未如此有效。但这不是 Newton 的论点。

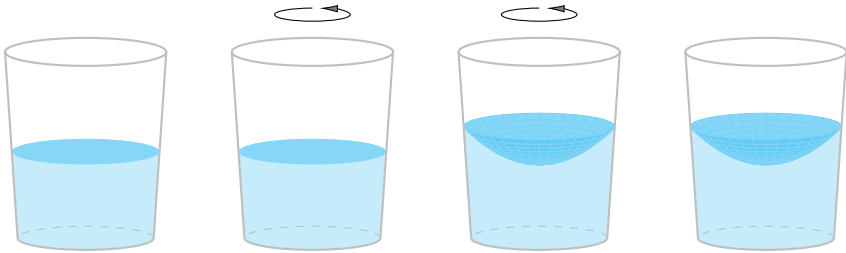


图 1.3: 水桶实验

Newton 采用经验证据，讨论了一个著名的水桶思想实验。将一盛有清水的桶静置，并用一根长绳悬吊（图中略去）使桶可以靠绳索的扭力而转动；随后，尽可能使桶很快地加速旋转至某一角速度 ω 并保持。起初，桶相对于我们旋转，但（大部分）水由于惯性仍将保持静止，并且液面是平坦的；待桶的运动通过摩擦力传导至水，使水开始渐渐地随桶旋转，液面逐渐凹陷。在某个时刻，水和桶都将以 ω 旋转，即相对静止，液面凹陷呈抛物面状：用等势面可算出液面方程在柱坐标系下为 $z = \omega^2 r^2 / 2a_g$ ；尽可能很快地停止旋转，这个时候水由于惯性仍然要旋转一会儿才会静止，液面仍然凹陷。虽说是思想实验，但其可行性还是挺大的：Newton 本人就声称“已经做过实验”，并对该结果表示肯定。

问题在于，究竟是什么原因使得液面凹陷？我们从经验中知道，可能是水在相对于某个参考系旋转，进而产生了离心作用，但这个参考系是谁呢？Newton 的水桶实验显示了该问题的微妙之处：如果像 Descartes 所要求的那样，运动是相对于周围物体的位置变化，那么我们必须说，起初水相对于环绕它的桶在旋转，但随后水与桶一同旋转，那么水相对于桶就是静止的。如果水的凹陷是因为旋转，那为什么当水与桶

船又是相对于河水运动的，而河水又是相对于地面运动的……Aristotle 的关系主义因以下事实而有所缓和：有一些可以作为首选参考的物体：位于“宇宙中心”的地球，以及天球（即星空，共同组合为天球音乐），特别是其中一颗固定的星星。因此，我们可以说，如果某物相对于地球运动，它就是在“绝对地”运动。然而，在古代宇宙学中，有两个首选框架：地球和固定恒星，而且这两个框架是相对于对方旋转的。中世纪的思想家们并没有错过这一点，他们详细讨论了恒星是围绕地球旋转还是地球围绕恒星旋转。值得注意的是，在 14 世纪，Buridan 基于理性的理由得出结论，两种观点都不比另一种更真实，而 Oresme 研究了地球的旋转，比 Copernicus 的日心说早了一个多世纪。

¹¹ “所以，只要空间被视为真正有别于运动的物体，那么就有必要把位置的定义和局部的运动，认定为是关于一些不运动的东西而言的，如单独的延伸或空间。”—— [?]，这与脚注 9 中 Descartes 的定义形成了鲜明对比。

相对静止时，液面却是凹陷的？这样从 Descartes 的思想就似乎有一个矛盾，表面的凹陷反而出现于“静止”而非“旋转”。因此，产生凹陷的旋转不是相对于水桶的旋转。那它还能相对于什么呢？是我们观测者还是……Newton 回答道，它是相对于空间本身的旋转。水面的凹陷是水的绝对运动的效果，即相对于绝对空间的运动，而非相对于周围物体的运动。Newton 企图利用这个实验说明绝对空间的存在性，即如果水在其中旋转会凹陷时，这个空间就是绝对背景，而“不受人为因素所决定”。

为正确理解它，我们应该阐明一个常见误解——不应该将相对主义，即运动只能在与其它物体的关系中被定义的想法与 Galileo 相对性相混淆。Galileo 相对性是指“匀速直线运动”先验地与静止无异的说法。也就是说，速度是相对于其他物体而言的；而另一方面，关系主义却认为，任何运动（无论如何曲折）都是先验的，与静止无法区分。Galileo 相对性的表述本身就假定了运动的非关系性定义。“直线和匀速”是相对于什么而言的呢？这应当视作惯性系的定义。

Newton 声称相对于绝对空间的运动是真实且物理的，某种意义上他做得过头了。他坚持认为即使是直线均匀运动也是绝对的，这引起一场激烈争论，因为惯性运动没有物理效应，因此，对这一类特殊的运动来说，水桶实验是失败的¹²。Newton 的动力学基础所需要的——亦正是这里讨论的——不是惯性运动的相对性：而是加速运动，如水桶里的水的旋转，是相对的还是绝对的。问题不是“是否存在一个其中速度可以定义的绝对空间”，而是“是否存在一个其中加速度可以定义的绝对空间”。Newton 的答案在水桶实验的支持下是肯定的。如果没有这个回答，Newton 第二定律将不再有意义。

三个世纪以来，没有人能够彻底驳斥 Newton 的论点。期间先后有诸多科学家和哲学家试图批判绝对空间，甚至强烈地反对超距引力的反对。Leibniz 和他的学派关于 Newton 的绝对运动及其所用的绝对加速度进行了激烈的争论¹³。随后，人们的疑虑从未真正消失过，一种挥之不去的感觉仍然存在，总觉着 Newton 的论证有问题¹⁴。19 世纪末，Ernst Mach 回到了这个问题上，做出了显赫的一步，他认为，水并非相对于某种绝对介质旋转，而是相对于宇宙的全部物质组分（星空背景）。我们知道，水凹陷并趋于溢出应当源于水的惯性，在旋转的参考系中出现了惯性力。它是所谓“运动学的部分”，而反之，假使真的有能力让“所有物质”都绕着水桶旋转，从相对性的推广上讲，确实应当产生同样的效果，尽管这个实验确实很难做。

我们甚至可更为激进地假设：一个毫无其它物质的空间中有一颗弹性球，由于没有任何参照，它的旋转便无法定义，那么随之产生的惯性效果显然不会出现。随着

¹²Newton 很清楚这一点，这在《原理》的推论 V 中有明确的说明，但他在导言中选择忽略这一点。或许他这样做只是为了简化他的论证，这对他同时代的人来说已经够难了。

¹³当然 Leibniz 还有其他抱怨的理由，比如两人在争夺微积分发明的优先权。看来历朝历代的科学家都有这个弱点。

¹⁴对于远距作用，Newton 主义在经验上的巨大胜利是难以磨灭的。或者还是可以“磨灭”的？毕竟前面讲过，Newton 的理论无法解释额外进动值。

我在周围加入的物质越多，它的旋转也容易通过参照而观测，进而才体现出了惯性（如逐渐扁平），直到我加入了“所有物质”，其惯性便“完全”体现了。Mach 如是说，“惯性需要在物质的相对运动中体现”，以驳斥绝对空间。与物质质量联系的“动力学的部分”此刻产生了惯性力的效果。随“物质的加入”而“愈发明显的”作用，我们知道，已经有一个被称为引力的相互作用具有极其相似的性质。从简而言，惯性（力）和引力确实都与“质量”的大小成正比。但是它们是正比于相同的“质”量吗¹⁵？

让我们回到 (1.1.1)，这里约去了相同的 m 。这本应只是 Newton 的搪塞，默认了相同而回避了进一步的哲学思索，但实际上我们完全可以据此定义出两种不同的质量出来：18~19 世纪，已经有许多物理学家认识到了这一点，即第二定律中定义的应叫惯性质量 (*inertial mass*):

$$\mathbf{F} = m_I \ddot{\mathbf{r}};$$

而万有引力定律中定义的应叫引力质量 (*gravitational mass*)¹⁶:

$$\mathbf{F} = -m_G \nabla \phi,$$

无心插柳柳成荫，这一点被 Einstein 敏锐地察觉到了，随后于 1907 年决定在今后寻找相对论囊括引力的方法时，贯穿这样一条准则：引力质量与惯性质量精确地成正比，且比例与物质本身无关。因此，选取适当的单位制，可使引力质量等于惯性质量，那此二质量便可不再作区分。这便萌芽了后人所称的等效原理 (*equivalence principle*)。正是等效原理保证了引力场的存在性和对试验质点的独立性。我将在 ?? 以及 ?? 节更加深入地讨论这些观点及其对 Einstein 的影响。

总之，Einstein 对水桶实验的新解释很简单，也很有说服力：

水相对于一个局域的物理实体旋转：引力场。

是引力场“告诉”物体是否在加速或旋转，而非绝对空间。没有像绝对空间这样的呆滞的背景实体，只有动态的物理实体。这些实体中包括场。在这些场中存在引力场部分。Newton 的水桶中液面的平坦或凹陷并非由水相对于绝对空间的运动决定的，它是由水和引力场之间的物理互动决定的。Einstein 关于引力的两条思路（为 Newton 引力找到场论、摆脱绝对加速度）在此相遇。关键在于，Newton 把重力场误认为是一个绝对空间。

那是什么使得 Einstein 有了这一想法？为什么 Newton 的加速度要相对于引力场来定义？Einstein 所关注的引力的特殊属性——等效原理给出了这个答案¹⁷。这一

¹⁵或者说，总是可以假定是同“种”质量，但它们的比例是否对任意物体都是一致的？

¹⁶Newton 力学中还将其分为主动 (active) 与被动 (passive) 两类。前者指作为场源时的质量，后者指作为试探质点的质量。本文此处指后者。

¹⁷引力是“特殊”的，因为 Newton 的绝对空间是引力场的一个构型。一旦我们摆脱了绝对空间的概念，引力相互作用就不再是特别的了。它是形成世界的场之一。但这与 Newton 和 Maxwell 的世界不太一样。

深邃的思想说明了，任何物质在引力场中应该具有相同的加速度，就像当年 Galileo 所说的那样。恰恰是 Galileo 所研究的那“矮”的局部现象，才能从引力的整体现象中剥离出重要的特征出来。

请把这纷繁的世界想象成一个“舞台”，物质在其中“舞蹈”着。现在让我们引入某种力场，对舞台中同一“位置”，任何“舞者”都将感受到相同的由力场带来的加速度。这很可能说明这个力场本身就是这“舞台”的内在性质，要研究这个力，自然就要研究这个“舞台”的几何学 (*geometry*)。

我们知道，惯性力的引入是为了使 Newton 定律在非惯性系中保持形式。也就是说，在选择了一个“不正确的参考系”的情况下引入一个“起源不明的力”解释运动学。这就是所谓的赝力 (*pseudo force*)。这些力的出现，只是因为观者没有站在惯性系的角度去观测。既然赝力和引力都与“质量”成正比，那也就有可能将引力本身解释为一种赝力。换种说法就是，引力是由于没有选对参考系而引起的。

Einstein 首先把这件事解释为对于从惯性系到非惯性系的坐标变换的一种协变性 (*covariance*)。我们通过其电梯思想实验来揭示它，它直接导致了支撑广相的物理直觉。

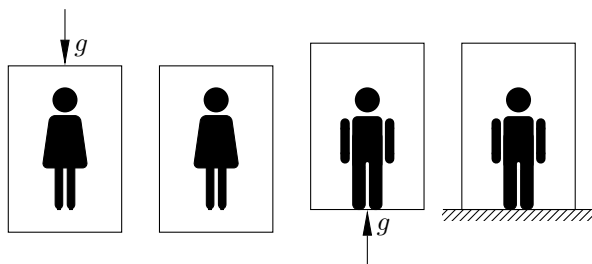


图 1.4: 电梯实验

设想一部封闭的电梯¹⁸和一位勇敢的女乘客。起初，这部电梯是由缆绳牵制而静置于空中的，乘客发现始终有一种作用让他待在电梯的地板上，这个力与它的体重成正比。我们可以说，这是乘客测量到了地球所产生的恒稳的引力势场

$$\phi(h, t) = a_g h,$$

但若缆绳不幸断裂，这部电梯将顷刻间具有相对大地朝下的加速度 a_g ，这样，原来的电梯就从 h 变到加速下坠的电梯参考系

$$\tilde{h} = h + \frac{1}{2}a_g t^2. \quad (1.2.9)$$

¹⁸Einstein 本身是瑞士专利局职员，其很多思想实验都与电机机械有关。为了描述一种自由落体的参考框架，他自然便会想到类似方型箱匣的事物：一部断线的有些骇人的电梯，并且当时电梯恰恰刚诞生不久。

在新的参考系下, Newton 第二定律就可以写成

$$\ddot{\mathbf{r}} = \mathbf{0}. \quad (1.2.10)$$

这说明, 乘客就似乎失重而漂浮了起来. 也可以把这个过程颠倒过来看. 假如根本没有地球, 飞船自由地漂浮, 那么男乘客也会同样漂浮其中, 自然也就没有出现什么引力现象. 此时若施加一个加速度 a_g 向上拖动电梯, 那么乘客应当会发现一个赝力, 如同引力一样把人拉向地板. 这样我们就成功在一个加速的参考系中用 (1.2.9) 式来“模拟”引力场. 关键在于, 在这些实验中, 乘客并不知道外界是否有地球, 亦或电梯 (飞船) 是否加速, 那么赝力和引力对其而言是难以区分的, 说不清楚总的作用之中“有哪些是惯性力, 有哪些是引力”.

下面将用 Newton 宇宙论再介绍另一个更现代、更现实的“电梯实验”. 这个模型很简单. 在 Newton 物理学的背景下, 考虑一个由非常大的球形星系云组成的宇宙. 假设这些星系均匀地分布在空间中, 并保持着与时间有关的密度 $\rho(t)$, 星系间在引力作用下相互吸引. 设 C 是云的中心. 考虑一个星系 A (例如, 咱们的银河系), 与中心 C 的距离为 $r(t)$. 众所周知, C 周围半径为 r 的球体外的星系对 A 的引力互相抵消, 而该球体内的星系的引力与集中在 C 内的相同质量的质点的引力效果相同. 因此 A 所受引力为

$$\mathbf{F} = -G \frac{m_A \frac{4}{3}\pi |\mathbf{r}(t)|^3 \rho(t)}{|\mathbf{r}(t)|^3} \mathbf{r}(t),$$

结合第二定律, 此即

$$\ddot{\mathbf{r}} = -\frac{4}{3}G\pi\rho(t)\mathbf{r}(t).$$

若质量在空间上保持常数, 则其密度标度将一致依赖于 $|\mathbf{r}|^{-3}$, 即 $\rho(t) = \rho_0 |\mathbf{r}(t)|^{-3}$. 因此

$$\ddot{\mathbf{r}} = -\frac{4}{3}\pi G\rho_0 \frac{\mathbf{r}(t)}{|\mathbf{r}(t)|^3} = -\frac{c}{|\mathbf{r}(t)|^3} \mathbf{r}(t), \quad (1.2.11)$$

这里定义的 $c = 4\pi G\rho_0/3$ 当然是个常数. 在后续内容中读者将知道, 方程 (1.2.11) 实际上就是支配宇宙膨胀的 Friedmann 方程, 因为这与在平直近似下从广相中得到的方程相同.

所考虑的 Newton 模型中, 星系 C 位于宇宙的中心且定义了一个惯性系, 而星系 A 不在中心且不是惯性的. 假设星云如此之大, 以至于从 C 或 A 无法观察到它的边界, 那么如果你在这两个星系中的一个, 你怎么能知道你在哪个星系中? 即, 你怎么能知道你是在惯性系 C 中还是在加速系 A ? 答案很明显: 你不能. 由于整个星云均匀地膨胀或收缩, 所以从所有星系来看, 本地天空的画面都是以相同方式均匀地膨胀或收缩的. 而且, 你也不能通过局域实验, 检测出你是在惯性星系 C 还是在加速星系 A ! 事实上, 要检测出你是否在一个加速系中, 你必须观察惯性力. A 系

的加速度是 (1.2.11) 式, 则引入一个作用在所有移动质量上的惯性力

$$\mathbf{F}_{\text{惯性}} = \frac{c}{|\mathbf{r}(t)|^3} \mathbf{r}(t),$$

这就是允许我们检测出非惯性系的那种力. 此外, 所有质量应当受到一个指向 C 的引力

$$\mathbf{F}_{\text{中心}} = -\frac{c}{|\mathbf{r}(t)|^3} \mathbf{r}(t).$$

最后, 它们还受到其它局域的力 $\mathbf{F}_{\text{局域}}$, 进而它们在加速系 A 中的运动由

$$m\ddot{\mathbf{r}} = \mathbf{F}_{\text{局域}} + \mathbf{F}_{\text{惯性}} + \mathbf{F}_{\text{中心}} = \mathbf{F}_{\text{局域}}$$

决定. 这样, A 中局域的动力学看起来就像是在惯性系中一样. 从外界看来, 在 A 中坠落的石头将沿着抛物线运动, 但在加速系 A 看来就是一条直线. 因此, 你不可能判断是否处在宇宙的中心, 也不可能判断所在参考系是惯性的还是非惯性的.

我们该怎样理解这一“不可能”? 根据 Newton 理论, C, A 的动力学应当是截然不同的, 但这种不同却不是可物理观测的 (*physically observable*)! 在 Newton 的概念结构中, A 是非惯性系, 因为同时出现了引力和惯性力, 但二者背后藏有某种隐蔽的联系. 实际上, 情况是完全普适的: 在一个足够小的邻域内, 惯性力和引力在一个自由落体参考系中, 是精确抵消的. 我们再一次说明了等效原理的深邃性¹⁹.

很明显, 可不借助于所有这些不可物理观测的力: 理解该物理情形有个更好的办法, 那就是扔掉 Newton 所偏好的整体 (*global*) 参考系, 并认识到每个星系都有其自己的局部 (*local*) 惯性系. 局部惯性系可以定义为其中没有可观测惯性效应的参考系, 正如 Newton 所做的那样. 这样, 每个星系就都有其局部惯性系. 这些参考系由引力所确定, 也就是说, 正是引力在每个点决定了什么东西该是惯性的. 惯性运动应当像这样相对于局部引力场而言, 而非绝对空间.

进而地, 引力决定了不同星系参考系相对彼此坠落的方式. 引力场可以阐明了不同惯性系间的关系, 因为正是引力场决定了惯性运动. Newton 的“绝对运动”不是相对于绝对空间的运动, 而是相对于引力场所决定的参考系的运动. 或者说, 运动应当相对于引力场. 那么思想很明显了, 第二定律应该支配的正是物体相对于引力场的运动.

关于为何不能找到整体惯性系, 读者不妨作如下设想: 在地球某一局部找到 Einstein 电梯使得引力变成惯性力, 固然是很轻松的. 况且颠倒地来说, 待在地表上的我们都相当于一个“向上加速的参考系”, 而在这个参考系中我们观察到了“惯性力”.

¹⁹顺便提一句, Newton 这个天才是知道这件事的: “无论诸物体彼此间以何种方式运动, 若它们被沿着平行线的相等加速力所推动, 则它们都将继续彼此间的运动, 遵循的方式就如同没有那些力作用一样.” ——《原理》, 推论 VI “运动定律” [?]. Newton 用这个推论来计算太阳系中月球的复杂运动. 在地球系下, 惯性力和太阳引力在良好近似下相互抵消, 那么月球就遵循一个 Kepler 轨道.

然而，如果我们扩大视野，观察两个相距较远的电梯，将会立即发现问题：很显然，在球面的另一些地方，甚至是另一侧，这一部电梯的加速参考系当然就解释不了另一部电梯的情况了。Einstein 如是说，“只能在某一点及其附近能局部地找到这样等效的参考系”。想要找到普适的“参考系”，或许说明，这个世界的几何学要比欧氏几何复杂的多。在 ?? 节将会知道，我们可能需要弯曲的 Lorentz 几何。一旦引入了弯曲几何，那么就总可以找到参考系使得引力变得“不再受力”，如同 (1.2.10) 式那样。

综上所述我们发现，其实引力虽然在局部上与惯性力无法区分，但是只要稍微扩大一点范围，由于不同地方所受的引力方向不同，因而在普通的几何里找不到一个共同的等效参考系，进而就区分出了引力和惯性力。

此外，即使方向是相同的，也有可能区分。如果我们再扩大这部电梯，使得不同高度的点由于其与地心距离不同，而具有明显不同的加速度。即势场不再取 $\phi(h,t) = a_g h$ ，而是 $\phi(r,t) = -GM/r$ 。这部电梯的不同部分所受引力是不均匀的，因而也就会产生形变，具体来讲是一种竖直方向的拉伸。由于这有些类似于月球引力对地表海洋影响所产生的潮汐，那么这种形变的力也叫作潮汐力 (tidal force)。很明显，这个潮汐力是不能用惯性力解释的，因此这个方法同样分辨出了引力与惯性力。

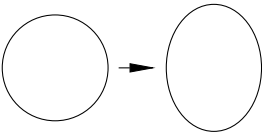


图 1.5: 潮汐效应

最后，在很小的尺度下，引力的强度相较于静电力弱得多：两个质子之间的静电斥力比它们之间的引力吸引大 1.2×10^{36} 倍（任何间隔上两种力都遵从平方反比律）。这在一定程度上暗示了经典引力理论在量子领域可能的失败。现在我们知道，自 Einstein 开始，至今仍未完全地调和广相同量子理论的各种矛盾。由于深入讲解这个话题必将需要读者了解相关的陌生概念，故我将不在此赘述，而只简单地指出：引力同经典电磁作用的统一理论实际上已经完成，其中 Kaluza-Klein 理论和 Einstein 的（经典）统一场论²⁰是完全经典的，而其它的诸如弦论、非交换几何（CLCB 模型）只部分地回答了量子引力的疑难。现在所谓的“大一统理论”通常指的是沿着标准模型的思路，将强相互作用同电弱作用统一的尝试。狭义上的量子引力理论只涉足引力的量子化问题，但这在数学上有巨大的困难；广义的量子引力理论或许是人们所希望的“万物理论”。此外，受黑洞热力学的启发，现在也有企图从统计物理角度寻找引力同热学共通之处的理论，诸如熵力理论。这些话题将部分地放入本书 Volume II 中去。

²⁰受到 Kaluza 理论的启发的 Teleparallel 引力，其数学语言采用 Weitzenbock 几何。