

高二物理竞赛 • 力学 (下), 热学, 几何光学

学而思

目录

| | |
|------------------------------------|-----------|
| 第一章 简谐振动 | 9 |
| 1.1 方程与谐振 | 9 |
| 1.1.1 简谐振动的定义 | 9 |
| 1.1.2 简谐振动的性质 | 9 |
| 1.1.3 简谐振动的判定 | 10 |
| 1.1.4 小振动 | 11 |
| 1.2 阻尼振动与受迫振动 | 12 |
| 1.2.1 阻尼振动 | 12 |
| 1.2.2 受迫振动 | 14 |
| 1.3 多自由度小振动 * | 18 |
| 1.3.1 基于线性代数与分析力学的简正坐标求解 | 20 |
| 1.3.2 基于对称性的简正模判定与简正频率求解 | 24 |
| 1.4 非线性摄动 * | 30 |
| 1.5 格波 | 30 |
| 第二章 万有引力 | 35 |
| 2.1 有心力下质点运动 | 35 |
| 2.1.1 运动的一般特征 | 35 |
| 2.1.2 有心力问题的求解 | 36 |
| 2.2 万有引力下天体运动 | 39 |
| 2.3 二体与潮汐 | 40 |
| 第三章 刚体 | 41 |
| 3.1 刚体的物理描述 | 41 |
| 3.2 平面平行运动 | 43 |
| 3.3 空间刚体运动 * | 44 |
| 第四章 弹性体 | 47 |
| 4.1 弹性体的物理描述 | 47 |
| 4.2 弹性棒, 弹性绳, 弹性膜与弹性体 | 51 |
| 4.2.1 弹性棒 | 51 |
| 4.2.2 弹性绳 | 52 |
| 4.2.3 弹性膜 | 53 |
| 4.2.4 弹性体 * | 54 |
| 4.3 弹性波 | 55 |
| 4.3.1 分离变量法 | 55 |

| | |
|---------------------------------|-----------|
| 4.3.2 变量代换法 | 56 |
| 4.3.3 多维情况 | 56 |
| 4.3.4 再论格波 | 58 |
| 第五章 流体 | 59 |
| 5.1 流体的物理描述 | 59 |
| 5.2 定常流动动力学 | 62 |
| 5.3 黏滞流体动力学 | 63 |
| 5.4 * 流体中的波 | 64 |
| 5.5 波的色散 | 64 |
| 第六章 热力学第一定律 | 65 |
| 6.1 热力学第零定律 | 65 |
| 6.2 热力学第一定律 | 66 |
| 6.2.1 准静态与非准静态过程 | 66 |
| 6.2.2 内能 | 67 |
| 6.2.3 功 | 68 |
| 6.2.4 热量 | 68 |
| 6.2.5 耗散 | 69 |
| 6.3 理想气体 | 70 |
| 6.3.1 理想气体的定义 | 70 |
| 6.3.2 理想气体物态方程 | 71 |
| 6.3.3 混合理想气体 | 73 |
| 6.3.4 理想气体的过程 | 74 |
| 6.4 开放系统的理想气体 | 78 |
| 6.4.1 静态平衡问题——重力场中的大气 | 78 |
| 6.4.2 能量守恒——伯努利方程 | 79 |
| 6.4.3 动量守恒——欧拉方程 | 80 |
| 第七章 热力学第二定律 | 83 |
| 7.1 循环过程 | 83 |
| 7.1.1 热机与热泵 | 83 |
| 7.1.2 热机循环 | 86 |
| 7.2 理想气体的熵 | 88 |
| 7.3 热力学第二定律 | 90 |
| 7.4 熵的计算 | 93 |
| 7.4.1 理想气体的熵 | 93 |
| 7.4.2 固定热容固体熵 | 94 |
| 7.4.3 大热容恒温热库 | 94 |
| 7.4.4 传热熵 | 94 |
| 7.4.5 混合熵 | 94 |
| 7.5 热力学函数与其特性 | 96 |
| 7.5.1 四个热力学函数与四个状态参量 | 96 |
| 7.5.2 若干定理的证明 | 97 |
| 7.5.3 自由能的含义 | 102 |

| | |
|-------------------------------|------------|
| 7.5.4 化学势 | 103 |
| 7.6 近平衡态热力学 * | 104 |
| 7.6.1 线性输运现象 | 104 |
| 7.6.2 传热的熵产生 | 107 |
| 7.6.3 普遍理论与昂萨格倒易关系 | 107 |
| 7.6.4 推证热电耦合的普遍规律 | 108 |
| 第八章 液体与固体的性质 | 111 |
| 8.1 固体晶格论 | 111 |
| 8.1.1 经典晶格论 | 111 |
| 8.1.2 量子晶格论 * | 114 |
| 8.2 固体电子论 * | 119 |
| 8.2.1 德鲁特模型 | 119 |
| 8.3 液体的彻体性质 | 120 |
| 8.3.1 液体性质综述与其微观成因 | 120 |
| 8.3.2 热容与黏度 | 122 |
| 8.4 液体的表面性质 | 122 |
| 8.5 极端条件下的其他物态 | 122 |
| 第九章 相与相变摘要 | 125 |
| 9.1 相平衡 | 125 |
| 9.2 气液相变 | 125 |
| 第十章 统计物理基础 | 127 |
| 10.1 数学基础 | 127 |
| 10.1.1 概率与独立性 | 127 |
| 10.1.2 随机变量及其数字特征 | 128 |
| 10.1.3 信息熵 | 129 |
| 10.2 统计假设 | 133 |
| 10.3 麦克斯韦分布律 | 134 |
| 10.4 能均分定理 | 135 |
| 10.5 功, 热, 熵 | 135 |
| 第十一章 光波与光线 | 137 |
| 11.1 界面上的反射与折射 | 137 |
| 11.1.1 光波与光线 | 137 |
| 11.1.2 菲涅尔公式 | 139 |
| 11.2 光线方程 | 143 |
| 11.2.1 光线方程与折射定律 | 143 |
| 11.2.2 光力类比 | 144 |
| 11.3 费马原理 | 145 |
| 第十二章 光学成像 | 147 |
| 12.1 傍轴光成像 | 147 |
| 12.1.1 物与像 | 147 |
| 12.1.2 球对称成像系统与符号法则 | 148 |

| | |
|----------------------------|------------|
| 12.1.3 光具组成像 | 150 |
| 12.2 理想成像系统 | 151 |
| 12.2.1 作图法 | 152 |
| 12.2.2 基点基面性质 | 153 |
| 12.2.3 实例与望远系统 | 154 |
| 12.3 更多讨论 * | 156 |
| 12.3.1 理想成像本质 | 156 |
| 第十三章 光学仪器 | 159 |
| 13.1 光度学 | 159 |
| 13.1.1 色度学 * | 159 |
| 13.1.2 国际单位与光度函数 | 163 |
| 13.1.3 亮度与照度 | 165 |
| 13.1.4 光度学守恒量 | 165 |
| 13.2 光阑与光瞳 | 166 |
| 13.2.1 光阑 | 166 |
| 13.2.2 光瞳 | 166 |
| 13.3 常见仪器简介 | 166 |
| 13.3.1 人眼 | 166 |
| 13.3.2 显微镜 | 166 |
| 13.3.3 望远镜 | 167 |

第一章 简谐振动

1.1 方程与谐振

1.1.1 简谐振动的定义

振动是最常见的物理现象. 而振动中的最简单 (simple) 而和谐 (harmonic) 者谓之简谐振动 (simple harmonic oscillation). 对谐振子 (harmonic oscillator) 的学习与研究是会贯彻整个物理理论不同层次内容的始终的. 现在是经典力学, 以后会上升到场论, 量子力学与量子场论的高度.

简谐振动是指一个物理量 Q 随时间围绕其平衡位置做上下的波动. 其形式符合:

$$Q = Q_0 + \Delta Q \cos(\omega t + \varphi)$$

我们经常会有用复数表示振动的习惯, 其做法是在三角函数 \cos 与其宗量 (argument) $\phi = \omega t + \varphi$ 构成的项后添加一个虚的 $i \sin \phi$ 项, 于是新的写法变成:

$$\tilde{Q}(t) = Q_0 + \Delta Q e^{i(\omega t + \varphi)} ; Q = \Re(\tilde{Q})$$

又或者:

$$\tilde{Q}(t) = Q_0 + \Delta \tilde{Q} e^{i\omega t} ; \Delta \tilde{Q} = \Delta Q e^{i\varphi}$$

以上各个常量中, Q_0 是平衡位置, ΔQ 叫振幅 (amplitude), 宗量 $\omega t + \varphi$ 叫做相位 (phase), ω 叫角频率 (angular frequency), φ 叫初相位. \tilde{Q} 为复化的复物理量, 而 $\Delta \tilde{Q}$ 叫做复振幅 (complex amplitude).

复数表示最大的一个好处就在于很方便计算物理量的线性组合与导数. 事实上, 如果合理选取 $Q_0 = 0$:

$$\dot{Q} = -\omega \Delta Q \sin \phi ; \dot{\tilde{Q}} = i\omega \tilde{Q} \Rightarrow \dot{Q} = \Re(\dot{\tilde{Q}})$$

$$\ddot{Q} = -\omega^2 \Delta Q \cos \phi ; \ddot{\tilde{Q}} = -\omega^2 \tilde{Q} \Rightarrow \ddot{Q} = \Re(\ddot{\tilde{Q}})$$

1.1.2 简谐振动的性质

以一个质点水平坐标 x 围绕 $x = 0$ 左右做谐振为例. 其运动方程写作¹:

$$x = A e^{i(\omega t + \varphi)}$$

那么其速度与加速度为:

$$v = i\omega A e^{i(\omega t + \varphi)}, a = -\omega^2 A e^{i(\omega t + \varphi)}$$

我们发现在运动过程中的任意一个时刻, 加速度都是指向平衡位置的, 与偏离平衡位置的位移是成正比的:

$$a = -\omega^2 x$$

¹ 以后我们对物理量和物理量的复化只在必要的时候加以区分, 看到复数只需要认为省写了取实部这一例常操作罢了.

而任意一个时刻既然 x 是余弦函数, v 是正弦函数, 它们就满足其绝对值大小的“此消彼长”关系²:

$$v^2 + \omega^2 x^2 = \omega^2 A^2$$

相位是很重要的物理概念, 相(phase), 状态也, 相位则是一个可以用来表示状态的数. 反过来, 我们可以根据物体在一个时刻的状态反过来确定这个时刻的相位. 如果位移是 x 而振幅为 A , 则:

$$\phi = \arcsin \frac{x}{A}, \arcsin \frac{x}{A} \in \left\{ \arcsin \frac{x}{A} + 2n\pi \mid n \in \mathbb{Z} \right\} \cup \left\{ \pi - \arcsin \frac{x}{A} + 2n\pi \mid n \in \mathbb{Z} \right\}$$

同理如果已知速度 v 和速度振幅(velocity amplitude) ωA , 那么相位为:

$$\phi = \arccos \frac{v}{\omega A}, \arccos \frac{v}{\omega A} \in \left\{ \arccos \frac{v}{\omega A} + 2n\pi \mid n \in \mathbb{Z} \right\} \cup \left\{ -\arccos \frac{v}{\omega A} + 2n\pi \mid n \in \mathbb{Z} \right\}$$

相位的取法上具有多值性. 但是相位随时间的变化我们约定必须是连续的, 事实上它随时间线性增加. 也就是说如果前一个状态下相位为 ϕ_1 , 后一个状态下相位为 ϕ_2 , 那么这两个状态间历时:

$$\Delta t = \frac{\phi_2 - \phi_1}{\omega}$$

1.1.3 简谐振动的判定

事实上以上的两个关系都可以成为体系做简谐振动的判据, 它们为:

- 线性回复判据: 如果一个随时间演化的变量的二阶导数正比于变量本身, 且符号相反, 可以判断变量做简谐振动:

$$\ddot{q} = -\omega^2 q \Rightarrow q = A \cos(\omega t + \varphi)$$

- 此消彼长判据: 如果一个随时间演化的变量的一阶导数与自己的正系数平方和为动力学守恒量(一般就是能量). 可以判断变量做简谐振动:

$$\dot{q}^2 + \omega^2 q^2 = \omega^2 A^2 \Rightarrow q = A \cos(\omega t + \varphi)$$

证明如下:

线性回复 \Rightarrow 此消彼长:

首先进行代换:

$$\ddot{q} = \frac{d}{dt}(\dot{q}) = \frac{dq}{dt} \frac{d}{dq}(\dot{q}) = \frac{\dot{q}d\dot{q}}{dq}$$

将上式代入 $\ddot{q} = -\omega^2 q$:

$$\dot{q}d\dot{q} + \omega^2 \cdot q dq = d\left(\frac{1}{2}\dot{q}^2 + \frac{1}{2}\omega^2 q^2\right) = 0 \Rightarrow \dot{q}^2 + \omega^2 q^2 = C$$

令 $A = \sqrt{C/\omega^2}$ 即得到此消彼长判据.

此消彼长 \Rightarrow 简谐振动:

考虑 \dot{q} 的正根即可, 负根结果也是简谐振动:

$$\begin{aligned} \dot{q} &= \frac{dq}{dt} = \omega \sqrt{A^2 - q^2} \\ \Rightarrow \quad \frac{dq}{\sqrt{A^2 - q^2}} &= \omega dt \end{aligned}$$

两边同时积分, 积分常数写到右侧记作 φ , 得到:

$$-\arccos \frac{q}{A} = \omega t + \varphi \Rightarrow q = A \cos(\omega t + \varphi)$$

²注意到这里(往下)要避免使用复数, 因为一个复数的实部平方不会等于其平方的实部 $v^2 \neq \Re(\tilde{v}^2)$

1.1.4 小振动

值得注意的是,很多情况下体系的运动并不是严格的简谐振动.十分常见的一种情况是**小振动**(small oscillation).事实上,如果我们研究体系为完整而稳定的一自由度体系,广义坐标为 q ,而且势能函数 $V(q)$ 存在一个极小值:

$$V'(q_0) = 0, \quad V''(q_0) > 0$$

那么根据上一章的阐述,这个 q_0 就是体系的稳定平衡位置.那么,任何偏离这个平衡位置的系统运动只要偏离的值足够小,即做坐标变换 $\delta = q - q_0$ 为无穷小量,那么就一定为简谐振动.这是因为势能可以由泰勒展开为:

$$V = \frac{1}{2}V''(q_0)\delta^2 + \frac{1}{3!}V'''(q_0)\delta^3 + \dots$$

只要 δ 足够小,三阶项就必然远小于二阶项.从而可以只保留第一项.同理这也适用于动能,它必然正比于 δ 的平方,系数则与平衡位置有关:

$$T = \frac{1}{2}M(q_0)\dot{\delta}^2$$

从而这个体系的能量函数(哈密顿量)就被近似为了:

$$H = T + V = \frac{1}{2}M(q_0)\dot{\delta}^2 + \frac{1}{2}V''(q_0)\delta^2$$

这就直接符合了“此消彼长”判据.从而小振动的角频率为:

$$\omega = \sqrt{\frac{V''(q_0)}{M(q_0)}}$$

例如,在典型的单摆问题中,取摆线与竖直方向的夹角 θ 为广义坐标.摆球的重力势能以平衡位置为原点表示为:

$$V(\theta) = mgl(1 - \cos\theta)$$

而动能为:

$$T = \frac{1}{2}ml^2\dot{\theta}^2 \Rightarrow M(\theta) = ml^2$$

从而只需要带入以上公式,就得到:

$$\omega = \sqrt{\frac{V''(0)}{M(0)}} = \sqrt{\frac{mgl}{ml^2}} = \sqrt{\frac{g}{l}}$$

这是因为事先看出来了稳定平衡点为 $\theta = 0$.简单的小振动问题的思路都大抵如此.第一步是写出能量函数来,它决定体系的所有可能的动力学演化.第二步要找到稳定平衡位置,往往一样地是通过势能函数的增减凹凸性来判断,往往也要结合体系的对称性来观察.最后便是在平衡位置处把动能势能都近似为简单的平方项.唯一需要补充的是,除了利用二阶导数来计算这一个平方项,对常见函数写泰勒展开式也是十分常见的思路.

单摆问题的周期从而就是:

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{l}{g}}$$

但是一定要注意这个振动仅仅对小振动成立.例如在摆角为 $10^\circ = 0.087\text{rad}$ 时会有约 2% 的误差.如何计算这个误差?之后的非线性振动将讨论这个问题.

1.2 阻尼振动与受迫振动

1.2.1 阻尼振动

阻尼振动, 狹义地是指普通的弹簧型谐振子, 但是让振子运动的空间充满介质而产生湿摩擦 (流体摩擦). 从而产生一个正比于速度的阻尼的情形. 其牛顿定律为:

$$m\ddot{x} = -kx - \gamma\dot{x}$$

这里的 γ 即**阻尼系数**(damping coefficient). 习惯上把 m/γ 称作**弛豫时间**(relaxation time). 最后还定义**衰减参数**(attenuation parameter) 或**损耗参数**(loss parameter) 为:

$$\beta = \frac{\gamma}{2m}$$

而无阻尼情况下的**固有频率**(natural frequency), 这里是角频率, 为:

$$\omega_0 = \sqrt{\frac{k}{m}}$$

这样原来的方程又被重新写为:

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = 0$$

我们定义广义的**阻尼振动**(damped oscillation), 只需要一个运动的任意自由度 q , 符合或者可以近似为以上式子形式的动力学方程, 那么这个自由度上的运动就是一个阻尼振动. 所有的阻尼振动只有两个常数特征, 一个是固有角频率, 一个是损耗的大小.

阻尼振动的求解需要分为三种情况. 我们统一猜解:

$$x = A e^{i\omega t}$$

带入原方程以后发现 ω 需要满足:

$$\omega^2 - 2i\beta\omega - \omega_0^2 = 0$$

从而:

- **欠阻尼**(underdamped)

常见的小阻尼情况, 此时 $0 < \beta < \omega_0$. 通过对以上方程求解得到两个含有正虚部的关于虚轴对称的根:

$$\omega_{1,2} = i\beta \pm \sqrt{\omega_0^2 - \beta^2}$$

这样就得到了原阻尼振动方程的通解:

$$x = e^{-\beta t} [A \cos(\sqrt{\omega_0^2 - \beta^2} t) + B \sin(\sqrt{\omega_0^2 - \beta^2} t)] = C e^{-\beta t} \cos(\sqrt{\omega_0^2 - \beta^2} t + \varphi)$$

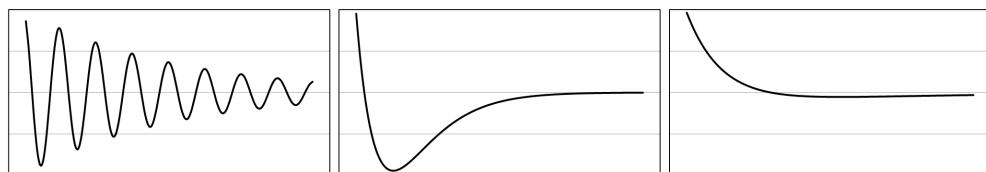


图 1.1: 欠阻尼, 临界阻尼与过阻尼

- **过阻尼**(overdamped)

此时 $\beta > \omega_0$. 从而振动和回复力被极大地抑制了. 主要是一个在大阻力和初速度下缓慢回到平衡位置的运动. 以上方程两根此时为两个正虚根:

$$\omega_{1,2} = i(\beta \pm \sqrt{\beta^2 - \omega_0^2})$$

这样就得到了原阻尼振动方程的通解:

$$x = e^{-\beta t} [Ae^{\sqrt{\beta^2 - \omega_0^2}t} + Be^{-\sqrt{\beta^2 - \omega_0^2}t}] = Ce^{-\beta t} \cosh(\sqrt{\beta^2 - \omega_0^2}t + \varphi)^3$$

- **临界阻尼**(critical damped)

这对应 $\beta = \omega_0$ 的特殊情况. 可以通过微分方程的理论, 或者采用求 $\beta \rightarrow \omega_0$ 极限的方式得到此时的运动方程. 它为:

$$x = e^{-\beta t}(A + Bt)$$

临界阻尼在实际生活中的应用为: 它是较快能够让振子回到平衡位置的理想阻尼大小. 这一点不难理解, 不难看出, 欠阻尼情况的振幅随时间的衰减方式为 $e^{-\beta t}$, 也就是适当增大 β 有利于振幅尽快衰减. 但是, 过阻尼时, 长时间后位移与时间的关系为

$$x \sim e^{-\beta t} \cosh(\sqrt{\beta^2 - \omega_0^2}t) \sim e^{-\beta t} e^{\sqrt{\beta^2 - \omega_0^2}t} \sim e^{-\frac{\omega_0^2 t}{\beta + \sqrt{\beta^2 - \omega_0^2}}}$$

这就不再难发现, 此时增加 β 只会使得 t 前的衰减系数变小而不利于尽快衰减其位移. 从而, 无论是欠阻尼还是过阻尼, 临界阻尼是它们能够尽快回到平衡位置的极限. 此时原来的固有频率对应的周期就是它回到平衡位置的特征时间.

对于欠阻尼情况, 尤其是衰减因子 $\beta \ll \omega_0$ 的小阻尼情况, 还有一些重要的性质值得挖掘. 下面都默认是这样的前提:

从之前的运动方程解不难看出来, 我们可以把小阻尼情况下的运动方程看作是一个振幅不断随着时间衰减的简谐振动. 其振幅衰减函数和阻尼化角频率为:

$$A(t) = A_0 e^{-\beta t}$$

$$\omega_d = \sqrt{\omega_0^2 - \beta^2} \approx \omega_0$$

要讨论其衰减的原因, 还有一种新的角度, 便是考虑其能量. 当振幅为 A 时, 其动能, 势能的和为:

$$E = T + V = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2$$

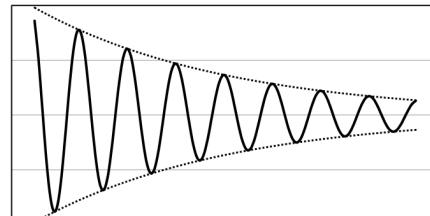


图 1.2: 小阻尼下的振幅衰减

现在它不是守恒的, 仅仅是在少数几个周期里面“近似”守恒. 但是我们可以考虑求它在一个周期内的平均:

$$\bar{E} = \frac{1}{2} \cdot \frac{1}{2}m\omega_d^2 A^2 + \frac{1}{2} \cdot \frac{1}{2}kA^2 = \frac{1}{2}m \frac{\omega_0^2 + \omega_d^2}{2} A^2 \approx \frac{1}{2}m\omega_0^2 A^2$$

现在我们计算单位时间减少的能量的平均值. 它其实就是阻尼造成的负功功率的平均值:

$$\bar{P} = \bar{\gamma}v \cdot \bar{v} = \frac{1}{2}\gamma\omega_d^2 A^2 \approx \frac{1}{2}\gamma\omega_0^2 A^2 = \frac{1}{2}m\omega_0^2 A^2 \cdot 2\beta$$

³也有可能是 \sinh

通过损耗功率等于能量的负导数,也可以得到振幅随着时间衰减的具体结果.我们现在关心一个重要的量,它实际上是反应振子振动单位时间所减少的能量占由于以 ω_0 振动导致的总能量转化速率的比例,我们把上式换一种写法:

$$\bar{P} = \frac{\omega_0 \bar{E}}{Q}$$

以上引入的 Q 就是著名的**品质因数**(quality factor). 它越大,表示损耗越低,振子的阻尼就越小,就越接近“谐振”. 实际上它算法为:

$$Q = \frac{\omega_0}{2\beta}$$

品质因数经常与另一个因子一起提出,两者是简单的倒数关系,更精确地,我们定义:

$$\sin \delta = \frac{1}{2Q} = \frac{\beta}{\omega_0}$$

δ 被称作**损耗角**(loss angle). 一定要注意这两个概念是普遍的. 谐振子模型是很多实际问题的抽象,它们作为谐振子都有自己的品质因数或损耗角,用以描述过程的一些共性,而且,不光可以用能量的损失的共性来定义品质因数和损耗角,还有很多重要的方式. 详情见后.

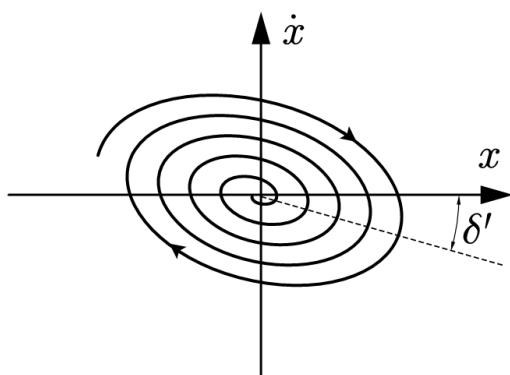


图 1.3: 阻尼振动的相图

现在我们来考虑简谐振动的**相图**(phase diagram). 在动力学中,这总是指把单个过程的运动,画成以其广义坐标和广义速度(或动量)为坐标轴的**相空间**(phase space)中的曲线形成的图像. 阻尼振动的代表性解为:

$$x = A e^{-\beta t} e^{i\omega_d t}$$

在这个解下,其速度为:

$$\dot{x} = (i\omega_d - \beta) A e^{-\beta t} e^{i\omega_d t}$$

两个量都在随时间衰减. 从而在相图上就体现为一随着时间不断向着原点靠近的螺旋线. 如果没有阻尼,那么标准简谐振动的相轨迹就是正椭圆:

$$\frac{x^2}{A^2} + \frac{\dot{x}^2}{\omega_0^2 A^2} = 1$$

但是,有阻尼时除了是螺旋式的椭圆,它还不是“正”的. 事实上,如果去除衰减因子,即将 x 和 \dot{x} 乘以一个 $e^{\beta t}$ 以抵消其衰减:

$$X = x e^{\beta t} = A e^{i\omega_d t}, \quad V = \dot{x} e^{\beta t} = i(\omega_d + i\beta) A e^{i\omega_d t}$$

两者之间也不是恰好相差相位 $\pi/2$,而是有一个额外相位差:

$$\delta = \arctan \frac{\beta}{\omega_d} = \arcsin \frac{\beta}{\omega_0}$$

可以发现这个相位差就是损耗角. 也就是说,损耗角的另一种理解方式是位置与速度振动过程中的相位夹角. 作为结果,其相图中的椭圆也要倾斜一个角度 δ' . 可以证明,损耗角越大,这个倾斜角度也会越大.

1.2.2 受迫振动

如果在谐振子上施加一个周期性的策动外力. 我们总是研究基本的简谐式的外力⁴:

$$F(t) = F_0 e^{i\omega t} \quad \text{i.e.} \quad F(t) = F_0 \cos \omega t$$

⁴傅里叶分析理论告诉我们,任何函数都可以分解为无穷多助三角函数的线性组合. 故对于线性系统原则上就可以用叠加原理处理任意任意强迫力的运动求解:只要我们把简谐强迫力找到解法.

这个运动就称作受迫振动(driven oscillation). 这样子原来的方程就变为:

$$m\ddot{x} + \gamma\dot{x} + kx = F_0 e^{i\omega t}$$

$$\Rightarrow \ddot{x} + 2\beta\dot{x} + \omega_0^2 x = \frac{F_0}{m} e^{i\omega t}$$

很明显, 可以猜一个谐振解. 它是上述微分方程的特解(稳态解). 通解还需要加上上一部分讨论的阻尼振动解(齐次方程的解是暂态解). 注意其振动频率应当是策动力的频率 ω 而不是体系的固有频率 ω_0 :

$$x = A e^{i\omega t} \Rightarrow (\omega_0^2 - \omega^2 + 2i\beta\omega)A = \frac{F_0}{m}$$

故, 复振幅的解法也是相当地直接:

$$A = \frac{F_0/m}{\omega_0^2 - \omega^2 + 2i\beta\omega}$$

作为复振幅, 上 A 表达式其实同时包含了两个因素, 一个是振幅, 一个是相位. 现在我们分离它们:

$$A = |A| e^{i\varphi}$$

$$|A| = \frac{F_0/m}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\beta^2\omega^2}}$$

$$\varphi = -\arctan \frac{2\beta\omega}{\omega_0^2 - \omega^2} \quad \text{or} \quad \varphi = -\left(\pi - \arctan \frac{2\beta\omega}{\omega^2 - \omega_0^2}\right)$$

这就构成了两个关于 ω 的函数. 这也是实验上我们特别关心的两个函数, 它们反映在稳态受迫振动的时候, 一定外力下振子产生的振幅与相位究竟是怎样取决于策动力的角频率的. 这就是著名的幅频特性(amplitude-frequency characteristic) 和相频特性(phase-frequency characteristic). 利用之前定义的品质因数 Q , 再命 $x = \omega/\omega_0$. 整理得:

$$|A| = \frac{QF_0/m\omega_0^2}{\sqrt{Q^2(1-x^2)^2 + x^2}}$$

$$\varphi = -\arccos \frac{Q(1-x^2)}{\sqrt{Q^2(1-x^2)^2 + x^2}}$$

我们依然只考虑小阻尼情况. 此时 Q 较大. 比如图1.4所示, $Q = 5, 10, 20$ 等.

那么受迫振动又被分为三种情形:

- 低频区为弹性控制区.

在 $\omega \ll \omega_0$ 时, $m\ddot{x}, \gamma\dot{x}, kx$ 三项中最后一项最大, 主要是弹簧的弹性力与策动力去平衡. 从而体现出以下计算结果:

$$|A| \approx \frac{F_0}{m\omega_0^2} = \frac{F_0}{k}$$

$$\varphi \approx 0$$

也就是策动力与位移几乎同相, 振子有充分的时间去实现平衡.

- 高频区为惯性控制区.

在 $\omega \gg \omega_0$ 时, $m\ddot{x}, \gamma\dot{x}, kx$ 三项中第一项最大, 主要是高速变化的策动力去直接产生振子加速度. 从而体现出以下计算结果:

$$|A| \approx \frac{F_0}{mx^2\omega_0^2} = \frac{F_0}{mw^2}$$

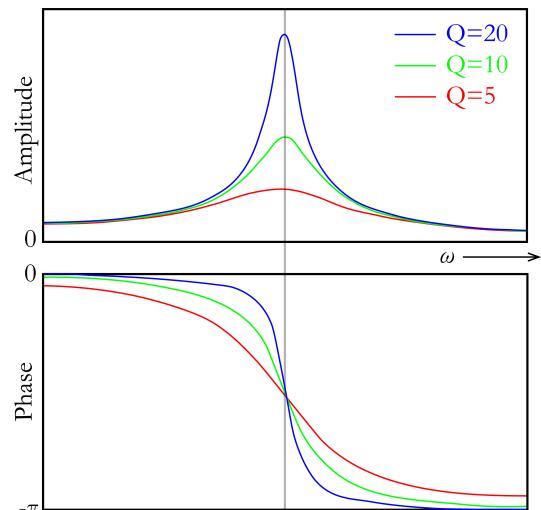


图 1.4: 幅频, 相频特性

$$\varphi \approx -\pi$$

也就是策动力与位移几乎反相(亦反向).振子“随波逐流”,弹性力和阻尼力没有时间跟上策动力的节奏.

- 中频区为共振区.

在 ω 在 ω_0 附近,即量级接近时.三个项 $m\ddot{x}$, $\gamma\dot{x}$, kx 也量级相当.如果把结果画成图1.4.可以发现十分典型的现象是,振幅在某频率处取到了极值.而相位差(绝对值)在某频率处从小于 $\pi/2$ 增加到大于 $\pi/2$.

共振(resonance) 是谐振子系统中十分典型的现象.一个共振发生的时候有两个量是值得关注的:

第一个值得关注的共振参数,是共振的峰的位置.即**共振频率**(resonance frequency).但是这个因素又被分为位移共振和速度共振,即位移振幅 $|A|$ 最大或者速度振幅最大的角频率值.或者可以证明,等价的是,**振幅共振**(amplitude resonance) 和**相位共振**(phase resonance).前者指的是响应 x 最大的点:

$$x^2 = 1 - \frac{1}{2Q^2} \approx 1 \Rightarrow \omega \approx \omega_0$$

后者则指响应和策动力完全相位差为 $\pi/2$ 的情况:

$$x^2 = 1 \Rightarrow \omega = \omega_0$$

可见两个共振在小阻尼情况下几乎是重合的.在振幅共振的情况下,振幅最终为:

$$|A| = Q \cdot \frac{F_0}{m\omega_0^2} \cdot \sqrt{\frac{4Q^2}{4Q^2 - 1}} \approx Q \cdot \frac{F_0}{m\omega_0^2}$$

如果以这个角频率的外力 F_0 直接作用在自由的质点上,或者是考虑让一个没有阻尼的振子进行固有振动并使弹簧给振子的力就是 F_0 .我们发现振子振幅应当是 $F_0/m\omega_0^2$.这就是说共振时振幅成为了这个振幅的 Q 倍.说明了如果让体系从静止开始,自某一时刻开始受到周期性的外力而产生振动,如果频率接近固有频率,体系会十分反常的振幅越来越大,直到振幅是自由振幅的 Q 倍.这也是品质因数的第二种理解方式:共振的放大倍数.

在看相位共振时的策动力-相位差为 $\pi/2$ 的现象.这说明了什么?这说明策动力与速度完全就是同相的.带入 $x = 1$ 容易计算此时位移振幅为 $|A| = QF_0/m\omega_0^2$,等效地可以写成 $F_0 = m\omega_0^2|A|/Q$.那么速度振幅就是 $\omega_0|A|$.从而策动力对速度单位时间做的功就是:

$$W = \overline{F(t)\dot{x}(t)} = \frac{1}{2}F_0\omega_0|A| = \frac{1}{2}m\omega_0^3|A|^2/Q$$

这与上一部分得到了相似的结果.因为振动的能量其实就是:

$$\bar{E} = \frac{1}{2}m\omega_0^2|A|^2$$

也就是说品质因数反映了耗散的能量与总能量(动能势能互相转换)的比例:

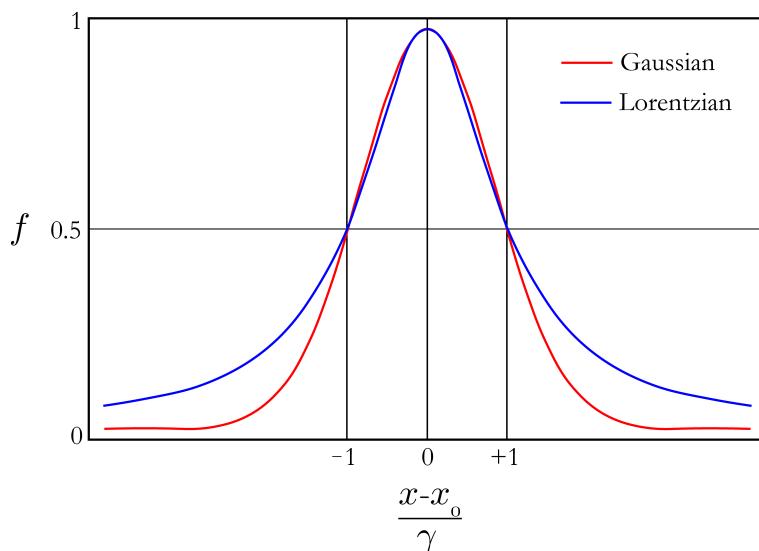
$$Q = \frac{\omega_0 \bar{E}}{W}$$

第二个值得关注的共振参数,是共振的峰的宽度.谐振子模型给出的幅频曲线,与著名的**柯西分布**(Cauchy distribution) 对应的函数曲线有密切联系,柯西分布又常常因为在理论物理学中的广泛使用而称作**洛伦兹型**(lorentzian shape)⁵.所谓柯西分布最初来自概率分布模型,它是指以下分布函数:

$$\text{pdf.}^6 = \frac{1}{\pi[\gamma^2 + (x - x_0)^2]}$$

⁵比如谱线的形状,即线型,就常被视作洛伦兹线型.

⁶probability distribution function,即概率分布函数的缩写.

图 1.5: 高斯型和洛伦兹型⁷

这样一个分布函数是需要有归一化的要求的:

$$\int_{-\infty}^{+\infty} \text{pdf. } dx = 1$$

但是现在的幅频特性, 明显与洛伦兹型是有出入的. 第一是, 我们需要把因变量修改为 $I = |A|^2$ 而不是 $|A|$. 这代表振动的强度而不是振幅. 第二是, 不同于柯西分布的分母为二次函数, 这里的分母却作为了 x 的四次函数. 是 ω 的四次函数. 虽然 $I(\omega)$ 与柯西分布有别, 而且也不归一化. 但是却体现出类似的性质. 它们都是有一个峰, 并在峰两侧按照较慢的多项式的方式衰减 (而不是像高斯型那样指数衰减). 而且我们都用半高峰宽(FWHM, full width at half maximum) 来描述峰的宽度. 它被定义为函数值, 或者代表的强度减少为峰值的一半对应的两点之间的区间宽度. 对于标准的洛伦兹型, 容易发现半高峰宽就是:

$$\text{FWHM} = 2\gamma$$

谐振子的幅频曲线的半高峰宽就更加难算. 我们此时会采取一个近似技巧, 不难验证其正确性. 注意到在洛伦兹型中取到半高的点处分母中的两项是相等的, 于是有理由相信在 $I(x)$ 函数情况下, 直接让分母两项相等就也近似是在半高点处:

$$\begin{aligned} I(x) &= I_0 \Big/ \left[(1 - x^2)^2 + \frac{x^2}{Q^2} \right] \Rightarrow I_{\max} \approx I_0 \\ I &= \frac{I_{\max}}{2} \Rightarrow 1 - x^2 \approx \pm \frac{x}{Q} \end{aligned}$$

这就可以确定, 取到半高的点 x_+, x_- 为:

$$x_+ = \sqrt{1 + \frac{1}{4Q^2}} + \frac{1}{2Q} ; \quad x_- = \sqrt{1 + \frac{1}{4Q^2}} - \frac{1}{2Q} \Rightarrow x_+ - x_- = \frac{1}{Q}$$

而 $\omega_+ = x_+ \omega_0$, $\omega_- = x_- \omega_0$. 从而半高峰宽就是:

$$\text{FWHM} = \Delta\omega = \omega_+ - \omega_- = \frac{\omega_0}{Q}$$

我们又发现了品质因数的一种常见定义方式, 它被定义为谱线的锐度(sharpness), 即峰处角频率与半高峰宽的比值:

$$Q = \frac{\omega_0}{\Delta\omega}$$

⁷高斯型是指高斯分布:

$$\text{pdf.} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

1.3 多自由度小振动 *

多自由度小振动,首先得是一个多自由度平衡问题.在上一章带约束的完整系统的静力学平衡的基础上.如果体系还存在 f 个自由度,与之对应的有 f 个广义坐标 q_i .那么我们指出过,体系的平衡条件为:

$$Q_i = \sum_k \mathbf{F}_k \cdot \frac{\partial \mathbf{r}_k}{\partial q_i} = 0$$

但是在绝大多数情况下,我们处理的是保守体系的问题.所以这种场合下又有势能函数 $V(q_i)$.广义力也能从势能中得到:

$$Q_i = \frac{\partial V}{\partial q_i}$$

这完全就能得到保守体系的平衡条件:

$$\frac{\partial V}{\partial q_i} = 0$$

又是在绝大多数情况下,势能函数是性质非常好的函数⁸.从而可以在平衡位置附近泰勒展开.把平衡位置记作 q_{i0} ,而偏离平衡位置的位移为 δ_i ,平衡位置处势能记作 $V_0 = V(q_{i0})$,由于平衡位置一阶导数为零,故:

$$V(q_{i0} + \delta_i) = V_0 + \frac{1}{2} \sum_{i,j} \frac{\partial^2 V}{\partial q_i \partial q_j} \Big|_{q_{i0}} \delta_i \delta_j + \frac{1}{3!} \sum_{i,j,k} \frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} \Big|_{q_{i0}} \delta_i \delta_j \delta_k + \dots$$

在忽略三阶小量(不会成为领头项,条件下面再讨论)的前提下,再通过去掉常数 V_0 ,不影响势能的效果.我们发现势能实际上就成为了一个二次型:

$$V = \frac{1}{2} \sum_{i,j} \frac{\partial^2 V}{\partial q_i \partial q_j} \Big|_{q_{i0}} \delta_i \delta_j = \frac{1}{2} \sum_{i,j} V_{ij} \delta_i \delta_j$$

而动能,由于在平衡位置附近小范围内做小的运动,当然也是一个关于广义速度的二次型,系数为常数 T_{ij} :

$$T = \frac{1}{2} \sum_{i,j} T_{ij} \dot{\delta}_i \dot{\delta}_j$$

关于系数 T_{ij} , V_{ij} .一个要求是它们必须对称:

$$T_{ij} = T_{ji} ; \quad V_{ij} = V_{ji}$$

其中的道理非常简单,例如只有两个广义坐标的变化 δ_1 , δ_2 .那么 $V_{12} = a \neq V_{21} = b$ 所决定的势能二次型与 $V_{12} = V_{21} = (a+b)/2$ 决定的二次型根本就是同一个:

$$V_{12} = a \neq V_{21} = b : V = \frac{1}{2} V_{11} \delta_1^2 + \frac{1}{2} V_{22} \delta_2^2 + \frac{1}{2} a \delta_1 \delta_2 + \frac{1}{2} b \delta_2 \delta_1 = \frac{1}{2} V_{11} \delta_1^2 + \frac{1}{2} V_{22} \delta_2^2 + \frac{1}{2} (a+b) \delta_1 \delta_2$$

$$V_{12} = V_{21} = \frac{a+b}{2} : V = \frac{1}{2} V_{11} \delta_1^2 + \frac{1}{2} V_{22} \delta_2^2 + \frac{1}{2} \frac{a+b}{2} \delta_1 \delta_2 + \frac{1}{2} \frac{a+b}{2} \delta_2 \delta_1 = \frac{1}{2} V_{11} \delta_1^2 + \frac{1}{2} V_{22} \delta_2^2 + \frac{1}{2} (a+b) \delta_1 \delta_2$$

反过来,如果给一个二次型,我们也总可以把系数写成对称的形式,比如在动能中,毕竟偏导数本身也是可交换次序的:

$$T_{ij} = \frac{\partial^2 T}{\partial \dot{\delta}_i \partial \dot{\delta}_j} = \frac{\partial^2 T}{\partial \dot{\delta}_j \partial \dot{\delta}_i} = T_{ji}$$

⁸数学上叫做光滑函数.

举两个例子. 在耦合摆(coupled pendulum) 问题1.6中, 分开看是相同的两个单摆, 但却因为弹簧而发生弱的耦合, 设想弹簧原长为 x 为摆间距, 这样就使得两单摆按原来的位置平衡时, 弹簧也没有作用力. 但只要一发生运动就彼此影响, 设两摆球距离 $x' = \sqrt{(x + l \sin \theta_2 - l \sin \theta_1)^2 + (l \cos \theta_2 - l \cos \theta_1)^2}$. 这样作为整体体系动能和势能为:

$$T_{\text{cp}} = \frac{1}{2}ml^2\dot{\theta}_1^2 + \frac{1}{2}ml^2\dot{\theta}_2^2$$

$$V_{\text{cp}} = mgl(1 - \cos \theta_1) + mgl(1 - \cos \theta_2) + \frac{1}{2}k(x' - x)^2$$

再引入描述耦合强度的参量 $\varepsilon = kl/mg$, 小量近似以后, 动能, 势能被近似为二次型:

$$T_{\text{cp}} = \frac{1}{2} \cdot ml^2 \cdot \dot{\theta}_1^2 + \frac{1}{2} \cdot ml^2 \cdot \dot{\theta}_2^2$$

$$V_{\text{cp}} = \frac{1}{2} \cdot (1 + \varepsilon)mgl \cdot \theta_1^2 + \frac{1}{2} \cdot (1 + \varepsilon)mgl \cdot \theta_2^2 - \varepsilon mgl \cdot \theta_1 \theta_2$$

而在双摆(double pendulum) 问题1.7中, 两个摆直接悬挂在一起. 也会彼此影响. 体系动能和势能为:

$$T_{\text{dp}} = \frac{1}{2}ml^2\dot{\theta}_1^2 + \frac{1}{2}ml^2[\dot{\theta}_1^2 + \dot{\theta}_2^2 + 2\dot{\theta}_1\dot{\theta}_2 \cos(\theta_1 - \theta_2)]$$

$$V_{\text{dp}} = mgl(1 - \cos \theta_1) + mgl(1 - \cos \theta_1) + mgl(1 - \cos \theta_2)$$

小量近似以后, 动能, 势能被近似为二次型:

$$T_{\text{dp}} = \frac{1}{2} \cdot 2ml^2 \cdot \dot{\theta}_1^2 + \frac{1}{2} \cdot ml^2 \cdot \dot{\theta}_2^2 + ml^2 \cdot \dot{\theta}_1 \dot{\theta}_2$$

$$V_{\text{dp}} = \frac{1}{2} \cdot 2mgl \cdot \theta_1^2 + \frac{1}{2} \cdot mgl \cdot \theta_2^2$$

如果把 T_{ij} 和 V_{ij} 写成矩阵, 那么对应的矩阵叫做惯性矩阵(inertial matrix) 和弹性矩阵(elastic matrix):

$$[T_{ij}]_{\text{cp}} = ml^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} ; \quad [V_{ij}]_{\text{cp}} = mgl \begin{bmatrix} 1 + \varepsilon & -\varepsilon \\ -\varepsilon & 1 + \varepsilon \end{bmatrix}$$

$$[T_{ij}]_{\text{dp}} = ml^2 \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} ; \quad [V_{ij}]_{\text{dp}} = mgl \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

从而我们可以发现这两个问题的重大区别在于, 前一个问题动能规整, 势能却具有交叉项. 后一个问题势能规整, 动能却具有交叉项.

当我们遇到一个这样普遍的小振动问题时, 该如何得到其运动的特征. 有两种常见的做法: 一是通过换元来实现动能势能的简化, 这种方法需要较深厚的数学基础, 又最好通过拉格朗日方程来理解一些过程. 我们下面也写出具体的过程, 有兴趣的读者可以做进一步研究; 但也有第二种简易的方法来求解一些直接的量. 我们也在详细讨论的方法之后加上去:

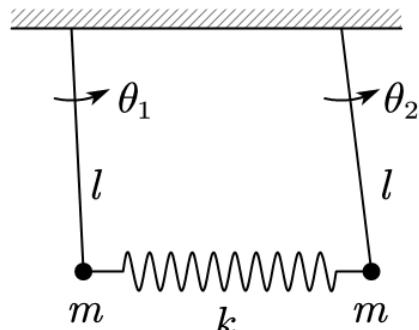


图 1.6: 耦合摆

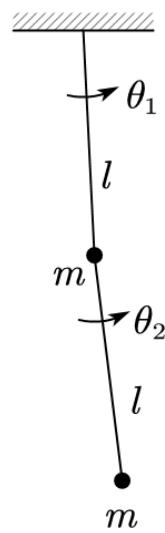


图 1.7: 双摆

1.3.1 基于线性代数与分析力学的简正坐标求解

对一个任意的二次型进行配方并不是难事. 例如 $F(x_1, x_2 \dots)$ 是一个二次型, 那么它必然可以写作:

$$F(x_1, x_2 \dots) = \frac{1}{2}f_{11}x_1^2 + x_1(f_{12}x_2 + f_{13}x_3 + \dots) + (\frac{1}{2}f_{22}x_2^2 + \dots)$$

即 x_1 自己的项, 和别的变量“耦合”的项以及与 x_1 无关的项. 那么做变量代换:

$$x'_1 = x_1 + \frac{f_{12}}{f_{11}}x_2 + \frac{f_{13}}{f_{11}}x_3 + \dots$$

便可以把“耦合”项配方到 x_1 自己的项内, 上式变成:

$$F(x_1, x_2 \dots) = \frac{1}{2}f_{11}x'^2_1 + F'(x_2, x_3 \dots)$$

之后只要继续对 F' 配方, 每次可以得到一个完全平方的项, 余项则是剩下的变量的二次型, 变量数至少减一. 这样就可以实现任意二次型的完全配方. 用这个方法可以选择动能或者势能进行化简. 最终形式, 由上讨论, 只包含平方项, 没有交叉项. 由于这样对应的矩阵是对角矩阵. 我们把这样的配方过程称作**对角化**(diagonalization):

$$T = \sum_i \frac{1}{2}T_i \dot{\delta}_i^2 \quad ; \quad V = \sum_i \frac{1}{2}V_i \delta_i^2$$

注意这里动能和势能的对角化是分开来进行的. 对应的系数 T_i 叫做**惯性系数**(inertial coefficient), 而 V_i 叫做**弹性系数**(elastic coefficient). 同时我们对这些系数必然有以下期待:

- 动能的对角化结果:
 - 非退化正定二次型: 所有 $T_i > 0$.
 - 退化正定二次型: 个别 $T_i = 0$.
 - 不可能出现某个 $T_i < 0$.
- 势能的对角化结果:
 - 非退化正定二次型: 所有 $V_i > 0$.
 - 可以负的二次型: 存在 $V_i < 0$.
 - 退化的二次型: 所有 $V_i \geq 0$ 且也存在 $V_i = 0$.

说明:

动能不可以出现负的惯性系数, 是因为在所有情况下动能都必须大于零. 但是在有的情况下惯性系数可以为零, 即, 即使广义速度不为零, 造成的动能为零, 称作**退化**(degenerate) 的情况. 这要么说明体系存在没有动力学效果的冗余自由度. 例如无限细的杆的自转角就是这样的例子. 去掉这个广义坐标也不影响能量的表达式. 要么说明这个自由度上的动能是一个高阶小量. 只是在目前平衡位置, 系数“偶然”地等于了零, 只要偏离这个位置去振动就会产生一定的动能, 抑或我们在实际问题中需要考虑更高阶泰勒展开产生的广义速度的高次项. 后面这种情况一般就与简谐振动无关了. 再次不予考虑. 故我们此后做假设: **所有小振动问题, 动能必须是非退化正定的.**

势能就更加复杂. 势能相对平衡位置本来就可正可负. 那么上面我们分为三种情况其实就包括了所有情况. 有一点是非常突出的, 如果沿某方向偏离平衡位置时, 势能降低了, 那么原平衡位置就一定是**非稳定平衡**(unstable equilibrium). 在仅仅考虑二次型的情况下, 一旦出现一个负的弹性系数这就成为了必然. 但是即使一个负的弹性系数都没有, 也不能因此就下断言原位置是稳定的. 因为还有二次型**退化**的情况. 它发生在某些弹性系数为零时, 此时高阶泰勒展开项如果是负的, 它依然可以不稳定, 普遍情况得考察多个领头

项.但是,有一点很好判断:就是如果所有弹性系数为正,那么原来的位置就一定是**稳定平衡**(stable equilibrium).因为此时任意的偏离平衡位置都使得势能升高.也只有稳定平衡才能作为小振动问题求解的先决条件.故我们此后做假设:**所有小振动问题,势能必须是非退化正定的.**

但是虽然单独对动能或势能对角化简单,但是两者却不会同时对角化.例如在双摆问题中,对于动能:

$$T_{dp} = \frac{1}{2} \cdot 2ml^2 \cdot \dot{\theta}_1^2 + \frac{1}{2} \cdot ml^2 \cdot \dot{\theta}_2^2 + ml^2 \cdot \dot{\theta}_1 \dot{\theta}_2 = \frac{1}{2} \cdot 2ml^2 \cdot \left(\dot{\theta}_1 + \frac{1}{2}\dot{\theta}_2 \right)^2 + \frac{1}{2} \cdot \frac{ml^2}{2} \cdot \dot{\theta}_2^2$$

但是如果做换元 $\vartheta_1 = \theta_1 + \theta_2/2$, $\vartheta_2 = \theta_2$, 动能倒是没有交叉项了,但原来好好的势能又产生交叉项了:

$$V_{dp} = \frac{1}{2} \cdot 2mgl \cdot \vartheta_1^2 + \frac{1}{2} \cdot \frac{3}{2}mgl \cdot \vartheta_2^2 - mgl \cdot \vartheta_1 \vartheta_2$$

所以,我们这里要追求的其实是**同时对角化**(simultaneous diagonalization) 的技巧.为了进一步研究,我们先来抽象对角化的过程.根据矩阵理论,任何一个坐标变换,其实是一个线性变换:

$$[q'_i] = [A_{ij}][q_j]$$

这个式子是以下两种写法的缩略形式:

$$q'_i = \sum_j A_{ij} q_j \quad \text{or} \quad \begin{bmatrix} q'_1 \\ q'_2 \\ \vdots \\ q'_n \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix}$$

它的逆变换:求出 $[A_{ij}]$ 的逆矩阵 $[B]_{ij}$, 就可以用新坐标表示旧坐标:

$$[q_i] = [B_{ij}][q'_j]$$

而动能,势能都可以用矩阵来表示⁹:

$$T = \frac{1}{2}{}^t[\dot{q}_i][T_{ij}][\dot{q}_j] \quad ; \quad V = \frac{1}{2}{}^t[q_i][V_{ij}][q_j]$$

这样只要把之前的变换代入这两个二次型,就得到了变换之后的二次型:

$$\begin{aligned} T &= \frac{1}{2}{}^t[\dot{q}'_i]^t[B_{ij}][T_{jk}][B_{kl}][\dot{q}'_l] = \frac{1}{2}{}^t[\dot{q}'_i][T'_{il}][\dot{q}'_l] \\ V &= \frac{1}{2}{}^t[q'_i]^t[B_{ij}][V_{jk}][B_{kl}][q'_l] = \frac{1}{2}{}^t[q'_i][V'_{il}][q'_l] \end{aligned}$$

这说明其实在坐标变换以后,惯性和弹性矩阵是按照右乘一个矩阵,再左乘它的转置来进行的.这种对矩阵的变换数学上叫做**合同变换**(congruent transformation):

$$[T'_{il}] = {}^t[B_{ij}][T_{jk}][B_{kl}] \quad ; \quad [V'_{il}] = {}^t[B_{ij}][V_{jk}][B_{kl}]$$

要把 $[T]$ 和 $[V]$ 同时对角化,其实问题的诀窍是求解**本征值问题**(eigenvalue problem):

$$([V_{ij}] + \lambda[T_{ij}])[x_j] = [0]$$

即,确定是否存在常数 λ^{10} 和非零向量 $[x_j]$ 使得两个矩阵组合以后恰好作为系数使得这个向量作为以上方程的解.乍一看这似乎很抽象.其实这背后的物理意义可以做这样的理解:既然是小振动,就有可能找到一些简单的运动模式,此时所有坐标上都是做不同振幅的简谐振动,即:

$$[q_i] = [x_i] \cos \omega t$$

⁹用 ${}^t[q_i]$ 表示转置,即把列向量写作行向量.

¹⁰后面会知道一定非零.

如何写出这种情况下的动力学方程?一种完全严谨的做法就是利用上一章讨论的拉格朗日方程:

$$L = T - V = \frac{1}{2}[\dot{q}_i][T_{ij}][\dot{q}_j] - \frac{1}{2}[\dot{q}_i][V_{ij}][q_j]$$

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} = \frac{\partial L}{\partial q_i} \Rightarrow [T_{ij}][\ddot{q}_j] = -[V_{ij}][q_j]$$

将简谐振动解带入,就得到了:

$$([V_{ij}] - \omega^2[T_{ij}])[x_j] = [0]$$

从而原来的本征值问题中的 λ 代表的含义就是振动角频率平方 ω^2 .那么这个本征值问题又如何求解?首先,上式中方程组存在解且非零的条件为系数行列式为零:

$$\det([V_{ij}] - \lambda[T_{ij}]) = 0$$

这就构成了关于 λ 的 n 次方程,他就一定有 n 个根,不过可以有重根.把这 n 个根带入,我们就可以把对应的 n 个本征向量¹¹ $[x_j]$ 找到,这仅仅是个体力活:求解至多 n 次 n 元线性方程组.一旦求解完毕,我们就找到了 n 个 λ 和 n 个 $[x_j]$,可以记作:

$$([V_{ij}] - \lambda_k[T_{ij}])[B_{j,k}] = [0]$$

最后还需证明一点,就是这 $k = 1, 2 \cdots n$ 个本征值对应的本征向量不仅是独立的,而且还是正交(orthogonal)的.注意这里的正交是指:

$$k \neq k' \Rightarrow {}^t[B_{i,k}][T_{ij}][B_{j,k'}] = 0$$

这个正交性足以说明独立性,毕竟如果 $[B_{i,k}]$ 与 $[B_{i,k'}]$ 不独立,那么前者(非零)就是后者(非零)的常数(非零)倍,以上 ${}^t[B_{i,k}][T_{ij}][B_{j,k'}]$ 就会变成常数乘以后者产生的动能,显然是非零的.所以一旦上式为零,两个矢量正交就足以说明它们不会同向.

证明正交性的依据还是它们的定义.首先我们发现,如果任意两个矢量恰好成为不同本征值下的本征向量:

$$([V_{ij}] - \lambda_1[T_{ij}])[x_{j,1}] = [0], ([V_{ij}] - \lambda_2[T_{ij}])[x_{j,2}] = [0], \quad \lambda_1 \neq \lambda_2$$

构造式子 ${}^t[x_{i,1}][V_{ij}][x_{j,2}]$.一方面先计算前面的向量乘弹性矩阵,一方面也可以先算后面的向量乘弹性矩阵:

$$\begin{aligned} {}^t[x_{i,1}][V_{ij}] &= \lambda_1 {}^t[x_{i,1}][T_{ij}], \quad [V_{ij}][x_{j,2}] = \lambda_2[T_{ij}][x_{j,2}] \\ \Rightarrow {}^t[x_{i,1}][V_{ij}][x_{j,2}] &= \lambda_1 {}^t[x_{i,1}][T_{ij}][x_{j,2}] = \lambda_2 {}^t[x_{i,1}][T_{ij}][x_{j,2}] \\ \lambda_1 \neq \lambda_2 \Rightarrow [x_{i,1}][T_{ij}][x_{j,2}] &= 0 \end{aligned}$$

这实际上是证明了,由 n 个本征向量 $[B_{i,k}]$ 形成的空间根据不同的本征值彼此正交,即使在某个本征值下有多个本征向量,总可以在它们线性组合形成的空间中选取相互正交的向量作为本征向量.这样我们就得到了一组特殊的本征向量,它们符合:

$${}^t[B_{i,k}][T_{ij}][B_{j,k}] = 0 \quad \text{if } k \neq k'$$

而当 $k = k'$ 时上述值一定大于零,这些结果,后面就可以看到,其实就是惯性系数:

$${}^t[B_{i,k}][T_{ij}][B_{j,k}] = T_k$$

现在我们直接把这 n 个 n 行的列向量横向按顺序并置构成矩阵:

$$[B_{ij}] = [[B_{i,1}], [B_{i,2}] \cdots [B_{i,n}]]$$

¹¹注意本征向量的存在性,非重根是显然的,但是如果有个 s 重根,是否也有 s 组独立的解向量?这不是显然的,需要证明,在次略去.一般线性代数的教材上都会有证明方法,不过一般会默认是 $[T_{ij}]$ 为单位矩阵的简单情况,其实这不影响证明过程.

这样以上关系其实就表示为了:

$${}^t[B_{ij}][T_{jk}][B_{kl}] = \text{Diag}(T_i)$$

其中 $\text{Diag}(T_i)$ 表示由 T_i 作为元素的对角矩阵. 同理我们如果考察弹性矩阵, 也有:

$${}^t[B_{i,k}][V_{ij}][B_{j,k'}] = 0 \quad \text{if } k \neq k'$$

$${}^t[B_{i,k}][V_{ij}][B_{j,k}] = \lambda_k T_k = \omega_k^2 T_k = V_k$$

同理用矩阵表示:

$${}^t[B_{ij}][V_{jk}][B_{kl}] = \text{Diag}(V_i)$$

这样我们之前的愿望就实现了, 即动能矩阵与势能矩阵的同时对角化.

让我们来对双摆体系利用上述套路来求解: 首先是对应的本征值问题的系数行列式为零:

$$\begin{aligned} [V]_{ij} &= mgl \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} ; \quad [T]_{ij} = ml^2 \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \\ \det([V_{ij}] - \lambda[T_{ij}]) = 0 &\Rightarrow \det \begin{bmatrix} 2g/l - 2\lambda & -\lambda \\ -\lambda & g/l - \lambda \end{bmatrix} = 0 \\ &\Rightarrow 2(g/l - \lambda)^2 = \lambda^2 \end{aligned}$$

从而得到特征值:

$$\lambda_1 = \omega_1^2 = (2 - \sqrt{2})\frac{g}{l} ; \quad \lambda_2 = \omega_2^2 = (2 + \sqrt{2})\frac{g}{l}$$

带入本征值问题, 分两个本征值来求解本征向量:

$$\begin{aligned} ([V_{ij}] - \lambda_k[T_{ij}])[B_{j,k}] &= 0, \quad k = 1, 2 \\ \begin{bmatrix} 2(\sqrt{2}-1) & -(2-\sqrt{2}) \\ -(2-\sqrt{2}) & \sqrt{2}-1 \end{bmatrix} \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -2(\sqrt{2}+1) & -(2+\sqrt{2}) \\ -(2+\sqrt{2}) & -(\sqrt{2}+1) \end{bmatrix} \begin{bmatrix} B_{12} \\ B_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Rightarrow \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} &= \begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix}, \quad \begin{bmatrix} B_{12} \\ B_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ -\sqrt{2} \end{bmatrix} \end{aligned}$$

这就给出了一个坐标变换矩阵:

$$[B_{ij}] = \begin{bmatrix} 1 & 1 \\ \sqrt{2} & -\sqrt{2} \end{bmatrix}$$

其逆矩阵:

$$[A_{ij}] = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{2}}{4} \\ \frac{1}{2} & -\frac{\sqrt{2}}{4} \end{bmatrix}$$

这就是给出了如下的坐标变换:

$$\begin{cases} \varphi_1 = \frac{1}{2}\theta_1 + \frac{\sqrt{2}}{4}\theta_2 \\ \varphi_2 = \frac{1}{2}\theta_1 - \frac{\sqrt{2}}{4}\theta_2 \end{cases}, \quad \begin{cases} \theta_1 = \varphi_1 + \varphi_2 \\ \theta_2 = \sqrt{2}\varphi_1 - \sqrt{2}\varphi_2 \end{cases}$$

代入便知, 初始的动能和势能分别变为:

$$T = \frac{1}{2} \cdot 2ml^2 \cdot \dot{\theta}_1^2 + \frac{1}{2} \cdot ml^2 \cdot \dot{\theta}_2^2 + ml^2 \cdot \dot{\theta}_1 \dot{\theta}_2 , \quad V = \frac{1}{2} \cdot 2mgl \cdot \theta_1^2 + \frac{1}{2} \cdot mgl \cdot \theta_2^2$$

$$\Rightarrow T = \frac{1}{2} \cdot 2(2 + \sqrt{2})ml^2 \cdot \dot{\varphi}_1^2 + \frac{1}{2} \cdot 2(2 - \sqrt{2})ml^2 \cdot \dot{\varphi}_2^2 , \quad V = \frac{1}{2} \cdot 4mgl \cdot \varphi_1^2 + \frac{1}{2} \cdot 4mgl \cdot \varphi_2^2$$

这就可以看作两个独立的振子, 完成了解耦(decouple). 各自的角频率为:

$$\omega_1 = \sqrt{\frac{4mgl}{2(2 + \sqrt{2})ml^2}} , \quad \omega_2 = \sqrt{\frac{4mgl}{2(2 - \sqrt{2})ml^2}}$$

与之前本征值确定的角频率是一样的.

现在我们来对这个结果产生的物理意义进行解释. 通过坐标变换, 我们发现动能势能形式变得简单了, 所以新的坐标 q'_i 称作简正坐标(normal coordinates). 对应的频率 ω_i 称作简正频率(normal frequencies):

$$\omega_i = \sqrt{\frac{V_i}{T_i}} \quad \text{or} \quad V_i = T_i \omega_i^2$$

于是我们产生了一个疑问, 为什么只要动能势能的形式变得简单, 就可以判定体系可以等效为独立的谐振子的组合? 其实它简单地来自于拉格朗日方程:

$$L = T - V = \sum_i \frac{1}{2} T_i (\dot{q}'_i)^2 - \omega_i^2 q'^2_i$$

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}'_i} = \frac{\partial L}{\partial q'_i} \Rightarrow \ddot{q}'_i = -\omega_i^2 q'_i$$

这就足以给出其解为简谐振动:

$$q'_i = C_i \cos(\omega_i t + \varphi_i)$$

再回到体系的初始广义坐标, 它就会等于这些各自做不同频率振动的简正坐标的组合, 从而体现出非常复杂的非简谐运动特性¹². 具体各个广义坐标的振幅和初相位如何, 由体系的初始条件决定. 但是可以想象, 存在一些简单的振动模式, 它的初始条件决定了, 最后只有一个简正坐标 q'_k 在做振动, 其他的坐标振幅为零. 那么最后原来的坐标也会具有单一的频率 ω_k , 从而每一处运动都是简谐振动, 其实它们为:

$$[q_i] = [B_{ij}] [q'_j] = [B_{i,k}] \cdot C_k \cos(\omega_k t + \varphi_k)$$

所以我们要把体系以统一的频率, 所有地方都做简谐振动的这种振动模式称作简正模(normal mode). “简正”这个词语三次出现在相关术语中, 它可以浅显地被理解为“简单”+“正确”. 也可以更精确地被理解为“简谐”+“正交”. 毕竟在中文语境下, 简谐本意就是“简单”+“和谐”, 而“正确”的“正”本身也含有方向的意味. 英文语境下的“normal”也恰好有着异曲同工之妙.

1.3.2 基于对称性的简正模判定与简正频率求解

如何规避掉较难的数学和分析力学知识? 用简单而优雅的方式快速确定一个多自由度简谐振动的绝大多数特征? 我们可以采用与对称性结合的方式来求解其运动的微分方程.

就比如对于具有更好的对称性的耦合摆问题. 我们不去追求用能量函数来解决问题, 这样势必涉及到分析力学方法. 而是独立地去推导两个质点的牛顿定律. 很容易给出它们为:

$$\begin{cases} \ddot{\theta}_1 = -\omega_0^2(1 + \varepsilon)\theta_1 + \omega_0^2\varepsilon\theta_2 \\ \ddot{\theta}_2 = +\omega_0^2\varepsilon\theta_1 - \omega_0^2(1 + \varepsilon)\theta_2 \end{cases}$$

¹²要知道傅里叶理论可以证明, 用不同频率的简谐振动组合可以无限接近任意光滑函数.

其中 $\omega_0 = \sqrt{g/l}$. 至于这个方程的求解, 猜一个本征模解是常规操作:

$$\begin{cases} \theta_1 = A_1 e^{i\omega t} \\ \theta_2 = A_2 e^{i\omega t} \end{cases}$$

但是在猜解之后, 我们不一定要通过带入原方程, 命系数行列式为零, 再次带入本征频率来求解方程的曲折方程来求解振幅 A_1 和 A_2 的比. 在本题中有一点是显然的. 就是体系具有鲜明的左右对称性. 这使得我们猜想: 是不是在一种本征模式下, 左右两个物体的位移完全相等. 而另一个模式下, 两个物体的位移完全相反?

这的确发生了. 在经典物理情况下, 这其实就对应了质心的左右振动和相对质心的二体在弹性内力下的振动. 一般把两个物体位移相等的模式称作**对称模式**(symmetric mode), 而位移相反的模式称作**反对称模式**(antisymmetric mode). 即:

$$\begin{cases} \theta_{1S} = A e^{i\omega_S t} \\ \theta_{2S} = A e^{i\omega_S t} \end{cases}, \quad \begin{cases} \theta_{1A} = A e^{i\omega_A t} \\ \theta_{2A} = -A e^{i\omega_A t} \end{cases}$$

代入原来的方程就可以看出来: 对称模式给出了:

$$-\omega_S^2 A = -\omega_0^2(1 + \varepsilon)A + \omega_0^2 \varepsilon A \Rightarrow \omega_S = \omega_0$$

$$-\omega_A^2 A = -\omega_0^2(1 + \varepsilon)A - \omega_0^2 \varepsilon A \Rightarrow \omega_A = \omega_0 \sqrt{1 + 2\varepsilon}$$

可以发现, 问题的关键在于如何看出振动模式下各个振幅之间的比例关系. 这个问题似乎取决于问题具体的对称性. 那么, 何谓对称性? 它又是怎样决定体系的振动解的. 我们对一个更复杂的问题来做进一步研究:

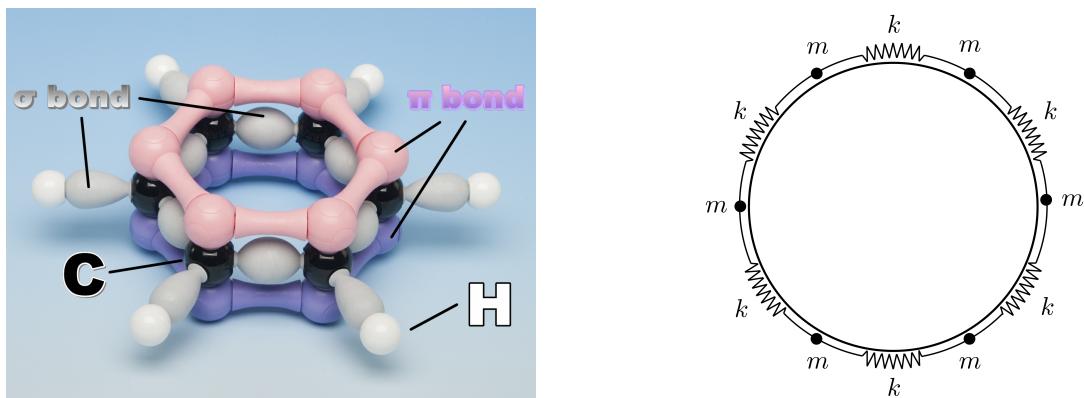


图 1.8: 芬: 6 原子环问题

对于苯分子的振动光谱问题主要的部分就可以做如下简化: 碳原子和氢原子作为一个整体来研究. 设他们是在一个刚性圆环上的质量为 m 的质点. 而彼此之间的相互作用简化为只有相邻的原子才有, 为劲度系数为 k 的线性弹簧. 体系还要求在一个平面上, 甚至一个圆环上作一维运动. 那么这个体系的能量和动力学方程为:

$$T = \sum_{i=1}^6 \frac{1}{2} m \dot{x}_i^2, \quad V = \sum_{i=1}^6 \frac{1}{2} m (x_{i+1} - x_i)^2$$

$$m \ddot{x}_i = -2kx_i + kx_{i+1} + kx_{i-1}, \quad i = 1, 2 \dots 6$$

在上式中我们约定 $x_0 = x_6$, $x_1 = x_7$. 那么这个微分方程必然存在某种解.

我们讨论第一个问题:什么叫做对称性?表面上看上去,苯分子似乎具有三种对称性:一是绕竖直轴旋转 60° 的六重轴对称性:

$$C_6 : \quad x_1 \rightarrow x_2, x_2 \rightarrow x_3, x_3 \rightarrow x_4, x_4 \rightarrow x_5, x_5 \rightarrow x_6, x_6 \rightarrow x_1$$

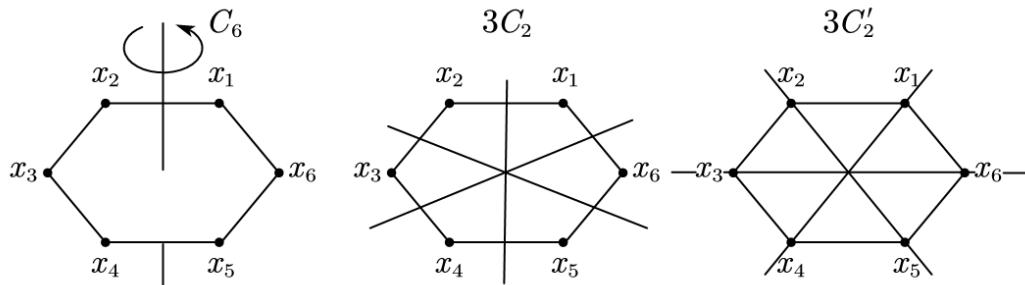


图 1.9: 6 原子环对称性

第二种是以六边形对边连线中点构成的二重轴. 绕它们转 180° 后体系才会复原, 典型的代表性元素:

$$C_2 : \quad x_1 \leftrightarrow -x_2, x_3 \leftrightarrow -x_6, x_4 \leftrightarrow -x_5$$

第三种是六边形的对角线作为二重轴. 典型的代表元素:

$$C'_2 : \quad x_1 \leftrightarrow -x_5, x_2 \leftrightarrow -x_4, x_3 \leftrightarrow -x_3, x_6 \leftrightarrow -x_6$$

对称性永远都是指在某一种变换下, 体系的某一些量具有不变的特性. 那么我们发现, 体系的能量函数就具有这样的特性. 以上无论哪一个变换, 在交换了各个 x 的角标以后, 能量函数的表达式却恰好回到原来的表达式. 但是对于一些不属于这个体系的对称性操作, 比如单个 $x_1 \leftrightarrow x_2$. 能量的表达式就变了. 所以我们说体系的能量函数具有对称性.

但是体系的形状是不会具有对称性的. 之前的示意图完全是静止状态下的苯分子¹³. 但是我们现在就是要考虑它的振动. 作为能量函数的结果, 其振动解是否一定具有对称的形式? 答案显然是否定的, 因为体系的运动由动力学方程和初始条件来决定. 即使动力学方程是对称的, 只要其初始条件具有不对称性, 其解就至少有一些时刻不再对称. 这就是说, 对称性的原因不一定产生对称的结果. 这种现象就是一种广义上的对称性破缺(symmetry breaking) 现象.

但是, 我们指出: 如果把对称性操作理解为算符(operator), 它作用在一组现成的解上可以得到一种不同的运动. 例如典型的 C_6 对称性操作, 它把每一个原子的运动状态转移到了它的下一个原子上. 那么我们就可以找到体系演化规律和某个对称性算符的共同本征模式. 这就是说, 存在一些振动模式, 它即是某个本征频率 ω 下的模式, 又恰好在某个对称性操作下 (例如 C_6) 产生的运动和原来仅仅相差一个常数, 这个常数称作本征值(eigenvalue):

$$C[x_i] = \lambda[x_i]$$

原因是简单的: 首先如果找到了体系的一个 ω 下的本征模式, 那么直接用这个对称性作用在这个模式下任意次, 显然得到的新的运动全都是原来的动力学方程的解, 就连频率也不会改变, 它们就全都是对应到同一个本征频率下的本征模式, 称作简并(degenerate) 在一个频率上的模式. 那么可以考虑得到的所有模式的组合, 由于对称性操作连续作用有限次以后必然回到初始状态. 故模式数必然有限. 这样就一定可以找到合适的状态.

¹³基态体系的对称性会完全等于能量函数的对称性, 这一点也是可以被证明的.

例如,对于 C_6 ,初始状态如果是 $[x_i]_1$,那么至多连续作用 C_6 五次,第六次就回到初始状态了:

$$[x_i]_2 = C_6[x_i]_1, [x_i]_3 = C_6^2[x_i]_1, [x_i]_4 = C_6^3[x_i]_1, [x_i]_5 = C_6^4[x_i]_1, [x_i]_6 = C_6^5[x_i]_1$$

$$[x_i]_1 = C_6^6[x_i]_1$$

这样我们就可以构造出以下的共同本征函数:

$$[x_i]_{(0)} = [x_i]_1 + [x_i]_2 + \cdots + [x_i]_6 \Rightarrow C_6[x_i]_{(0)} = 1 \cdot [x_i]_{(0)}$$

$$[x_i]_{(1)} = [x_i]_1 \cdot (w^1)^1 + [x_i]_2 \cdot (w^1)^2 + \cdots + [x_i]_6 \cdot (w^1)^6 \Rightarrow C_6[x_i]_{(1)} = w^{-1} \cdot [x_i]_{(1)}$$

$$[x_i]_{(2)} = [x_i]_1 \cdot (w^2)^1 + [x_i]_2 \cdot (w^2)^2 + \cdots + [x_i]_6 \cdot (w^2)^6 \Rightarrow C_6[x_i]_{(2)} = w^{-2} \cdot [x_i]_{(2)}$$

...

$$[x_i]_{(5)} = [x_i]_1 \cdot (w^5)^1 + [x_i]_2 \cdot (w^5)^2 + \cdots + [x_i]_6 \cdot (w^5)^6 \Rightarrow C_6[x_i]_{(5)} = w^{-5} \cdot [x_i]_{(5)}$$

而 $w^6 = 1$, $w = e^{i\pi/3}$ 是六次单位根.把某个本征频率下的所有本征模式拿出来,对应的 C_6 算符的本征值也只可能在以上六种值:六次单位根的若干幂次中选取.

同理,如果考虑 C_2 或者 C'_2 对称性,它们的本征值则更少,作为二次单位根仅仅只有 1 或 -1 两种可能性.

接下来我们可以先分析单态(singlet).假设某一个角频率下仅仅只有一重简并,即只有一个本征模式.那么这个模式自己就必须同时成为 C_6 , C_2 , C'_2 三者的本征态.对于后两者有四种可能性:

$$C_2[x_i] = \pm [x_i], \quad C'_2[x_i] = \pm [x_i]$$

$$\Rightarrow \begin{bmatrix} -x_2 \\ -x_1 \\ -x_6 \\ -x_5 \\ -x_4 \\ -x_3 \end{bmatrix} = \pm \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}, \quad \begin{bmatrix} -x_5 \\ -x_4 \\ -x_3 \\ -x_2 \\ -x_1 \\ -x_6 \end{bmatrix} = \pm \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}$$

如果两个本征值分别是 +1 和 -1.那么对应的模式为:

$$[x_1, x_2, x_3, x_4, x_5, x_6] = x_1[+, -, +, -, +, -]$$

这样的模式恰好也是 C_6 本征值为 +1 的模式.对应的运动为三个相间的原子编组,另外三个也编组,两组原子反向运动的振动模式.用 C_2 和 C'_2 的本征值正负来标示这种模式,代入原来的动力学方程,得到:

$$-m\omega_{+-}^2 A = -2kA - kA - kA \Rightarrow \omega_{+-} = \sqrt{\frac{4k}{m}}$$

再来考虑两个本征值都是 -1 的情况,这样恰好有:

$$[x_1, x_2, x_3, x_4, x_5, x_6] = x_1[+, +, +, +, +, +]$$

也就是所有原子都沿相同方向作相同运动.对应 C_6 本征值为 +1.显然这样的运动就是整体作匀速旋转的运动:

$$\omega_{--} = 0$$

但是另外两种情况略有不同. 首先是两个本征值都是 +1 的情况. 此时得到:

$$[x_1, x_2, x_3, x_4, x_5, x_6] = x_1[+, -, 0, +, -, 0]$$

抑或是两个本征值分别为 -1 和 +1 的状态. 得到:

$$[x_1, x_2, x_3, x_4, x_5, x_6] = x_1[+, +, 0, -, -, 0]$$

我们发现, 这些运动直接并不是 C_6 的本征态. 但是这些运动有它存在的合理性. 事实上画图就会发现这些运动都是可以符合动力学规律的:

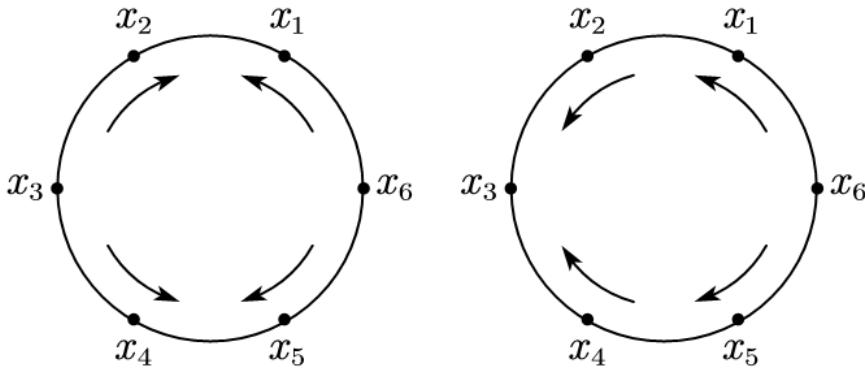


图 1.10: 二重简并的振动模式

每种模式的 x_1 的振动就能代表 x_1, x_2, x_4, x_5 四个质点的振动. 而 x_3, x_6 在这两种模式下都是静止的. 对于第一种模式, 对 x_1 列牛顿定律:

$$-m\omega_{++}^2 A = -2kA - kA \Rightarrow \omega_{++} = \sqrt{\frac{3k}{m}}$$

$$-m\omega_{-+}^2 A = -2kA + kA \Rightarrow \omega_{-+} = \sqrt{\frac{k}{m}}$$

正因为这些运动合理, 但又不是 C_6 的本征态, 所以其实它们根本就不像最初假设的那样, 是两个本征频率 ω_{++}, ω_{-+} 下的单态, 而是各存在二重简并. 显然我们把 C_6 作用在这两个态上各一次就是与它简并在同一个频率上的另一个独立振动模式, 如果设 $x_1 = Ae^{i\omega t}$:

$$[x_i]_{++} = [+, -, 0, +, -, 0]Ae^{i\omega_{++}t} \Rightarrow C_6[x_i]_{++} = [0, +, -, 0, +, -]Ae^{i\omega_{++}t}$$

$$[x_i]_{-+} = [+, +, 0, -, -, 0]Ae^{i\omega_{-+}t} \Rightarrow C_6[x_i]_{-+} = [0, +, +, 0, -, -]Ae^{i\omega_{-+}t}$$

恰好, 作用两次是徒劳的. 两次旋转以后的运动可以用前两次的运动组合形成:

$$C_6^2[x_i]_{++} = -[x_i]_{++} - C_6[x_i]_{++} : [-, 0, +, -, 0, +] = -[+, -, 0, +, -, 0] - [0, +, -, 0, +, -]$$

$$C_6^2[x_i]_{-+} = -[x_i]_{-+} - C_6[x_i]_{-+} : [-, 0, +, +, 0, -] = -[+, +, 0, -, -, 0] - [0, +, +, 0, -, -]$$

从而我们就完整地找到了体系的六个本征模式:

$$\omega_{--} = 0 : [x_i]_{--} = [+, +, +, +, +, +]vt$$

$$\omega_{-+} = \sqrt{\frac{k}{m}} : [x_i]_{-+} = [+, +, 0, -, -, 0]Ae^{i\omega_{-+}t}, C_6[x_i]_{-+} = [0, +, +, 0, -, -]Ae^{i\omega_{-+}t}$$

$$\omega_{++} = \sqrt{\frac{3k}{m}} : [x_i]_{++} = [+, -, 0, +, -, 0]Ae^{i\omega_{++}t}, C_6[x_i]_{++} = [0, +, -, 0, +, -]Ae^{i\omega_{++}t}$$

$$\omega_{+-} = \sqrt{\frac{4k}{m}} \quad : \quad [x_i]_{+-} = [+,-,+,-,+,-] A e^{i\omega_{+-} t}$$

事实上我们只看到了问题的一个局部. 关于复杂体系小振动背后的对称性分析是一个内容非常丰富的专题. 它涉及到关于对称性的群论(group theory) 和对于振动在对称性操作下演化的群表示论(group representation theory). 它在原子分子物理, 固体物理, 乃至粒子物理中都有着非常广泛的应用.

对于苯分子的振动模式我们也做了过多的简化, 下面的表是较完整地考虑碳原子和氢原子间共 12 个原子, 在三维空间中各做三自由度运动而形成的共 $36 - 6 = 30$ 个振动模式. 减 6 是因为有 6 个自由度是整体的平动和转动而不是振动. 根据对称性, 一共有 20 种振动模式. 其中有 10 种为二重简并. 分子的振动光谱一般在红外线波段:

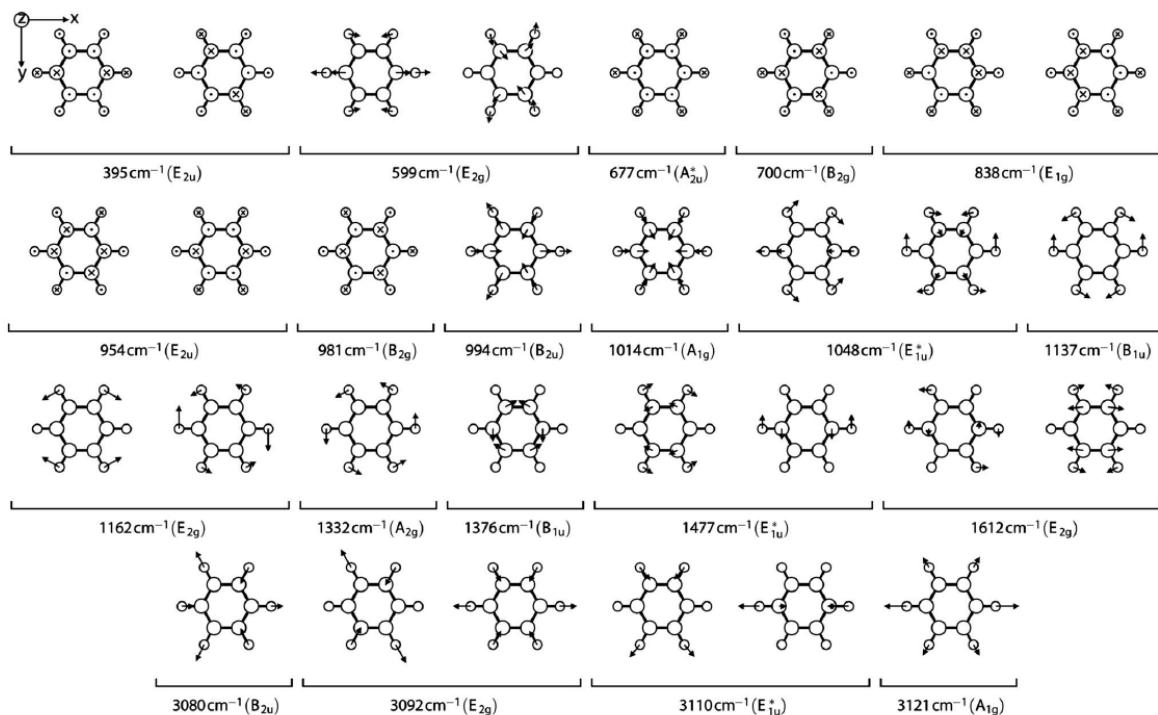


图 1.11: 苯分子的完整振动模式

利用原子间作用力的经验函数就可以得到各个本征角频率 ω . 而由量子力学中的原理, 振动的能量实际上是量子化的:

$$E_n = \left(n + \frac{1}{2} \right) \hbar \omega$$

苯分子的吸收光谱和发射光谱中对于相邻两个能级能量差的光子就会有共振吸收的现象:

$$\Delta E = h\nu = \hbar\omega$$

可见其实就是当电磁波的频率恰好与振动频率吻合时发生共振吸收:

$$\nu = \frac{\omega}{2\pi}$$

1.4 非线性摄动 *

1.5 格波

让我们考虑两个十分经典的**固体物体**(solid state physics) 问题: 声音的本质是什么? 固体热运动的本质是什么? 我们可能会得到一个惊人的答案: 两者的本质, 都是晶格的振动.

固体, 一般是晶体, 它的描述方法, 也需要照顾到电子的行为, 也需要照顾余下的原子实---它占据着主要的质量与动能---的运动形式. 如果把目光仅仅放在原子实上, 认为原子实是在格点的平衡位置附近做小振动. 而这个运动的动力学成因, 必然是来自于初始条件和原子之间可以类比为弹簧的线性作用力. 这样去思考问题的话, 也许具体的计算还是困难的, 但至少我们会有以下三个结论:

一: 这个运动一定是牵一发而动全身的. 一个原子的振动会带动其它原子的振动, 一个地方的振动会带动另一个地方的振动, 这其实就是固体可以传热和传声的原理. 而且不出意外的, 一般传热性能越好的固体, 声速也就越快. 天然材料中金刚石就是在两个方面都具有卓越性能的典型: 它的声速高达 12000m/s , 在元素单质方面仅仅次于在自然界不以单质形式存在的金属铍 (13000m/s), 是铁的两倍多. 而热导率则是达到了惊人的 $2000\text{W/m}\cdot\text{K}$, 为金属中导热性能优异的银的约五倍.

二: 有规则的这种运动形成了声波. 也就是宏观问题中看似连续的声波, 在微观看来其实就是一个个分离的原子在因为弹性力而做多自由度的小振动, 而不再具有连续性. 这种运动就叫做**格波**(lattice wave).

三: 无限不规则的声波进行叠加就形成了热运动. 基于这一点, 我们其实不应该把热运动分解为每一个原子各做一个独立简谐振动, 这样的统计方法是不能给出正确的关于固体的热学结果的, 因为这些简谐振动根本就不“独立”. 从爱因斯坦时代开始, 把热运动分解为真正声波的叠加, 并且对声波的能量进行量子化, 称作**声子**(phonon), 这才真正能开始解释关于固体的热容等等初步的实验结果.

这一节我们暂时无法得出这些深刻结果的计算过程. 但至少, 我们可以尝试在完全经典的情况下对两个简化问题给出完整的动力学求解方法.

第一个问题就是**一维原子链**(1D atomic chain) 问题. 把质量为 m 的物块和劲度系数为 k 的弹簧无限串接形成下图所示的体系.

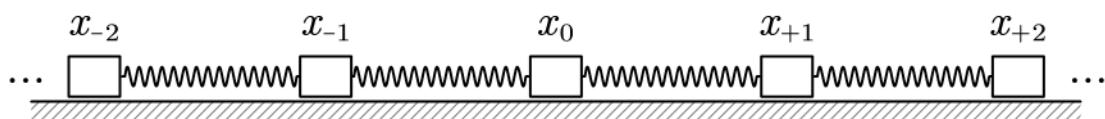


图 1.12: 一维原子链

这样的体系的动力学方程写作:

$$m\ddot{x}_i = -2kx_i + kx_{i+1} + kx_{i-1}$$

这样一个方程的求解其实非常类似于之前的对称性分析法. 这里体系的对称性为平移一个单元. 那我们就设想体系的状态为平移算符的本征态, 即每一个物块的振动都恰好是其前一个(左边的那个, 指标比它小一的那个)的 λ 倍. 只不过这里的 λ 其实是一个相位因子 $e^{i\varphi}$ 罢了. 那么我们猜一个本征频率为 ω , 在平移算符下本征值为 λ 的解, 显然 ω 就会是 λ 的函数:

$$x_i = \lambda^i A e^{i\omega t}$$

代入原方程, 命 $\omega_0 = \sqrt{k/m}$ 就可以得到:

$$\frac{\lambda + \lambda^{-1}}{2} = 1 - \frac{\omega^2}{2\omega_0^2}$$

若 λ 为实数, 不管是 $|\lambda| > 1$ 还是 $|\lambda| < 1$ 都不是我们希望得到的结果, 这样在 $i \rightarrow +\infty$ 和 $i \rightarrow -\infty$ 两个方向一个振幅指数衰减, 一个指数爆炸. 那么根据因果律和能量守恒可以判断, 指数爆炸那一侧是受到外力开始起振的那一侧, 这个波根本就不能在原子链上传播, 而是传着传着振幅就衰减掉了. 此时等号左侧是一个绝对值大于 1 的实数. 故我们发现当 $\omega \geq 2\omega_0$ 时对应的波无法在一维原子链上传播. 这个频率 $\omega_c = 2\omega_0$ 就叫做截止频率(cut-off frequency). 恰好为截止频率时, $\lambda = -1$, 也就是这样的波虽然可以形成, 但任何相邻两个原子振动方向彻底反向, 它其实是一种“驻波”, 原则上也没有在传播.

那么就只剩下 $\omega \leq \omega_c$ 的波可以在一维原子链上传播. 把 λ 写成 $e^{i\varphi}$. 我们得到:

$$\cos \varphi = 1 - \frac{\omega^2}{2\omega_0^2} \Rightarrow \omega = 2\omega_0 \left| \sin \frac{\varphi}{2} \right|$$

相邻两个原子的相位差应当是一个 $\varphi \in (-\pi, \pi)$ 内的数. 这是因为如果 A 比 B 相位大一个 $\pi + \delta$, 那么也可以解释为 A 比 B 相位小一个 $\pi - \delta$. 所以如果是截止频率的临界情况, 我们也只能说相邻的原子反相, 至于谁比谁大也是无从判断的, 这是格波的特性.

对于上式我们更喜欢写为色散关系(dispersion relation)的形式. 就是说把自变量改为波矢(wave vector). 设相邻原子距离为 l , 那么产生的相位差 φ 可以由波矢来产生:

$$Kl = \varphi, \quad K \in \left(-\frac{\pi}{l}, \frac{\pi}{l} \right)$$

上面这个 K 应当处于的空间称作布里渊区(Brillouin zone). 它告诉我们, 在一维原子链上传播的格波在频率和波矢上都是受限的. 其函数图像和表达式如图1.13.

色散曲线上每一个点都代表一种能够在一维原子链上传播的振动模式:

$$x_i = A e^{i(\omega t + iKl)}$$

此时两个极限是值得关注的:

一是, 在短波极限: 波长最短就是当 $Kl = \pi$ 时, 此时波长 $\Lambda = 2\pi/K = 2l$ 时, 向左和向右延伸的色散曲线到达最高点 Γ . 但尤其要注意两个 Γ 点实质上就是同一个点, 对应完全相同的状态. 可以想想把第一布里渊区卷做一个圆柱, 让两个 Γ 点重合. 对应的相邻原子彼此反相的“驻波”形式. 也就是说如果让 $K > 0$ 的右行波的 K 持续增加以跨越 Γ 点, 波就演化为左行波了. 一个中缘由在于 K 的正负指称的速度其实是相速度(phase velocity), 但在判断波的传播方向时, 其实群速度(group velocity)才是更加合理的选择. 它被定义为:

$$v_g = \frac{d\omega}{dK}$$

其中的原理我们将在流体的相关波动章节中简介.

二是, 在长波极限: $\Lambda \gg l$. 此时 Kl 时一个小量, 可以对 \sin 使用近似, 物理上代表相邻两个原子几乎是以相同的方式运动, 微观波变成了宏观的波:

$$\omega = \omega_0 Kl$$

这就给出了波的传播速度:

$$v = \frac{\omega}{K} = \omega_0 l = \sqrt{\frac{k}{m}} l$$

这个结果与宏观波速公式 $v = \sqrt{E/\rho}$ 是完全一致的. 且看之后弹性体章节的分析.

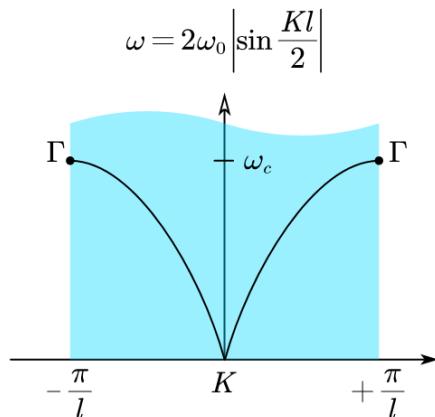


图 1.13: 色散关系

最后我们还看看双原子链(diatomic chain)与以上单原子链(monoatomic chain)产生的并不平凡的区别. 双原子链就意味着虽然其平移周期还是 l , 但每一个单元内部出现了两个待求解的坐标. 我们让两个原子质量分别为 M, m , 弹簧劲度系数依然为 k :

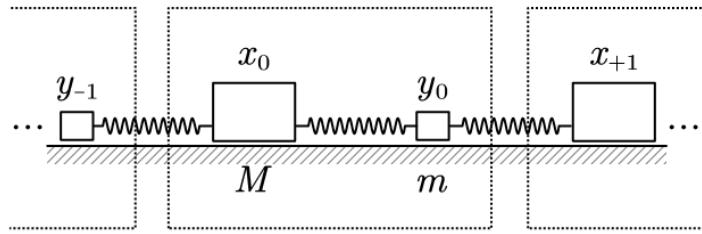


图 1.14: 双原子链

那么对应的动力学方程写作:

$$M\ddot{x}_i = -2kx_i + ky_i + ky_{i-1}$$

$$m\ddot{y}_i = -2ky_i + kx_i + kx_{i+1}$$

由于有单原子链的经验. 我们让 $\omega_0 = \sqrt{k/2M}$ ¹⁴, 而 $m = \beta M$. 并直接设解为:

$$x_i = A e^{i(\omega t + iKl)}, \quad y_i = B e^{i(\omega t + iKl)}$$

其中 A 和 B 是复振幅可以差一个相位. 那么这就给出:

$$(4\omega_0^2 - \omega^2)A + -2\omega_0^2(1 + e^{-iKl})B = 0$$

$$-2\omega_0^2(1 + e^{iKl})A + (4\omega_0^2 - \beta\omega^2)B = 0$$

这个方程组 A, B 的解不为零的条件就是:

$$(4\omega_0^2 - \omega^2)(4\omega_0^2 - \beta\omega^2) = 4\omega_0^4(1 + e^{iKl})(1 + e^{-iKl}) = 16\omega_0^4 \cos^2 \frac{Kl}{2}$$

在 β 是个小量的情况下, 我们可以近似地求解这个方程. 在任意 K 下这个方程与单原子链不同, 它 ω^2 总是有两个解. 第一个解发生在 ω 与 ω_0 量级相当的情况下, 此时 $4\omega_0^2 - \beta\omega^2 \approx 4\omega_0^2$. 近似可以得到:

$$\omega = 2\omega_0 \left| \sin \frac{Kl}{2} \right|$$

可以发现这个解与只含 M 的单原子链的结果并没有什么区别. 所以其实是这种模式下, 主要的动力学效果由 M 承担, m 则几乎平衡在两侧的 M 中间. 两个相邻的 M 之间的两段弹簧几乎就是直接串联在一起, 同时伸长同时缩短, 感受不到很轻的 m 的影响.

而另外一个解则发生在 $\omega \gg \omega_0$ 的极端情况下, 此时 $4\omega_0^2 - \omega^2$ 本是一个绝对值很大的负数, 它 $\approx -\omega^2$. 但是与另一个因子 $4\omega_0^2 - \beta\omega^2$ 相乘时又回到正的 ω_0^4 量级的等式右侧了. 从而乘的因子其实很接近零:

$$4\omega_0^2 - \beta\omega^2 \approx 0 \Rightarrow \omega \approx \frac{2\omega_0}{\sqrt{\beta}}$$

为了更精确地找到这个频率与 K 的函数关系, 我们把原方程的 $4\omega_0^2 - \omega^2$ 项除到等号右边, 并近似:

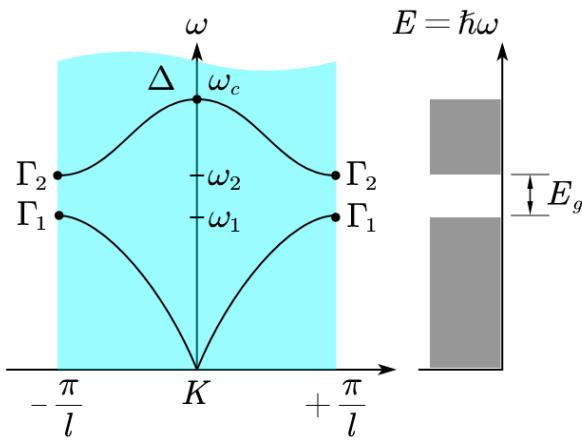
$$\omega \approx \frac{2\omega_0}{\sqrt{\beta}} \cdot \sqrt{1 + \beta \cos^2 \frac{Kl}{2}}$$

¹⁴即先忽视 m 的存在让弹簧直接串联.

这样的两组解看似近似,但是其实在特征点处的性质和坐标又是完全准确的. 特征点就是指波矢为零 $K = 0$ 和波矢为 $\pm\pi/l$ 的布里渊区中心和边界上的点. 我们把两组 $\omega(K)$ 关系画成色散曲线如图1.15. 对应的特殊点 $\Gamma_1, \Gamma_2, \Delta$ 对应的特殊频率为:

$$\omega_1 = 2\sqrt{\frac{k}{2M}}, \quad \omega_2 = 2\sqrt{\frac{k}{2m}}$$

$$\omega_c = 2\sqrt{\frac{k}{2\mu}} (\mu = \frac{Mm}{M+m})$$



现在终于可以发现, 多出来的第二组解, 对应了在原来的色散曲线上方又多出来一支. 这一支在 Δ 的行为具有代表性: 即使 $K = 0$, 即相邻两个单元之间完全没有相位差. 但是依然有很大的角频率. 也就是这个振动并没有传播, 而从角频率的表达式来看, 其实就是每一个单元内部两个物体在弹性力作用下做二体小振动. 而从 Γ_2 到 Δ 点这样一段色散曲线对应的各个波动模式下, 十分显著的一点就是轻的原子一定要主动振动起来而且与相邻的重原子根接近反相而不是同相. 最边缘的 Γ_2 点则只有轻的原子在振动. 这样的模式的频率, 其实与上一节讨论过的分子振动频率相当, 就在光的红外线频段, 故把这样的一支振动模式称作色散曲线的光学支(optical branch).

所以原来单原子链的保留支, 特点是从 $\omega = 0, K = 0$ 延续到布里渊区边缘的 Γ_1 , 就相应的被称作声学支(acoustic branch). 因为在长波极限 $K \rightarrow 0$ 下, 这个模式的轻重原子之间是没有相位差的, 而不是像光学支那样反相, 这样就相当于宏观地看整个原子链在较大范围内做宏观平移. 形成的波就是常规意义上的声波. 在声学支模式下, 轻的原子跟随着重的原子而运动, 相邻轻重原子更加接近同相而不是反相. 而极限点 Γ_1 对应的情况是只有重原子在振动, 轻原子全部静止了.

最后我们指出, 简单的计算产生的结果中还蕴含着意义非凡的一个物理图像. 那就是能带(energy band)结构的产生. 在这里我们着眼的是晶格上的波动, 色散曲线上每一个点都代表一种可能的波动模式, 其振幅, 在经典物理中, 由初始条件决定, 且是可以连续变化的. 但是量子理论预言, 这个能量实则在低能量时量子化:

$$E_n = \left(n + \frac{1}{2}\right) \hbar\omega$$

从而色散曲线的这一副图, 实际上也代表量子化的振动的元激发(elementary excitation)的能量单位与 K 的关系图. 更有甚者, 如果我们不是去考虑晶格振动的动能加势能的元激发, 而是去考虑在晶格间运动的电子, 它的能量也会产生非常类似的但更复杂的能带结构¹⁵, 毕竟电子在其中的描述也应当使用物质波, 它具有波矢 K 和相应的能量 $E(K)$. 而为什么这个 $E(K)$ 的函数也会产生类似的行为, 这一点可以纯粹从电子满足的薛定谔方程(Schrödinger equation) 和对称性分析出发得到. 在介质中的电子的波矢 K 其实也是受限的. 而同一个 K 对应的能量也可以有多个, 分别在不同分支的色散曲线上. 故我们把每一条色散曲线对应的可取能量区间叫做允带(allowed band). 而允带之间没有对应状态的能量区间叫做禁带(forbidden band). 两个相邻的允带之间的禁带宽度就是带隙(band gap).

这样一种全新的看待振动的能量, 电子的能量的理论就是能带论. 尤其是把电子视作物质波, 利用量子力学完整地计算固体中的电子的运动, 从而非常好地解释了历史上残留的金属电导率, 热导率相关的疑问并很好地带动半导体物理学的成熟的理论, 以它的主要缔造者: 瑞士-美国物理学家布洛赫命名为布洛赫理论(Bloch theory).

能带结构是一种介于自由电子(free electron) 和紧束缚电子(tight-binding electron) 之间的存在形式. 在真空中的自由电子满足:

$$E = \frac{\hbar^2 K^2}{2m}$$

¹⁵会往上产生更多能带.

其能量完全可以连续的改变.但是在束缚态上,如氢原子外的电子,简单地由玻尔模型,电子的能量被量子化为能级结构:

$$E = -\frac{me^4}{8n^2\varepsilon_0^2h^2} = -\frac{13.6\text{eV}}{n^2}$$

那么在金属中的电子,微观上自由:可以脱离金属原子的束缚而传导,但宏观上,本质又是被束缚的:被所有原子核共同产生的吸引力束缚在金属内部.所以就体现出一种折衷:紧束缚的能级展开为能带,具有一定的能量连续变化的范围.但是能带与能带之间依然是分立的,没有像真空那样连成一片.

第二章 万有引力

2.1 有心力下质点运动

天体运动中起到核心作用的相互作用力都是**有心力**(central force),事实上不光是天体运动,任意两个可以近似为质点的物体之间的相互作用力,根据牛顿第三定律的要求,其受力方向都必须沿着两个物体的连线方向.那么我们只要做两个要求,这就构成了一个有心力问题:

一是,这个力必须是保守力,即,它必须由势能生成:

$$V(\mathbf{r}_1, \mathbf{r}_2) : \quad \mathbf{F}_1 = -\nabla_1 V, ; \mathbf{F}_2 = -\nabla_2 V$$

根据我们之前的说法,根据对称性或牛顿第三定律,这个势能其实就是两个质点连线距离 R 的函数 $V(R)$.而以 1 为中心向 2 引 \mathbf{R} 矢量,则:

$$\mathbf{F}_2 = -V'(R)\mathbf{e}_R = F(R)\mathbf{e}_R$$

二是,中心物体 1 必须不动或者是近似不动. 1 不动是指有外力作用在 1 上以维持其静止.但 2 上不应该有这样的外力,而仅仅是在 1 对 2 产生的 \mathbf{F}_2 作用下做运动. 1 近似不动是比如考虑太阳系这种典型情况,太阳虽然受到多个行星对它的万有引力,但是由于自己质量过重从而近似是不动的. 即使是地球月亮构成的二体问题,地球和月亮并不一定能认为都不动,我们也有相应的转化为有心力问题的方法,见后. 最后当然,也有一些更简单的情况,2 受到一些更复杂的体系对它的力构成了有心力 $F(R)$,比如弹性绳对绳端质点的拉力.

在有心力场中运动的物体,其基本的牛顿定律可以写作:

$$m\ddot{\mathbf{r}} = F(r)\mathbf{e}_r$$

可以把这个方程作为一切讨论的起点. 或者进一步,我们先发现:

2.1.1 运动的一般特征

首先,有心力场中运动的质点必然在一个平面上运动.这一点既可以通过初始条件判断 \mathbf{r}_0 和 \mathbf{v}_0 的平面就是运动平面.也可以通过下面的角动量守恒来发现这一点.正是这样,在运动平面上建议极坐标是方便的,用极坐标写出的牛顿定律分量式为:

$$\begin{cases} m(\ddot{r} - \dot{\theta}^2 r) &= F(r) \\ m(\ddot{\theta}r - 2\dot{\theta}\dot{r}) &= 0 \end{cases}$$

除了牛顿定律以外,还有两个可以变通使用的结论: 角动量和能量都是守恒的:

角动量守恒的原因是有心力对力心没有力矩. 角动量的方向垂直于运动平面. 极坐标下大小写为:

$$L = m\dot{\theta}r^2 = \text{Const.}$$

而习惯上通常把质量 m 省略后的量单独提出来, 它实际上是面积速度的两倍:

$$\dot{\theta}r^2 = h = 2 \frac{dA}{dt}$$

也就是说, 面积速度不变是有心力问题的普遍特征, 这也是开普勒第二定律的一种推广形式.

第二个守恒量是哈密顿量:

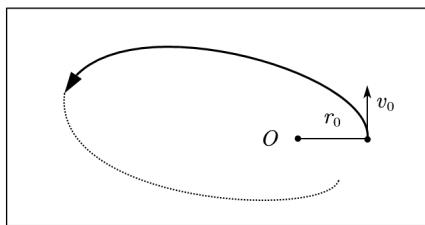
$$H = \frac{1}{2}m(\dot{r}^2 + \dot{\theta}^2 r^2) + V(r) = \text{Const.}$$

守恒到的值直接叫做能量 E . 通常我们喜欢把这个式子移项写作:

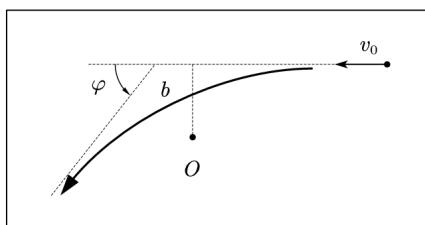
$$\dot{r}^2 + \dot{\theta}^2 r^2 = \frac{2[E - V(r)]}{m}$$

2.1.2 有心力问题的求解

有心力问题的求解可以指一定初始条件下, 轨迹方程 $r(\theta)$ 的求解. 也可能需要进一步求出运动方程 $r(t)$ 和 $\theta(t)$ 的具体形式来. 通常还有两种常见的提法:



一是定态问题(stationary state problem)的求解: 例如直接从距离力心 r_0 给一个垂直半径方向的速度 v_0 . 求解之后的运动. 不难理解, 只有引力(attraction)才能形成束缚的态, 斥力(repulsion)就不行. 而要想这个态在动力学上是“稳定”的, 后面就知道还需要这个力符合一定的条件. 这一个问题在微观上也存在对应的量子力学版本.



二是散射问题(scattering problem)的求解: 粒子从无穷远出发以一定的初速度 v_0 和瞄准距离(impact parameter) b 入射, 经过一个有心力影响后轨迹发生偏移. 求解过程中, 尤其是散射末态的粒子状态. 由于轨迹必然存在对称性, 出射过程基本上就是入射过程在另一个角度处的时间反演. 所以经典物理中关注的量主要是散射角 φ 的大小, 它被定义为散射前后运动方向的夹角. 不难看出, 不管是引力还是斥力都会造成散射. 但是散射问题对有心力也存在一个要求, 它在无穷远处必然对粒子没有影响, 即, 不管是势能还是力都要衰减到零. 这个要求实际上叫做短程力(short-ranged force). 但它的严格定义和个别性质也需要上升到量子力学的散射问题才能解释清楚. 比如万有引力其实就不属于短程力而是长程力(long-ranged force), 而超过平方的幂次反比例都是短程的. 尽管前者的散射问题也经常被拿来讨论. 散射问题不仅存在量子力学版本, 引力场下的光线偏转问题也是一个非常经典的广义相对论问题.

图 2.1: 定态问题与散射问题

而目前求解经典的有心力问题通常有两种思路:

1. 有效势能法

利用两个守恒量, 实际上就等于把原来的二阶微分方程求解的问题转化为了一阶微分方程的问题. 然后再利用角动量守恒式子把 $\dot{\theta}$ 用 r 来表示并代入能量守恒. 我们得到:

$$\frac{1}{2}m\dot{r}^2 = E - V(r) - \frac{mh^2}{2r^2} = E - V_{\text{eff}}(r)$$

第一点是, 问题立马变成了一个一维动力学问题. 第二点是, 这个问题等价于一个粒子 m 在一维势能场 $V_{\text{eff}}(r)$ 中做运动的问题. 正是因为这个原因, 我们把 $V_{\text{eff}}(r)$ 称作有心力问题的**有效势能**(effective potential):

$$V_{\text{eff}}(r) = V(r) + \frac{mh^2}{2r^2}$$

利用这种方法,物理意义上最显著的一点就是可以方便我们考察定态的存在性和动力学稳定性.显然,在一维运动中,仅有当在 $r \in (0, +\infty)$ 下存在“山谷”,即一阶导数等于零,二阶导数大于零的点才有可能造成稳定的振动,它就是我们要寻找的束缚态.而这一点强烈依赖于势能的具体形式.考察引入的“离心势能”项的导数,我们发现:

$$V_{\text{cf}} = +\frac{mh^2}{2r^2} \Rightarrow V'_{\text{cf}} = -\frac{mh^2}{r^3}, V''_{\text{cf}} = +\frac{3mh^2}{r^4}$$

故这是一个倾向于往外推质点的力,而本身具有让质点稳定的因素.那么为了使得存在 $V'_{\text{eff}} = 0$ 和 $V''_{\text{eff}} > 0$ 的点,就要求 $V' > 0$ 必然成立,即必须是引力.且 V'' 虽然可以小于零但不能太小.

我们考察幂次的引力 $F = -kr^n$,那么它对应的势能为¹:

$$V = \frac{kr^{n+1}}{n+1} (n \neq -1), \quad V' = kr^n, \quad V'' = nkr^{n-1}$$

结合平衡条件可以发现存在平衡点:

$$V' + V'_{\text{cf}} = 0 \Rightarrow R = \left(\frac{mh^2}{k} \right)^{\frac{1}{n+3}}$$

那么在这个点处的稳定性就取决于二阶导数:

$$V''|_{r=R} = \frac{mh^2}{R^4} (3+n)$$

由于除非 $h = 0$,即纯径向运动,系数必然是大于零的.故分为两类情况:

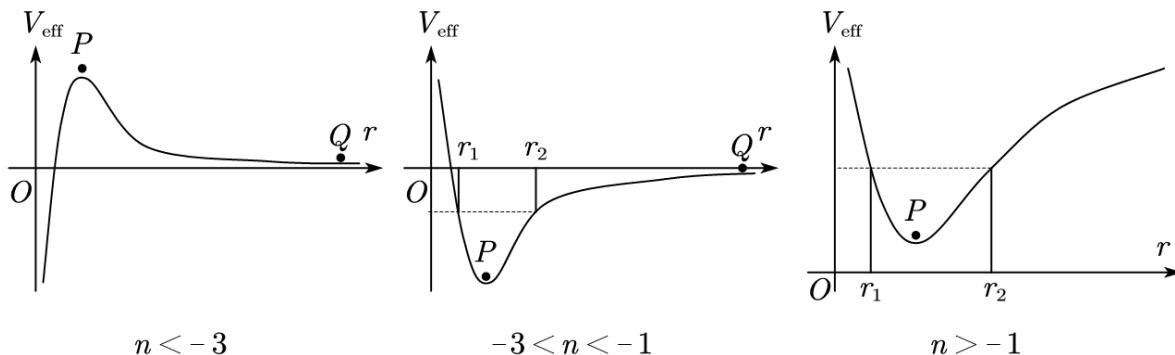


图 2.2: 三类幂律力

一是, $n \leq -3$, 即超过三次以上反比的力场,这样的力场不可能造成稳定的动力学运动.其势能函数图像如图2.2左.平衡是可以造成的,此时径向上始终保持 $r = R$,即停留在图中的 P 点.角向上 $\dot{\theta} = h/R^2$,其实就是在做匀速圆周运动.但是这个圆周运动动力学不稳定,一旦存在径向上微扰,要么轨道无限向力心坍缩,要么向无穷远转移而失去力心的束缚. $n = -3$ 的特殊情况值得引起关注.此时,如果 $h = \sqrt{k/m}$,那么势能恒等于零,对应大量不稳定的匀速圆周运动.而如果 $h > \sqrt{k/m}$ 或 $h < \sqrt{k/m}$,两者势能都是单调的不存在束缚态的可能性,前者会无限远离力心后者将无限接近力心.总之, $n < -3$ 都是不稳定的情况.

二是, $n > -3$ 下的所有力场都是稳定的.也就包括常见的平方反比例场.图2.2的中,右图都是这种情况. P 点具有稳定性.从而束缚态其实是指在 P 点左右的振动和对应的角向运动的合成.

对束缚态和散射态的求解能给出哪些共同的结论?首先让我们计算束缚态下 r_1 到 r_2 间振动的周期,也就是径向运动的动力学周期.这个问题和 $r(t)$ 的求解方法是共同的.只需要对之前的能量守恒式移项后积分:

$$\dot{r} = \sqrt{\frac{2[E - V(r)]}{m} - \frac{h^2}{r^2}}$$

¹ $n = -1$ 时, $V = k \ln r$

$$T = 2 \int_{r_1}^{r_2} \frac{dr}{\sqrt{\frac{2[E-V(r)]}{m} - \frac{h^2}{r^2}}}$$

其中 r_1 为过程中最靠近力心的点, 通称近心点(pericenter), 而 r_2 则为远心点(apocenter). 它们都是以上积分分母恰好为零的点, 使得该积分成为一个标准的瑕积分. 在 r_2 不是无穷大的情况下, 该积分总是有限的.

我们还可以计算在近心点和远心点运动过程中质点转过的角度. 这一点只需要找到角度的导数:

$$\dot{\theta} = \frac{h}{r^2}$$

并与之前的半径导数来做比:

$$\frac{d\theta}{dr} = \frac{h}{r^2 \sqrt{\frac{2[E-V(r)]}{m} - \frac{h^2}{r^2}}}$$

这就能得到角度:

$$\Delta\theta = \int_{r_1}^{r_2} \frac{h dr}{r^2 \sqrt{\frac{2[E-V(r)]}{m} - \frac{h^2}{r^2}}}$$

最后考虑散射问题的普遍结论. 首先, 幂律形式的力显然如果 $n \geq -1$, 这样的力对应的势能在无穷远处并不是 0. 将造成无穷远处具有一定初动能的物体来到近处时具有无穷大动能的发散问题. 故我们讨论的范围局限在 $n < -1$ 的幂律力, 引力斥力皆可. 十分类似地, 如果计算从无穷远来到无穷远去发生的位矢的转角, 为:

$$\Delta\theta = 2 \int_{r_1}^{\infty} \frac{h dr}{r^2 \sqrt{\frac{2[E-V(r)]}{m} - \frac{h^2}{r^2}}}$$

再进一步注意到由初始条件确定的的守恒量各为:

$$h = v_0 b \quad ; \quad E = \frac{1}{2}mv_0^2$$

代入化简, 我们要计算的偏转角 $\varphi = \pi - \Delta\theta$ 就成为了:

$$\varphi = \pi - 2 \int_{r_1}^{\infty} \frac{v_0 b dr}{r \sqrt{v_0^2(r^2 - b^2) - 2r^2 V/m}}$$

当 $n \geq -1$ 时偏转角的积分不会存在发散的问题. 永远是一个有限值.

2. 比耐方程法

另一种思路执着于求解二阶形式的方程, 但是通过变形转化为轨迹方程. 具体来说, 半径的一阶导数被我们化为:

$$\dot{r} = \frac{dr}{d\theta} \cdot \frac{d\theta}{dt} = \frac{h}{r^2} \cdot \frac{dr}{d\theta}$$

而二阶导数就能继续化为:

$$\ddot{r} = \frac{d}{d\theta}(\dot{r}) \cdot \frac{d\theta}{dt} = \frac{h}{r^2} \cdot \frac{d}{d\theta}\left(\frac{h}{r^2} \cdot \frac{dr}{d\theta}\right) = \frac{h^2}{r^4} \frac{d^2 r}{d\theta^2} - \frac{2h^2}{r^5} \left(\frac{dr}{d\theta}\right)^2$$

这样还不足以使得问题简化. 因为如果代入径向的牛顿定律, 得到的结果:

$$\frac{d^2 r}{d\theta^2} - r - \frac{2}{r} \left(\frac{dr}{d\theta}\right)^2 = \frac{r^4 F(r)}{mh^2}$$

还不见得是一个如何简单的方程. 比耐方程(Binet equation) 的不平凡之处在于它对半径进行了反演. 就是考虑一个倒数变换:

$$u = \frac{1}{r}$$

它可以把之前的运动轨迹的内侧翻到外侧去. 那么如果依然保持 θ 自变量不变, 轨迹的新导数为:

$$r = \frac{1}{u} , \quad \frac{dr}{d\theta} = -\frac{1}{u^2} \frac{du}{d\theta} , \quad \frac{d^2r}{d\theta^2} = -\frac{1}{u^2} \frac{d^2u}{d\theta^2} + \frac{2}{u^3} \left(\frac{du}{d\theta} \right)^2$$

代入之前的轨迹微分方程, 发现方程简单地变成了:

$$\frac{d^2u}{d\theta^2} + u = -\frac{F(\frac{1}{u})}{mh^2u^2}$$

这个简单的方程就可以帮我们快速解决一部分有心力问题的基础轨迹.

2.2 万有引力下天体运动

万有引力即以下**平方反比力**(inverse-square force):

$$F(r) = -\frac{GMm}{r^2}$$

任何平方反比力下的运动性质都是类似的. 我们以引力为正. 事实上经典物理喜欢用 $k^2 = GM$ 来指称中心力心产生力场的强度, 称作**高斯常数**(Gauss constant). 我们这里用 $k = GM$ 来表示引力的强度. 这样 $k < 0$ 就表示斥力. 而其它平方反比力, 比如库仑力也就归为同一种:

$$F(r) = \frac{Qq}{4\pi\varepsilon_0 r^2} = -\frac{km}{r^2} \Rightarrow k = -\frac{Qq}{4\pi\varepsilon_0 m}$$

于是, 在平方反比有心力下质点的运动问题, 就构成了经典力学历史上第一个用来解决的里程碑式的大问题, 它和它后来的各种变式, 统称为**开普勒问题**(Kepler problem).

为了以最快速度得到它的解, 我们直接使用比耐方程, 将力的函数用 u 表示并代入:

$$\frac{d^2u}{d\theta^2} + u = \frac{k}{h^2}$$

这就已经得到它的解了, 在合适的极坐标初始角度下写作:

$$u = \frac{k}{h^2}(1 + e \cos \theta)$$

其中 e 是根据初始条件确定的常数. 再倒过来便是:

$$r = \frac{h^2/k}{1 + e \cos \theta} = \frac{p}{1 + e \cos \theta}$$

不管 e, p 的正负和取值情况, 这都表示一条圆锥曲线方程. 可能是椭圆, 抛物线, 双曲线三者之一. 我们还因此得到了一个恒等式 $p = h^2/k$.

比耐方程固然简单. 但是它对我们解决一个天体运动的含初始条件的具体问题帮助不大. 为了得到一种行之有效的方法, 我们重新回到有效势能的方法. 现在我们代入势能函数:

$$V = -\frac{km}{r}$$

得到用 \dot{r} 和 $\dot{\theta}$ 做除法得到的微分方程:

$$\frac{d\theta}{dr} = \frac{h}{r^2 \sqrt{\frac{2E}{m} + \frac{2k}{r} - \frac{h^2}{r^2}}}$$

配方, 这便是:

$$\frac{d\left(\frac{k}{h^2} - \frac{1}{r}\right)}{\sqrt{\frac{2E}{mh^2} + \frac{k^2}{h^4} - \left(\frac{k}{h^2} - \frac{1}{r}\right)^2}} = d\theta$$

同理通过积分和合适的极坐标初始角:

$$r = \frac{\frac{h^2}{k}}{1 + \sqrt{1 + \frac{2h^2E}{k^2m}} \cos \theta} = \frac{p}{1 + e \cos \theta}$$

这个公式就完整透漏了其轨迹参量 p, e 与运动学积分不变量 h 和 E 的关系. 它们是:

$$p = \frac{h^2}{k}, \quad e = \sqrt{1 + \frac{2h^2E}{k^2m}}$$

通常这两个式子不喜欢这么些. 我们通过圆锥曲线长轴 A 的定义:

$$p = |1 - e^2|A$$

改写以上两个式子, 化简得到:

$$h = \sqrt{kp}, \quad E = \pm \frac{km}{2A}$$

这两个公式十分地重要, 它们直接联系了运动积分 h, E 和轨道参量 p, A . 通常, 通过初始条件就可以直接计算运动积分, 代入公式就可以完全确定轨道大体性质, 再根据几何关系就可以找到初始时刻在轨道上的具体位置. 最后就只需要对我们感兴趣的过程量或者状态量进行计算即可. 比如, 对于闭合的椭圆轨道, 其能量必然小于零. 那么上式中取负号以后我们可以计算一个周期需要的时间与轨道参量之间的关系:

$$T = \frac{\pi AB}{h/2} = \frac{2\pi AB}{\sqrt{kp}} = \frac{2\pi AB}{\sqrt{kB^2/A}} = 2\pi \sqrt{\frac{A^3}{k}}$$

2.3 二体与潮汐

第三章 刚体

3.1 刚体的物理描述

近代以前人们意识到了物质世界的连续性(continuum), 同时针锋相对地也提出了原子论(atomism). 原子最简单的模型就是质点, 而调和物质连续性与原子学说的中间模型就是刚体(rigid body) 模型. 刚体是不允许形变发生的系统. 由牛顿力学对质点的讨论推广到质点系的讨论, 使我们也很容易将相关结论进一步推广到刚体.

由于刚体上一点受力, 则整体同时运动起来, 这个模型与相对论力学体系是不兼容的. 具体来说, 相互作用必须以有限的速度传播, 否则就违背了因果律. 刚体不符合因果律这一时空的固有结构. 在很多相对论情境下将招致矛盾的结果.

刚体的物理学量是哪一些呢? 必要的内禀的属性是其质量的分布. 某一默认时刻 t_0 刚体占据了空间区域 Ω_0 , 由大量体积微元 dV (记做 $d^3\mathbf{R}_0$) 组成, 则刚体的总质量为:

$$m = \int_{\Omega_0} \rho(\mathbf{R}_0) d^3\mathbf{R}_0 = \int dm$$

随着刚体的运动, 原来在 \mathbf{R}_0 处的体积元现在在 t 时刻位于 \mathbf{r} 处, 整个刚体的运动由一个多元映射来定义:

$$f: \mathbf{R}_0, t \rightarrow \mathbf{r}(\mathbf{R}_0, t)$$

刚体的刚性的要求, 使得这个映射必须保持体积元的不变性:

$$f: d^3\mathbf{R}_0 \in \Omega_0 \rightarrow d^3\mathbf{r} \in \Omega; d^3\mathbf{R}_0 = d^3\mathbf{r} = dV$$

而且所有这个体积元内的所有内禀属性, 这里包括密度都不能变. 所以质量元 dm 也是不变的. 从而刚体具有不变的总质量. 马上就会发现, 质量几何(mass geometry) 对刚体的动力学来说也十分重要. 质量几何研究质量对特定原点 O 的各级矩(moment). 其中零级矩即为质量:

$${}^0M = m = \int_{\Omega} dm$$

一级矩是个矢量, 它定义了刚体的质心(center of mass) 的位置:

$${}^1M_i = mr_{Ci} = \int_{\Omega} r_i dm$$

$$\mathbf{r}_C = \frac{\int \mathbf{r} dm}{\int dm}$$

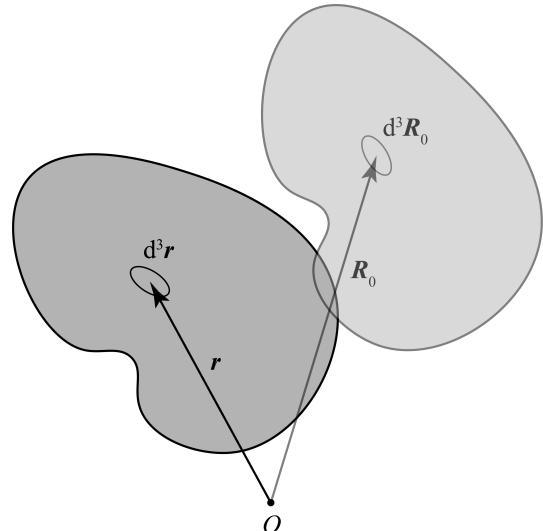


图 3.1: 刚体的描述

二级矩则是一个张量,它的九个分量代表**惯量积**(product of inertia):

$${}^2 M_{ij} = \int_{\Omega} r_i r_j dm$$

这些矩和原点的选取有关,随着刚体的运动也会不断变化,这三阶矩的信息对刚体动力学来说就是充分的了,通过后面的动力学可以发现,刚体的运动完全依赖于外力和这三阶矩.如果要研究广义相对论里的引力波辐射问题,更高阶的矩才变得重要起来.

刚体的运动可以被我们更精确地描述,在 t_0 时刻建立固定在刚体上,沿 x, y, z 三方向的单位矢量 e_1, e_2, e_3 ,那么刚体的运动同时也把三个矢量旋转到新的三个方向:

$$f: e_i \longrightarrow \epsilon_i$$

这三个矢量仍然要互相垂直,且长度为一.这在数学上导致了可以通过这三个矢量的导数定义**角速度**(angular velocity) 矢量的结果:

$$\dot{\epsilon}_i = \omega \times \epsilon_i$$

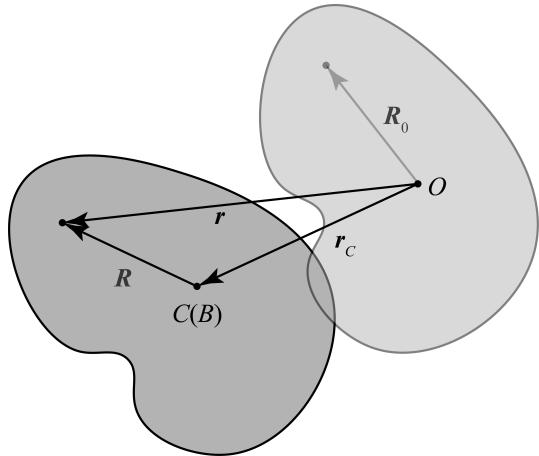


图 3.2: 基点法

十分类似于旋转参考系的变换的运动学,刚体的运动实际上就是有一个唯一的旋转参考系固连在刚体上,以后称作**刚体系**(reference system of rigid body),要研究的刚体上各个点速度实际上就是刚体系中定点的运动.于是习惯上我们采用**基点法**(method of base point) 来计算刚体上任意点的运动学量.定义**基点**(base point) 为 t_0 时刻 $\mathbf{R}_0 = \mathbf{0}$ 的点 B ,之后的位矢为 \mathbf{r}_B ,对应的基点速度加速度为 $\mathbf{v}_B, \mathbf{a}_B$,而刚提上待研究的点相对基点的位矢为 $\mathbf{r} - \mathbf{r}_B = \mathbf{R}$,那么该点的速度加速度即为:

$$\mathbf{v} = \mathbf{v}_B + \omega \times \mathbf{R}$$

$$\mathbf{a} = \mathbf{a}_B + \omega \times (\omega \times \mathbf{R}) + \dot{\omega} \times \mathbf{R}$$

出于力学的考虑,基点 B 一般就取做质心 C .作用在刚体上 \mathbf{r} 处的力 \mathbf{F} 固然对原点会有力矩 \mathbf{M} ,但是为了研究使刚体自身转动的效应,考虑到这个力同时也会使得质心运动起来,故我们重视这个力相对质心的力矩 \mathbf{M}_C :

$$\mathbf{M} = \mathbf{r} \times \mathbf{F}$$

$$\mathbf{M}_C = \mathbf{R} \times \mathbf{F}$$

刚体受到一个力系 \mathbf{F}_i 的作用,那么以下六个定理则来自于之前的动力学理论:

$$\begin{aligned} \sum_i \mathbf{F}_i &= \frac{dp}{dt} \quad ; \quad \sum_i \mathbf{F}_i = m \mathbf{a}_C \\ \sum_i \mathbf{v}_i \cdot \mathbf{F}_i &= \frac{d}{dt} \left(\frac{1}{2} m \mathbf{v}_C^2 + E_{kr} \right) \quad ; \quad \sum_i (\mathbf{v}_i - \mathbf{v}_C) \cdot \mathbf{F}_i = \frac{dE_{kr}}{dt} \\ \sum_i \mathbf{r}_i \times \mathbf{F}_i &= \frac{d}{dt} (\mathbf{r}_C \times m \mathbf{v}_C + \mathbf{L}_r) \quad ; \quad \sum_i \mathbf{R}_i \times \mathbf{F}_i = \frac{d\mathbf{L}_r}{dt} \end{aligned}$$

前两个式子实际上几乎没有区别,因为刚体的总动量其实就是质心动量 $\mathbf{p} = m \mathbf{v}_C$.后面几个式子则涉及到相对(平动)质心系的能量与角动量.它们与质量二级矩有着密切的联系,我们在第三讲3.3阐明.现在我们写出相对质心系能量角动量的定义:

$$E_{kr} = \int_{\Omega} \frac{1}{2} (\omega \times \mathbf{R})^2 dm$$

$$\mathbf{L}_r = \int_{\Omega} \mathbf{R} \times (\omega \times \mathbf{R}) dm$$

3.2 平面平行运动

一个底面磨平的物体贴在平坦的地面上运动给人以平面平行运动(plane-parallel motion)的概念. 其他的一些物体运动特征也相似于它: 黑板刷在黑板上的运动, 车轮在直行时的滚动... 它们的特征是: 刚体任何一个体积元都在一个特定的平面上运动, 而这些平面又彼此平行. 对于这样的运动基点法只需要给出质心 C 的位矢 \mathbf{r}_C 和对 t_0 时刻刚体转过的角度 θ 即可. 自由度就只有 3 个. 角速度与角加速度即可用角度的导数来给出 $\omega = \dot{\theta}$, $\beta = \ddot{\theta}$. 一般来说, 在垂直于这些平面方向要么由于动力学对称性不需要也不存在任何作用力. 要么这些力作为约束平面对物体的约束力而不在我考虑范围内. 所以我们对一些物理对象做如下修改:

- 所有参考点改成垂直平面过该点的参考轴. 并约定其正方向为垂直纸面向外. 这样的轴方向为 z 方向, 而运动平面的方向为 $x - y$ 平面.
- 与之对应地, 刚体绕轴的转动, 产生的角速度方向以逆时针为正. 这样角速度方向与约定的正方向一致. 以后在处理的时候, 往往被作为标量而处理.
- 关于转动惯量(rotational inertia): 现在就可以把质元到转动轴---不妨把它设为 $x = y = 0$ 的轴---的距离用来计算其二级矩, 对应的概念就叫做对这根轴的转动惯量:

$$I = \int \rho^2 dm = \int (x^2 + y^2) dm = M_{xx} + M_{yy}$$

正如上式所示, 它是 9 个惯量积中的特定 2 个的和.

- 轴的三种可能性: 一是空间中的定轴, 一是刚体上的固连轴, 还有可能是在空间中或是相对刚体都不静止的动轴. 后面就会发现, 定轴用角动量定律; 固连轴几乎都取过质心的轴, 有转动定律; 而动轴也有瞬时转动轴的特例, 可以证明有瞬时转动轴定律存在. 无论哪种情况, 转动惯量如上一点定义. 这样在定轴和瞬时转动轴情况下, 它有可能变成一个随时间变化的值 $I(t)$
- 动能的计算: 动能的定义不需要做任何修正. 而且概念本身也并不依赖与轴的选取. 只是, 其计算方法变得简单, 而且有两种算法:

- 如果是瞬时转动轴, 那么动能直接是:

$$T(t) = \frac{1}{2} I(t) \omega^2(t)$$

- 如果是定轴或过质心的轴, 事实上任何对定轴 O 计算刚体动能最终都必须由柯尼希定律转化为两项: 一项是质心 C 的轨道运动(orbital motion) 动能, 和相对质心的自旋运动(spin motion) 带来的动能. 这也是唯一正确的分解方式: 质心 C 换成其它相对刚体固定的点都是错的. 而刚体相对质心的转动惯量就不会是变化的了, 记作 I_C . 从而两个动能可以表示为:

$$T_C = \frac{1}{2} m v_C^2(t), \quad T_S = \frac{1}{2} I_C \omega^2(t), \quad T = T_C + T_S$$

- 角动量的计算: 角动量的定义就需要做修正了. 我们仅仅选取角动量的 z 分量, 这是因为要么刚体是薄的, 对平面上的点角动量本来就只有 z 分量, 要么刚体虽然有 $x - y$ 平面上的角动量但其动力学效果是产生一些约束它在 $x - y$ 平面上运动的我们不感兴趣的约束力. 所以角动量也和角速度一样可以被视作标量, 对任何点 P 的角动量 J 可以被定义为¹:

$$J_P = \mathbf{e}_z \cdot \int \boldsymbol{\rho}_P \times \mathbf{v} dm$$

式中, $\boldsymbol{\rho}_P$ 是从 P 轴垂直地指向每一个质元 dm 的矢径. 根据 P 点不同的选取方法, 有以下几种计算方法:

¹注意即使 P 在动, \mathbf{v} 也没有必要改为 $\mathbf{v} - \mathbf{v}_P$.

- 如果是瞬时转动轴,这种情况我们只在较少情况下使用.那么对它的角动量直接是:

$$J(t) = I(t)\omega(t)$$

- 如果是定轴,刚体角动量也必须由柯尼希定律转化为两项:一项是质心 C 的轨道角动量(orbital angular momentum) L , 和相对质心的自旋角动量(spin angular momentum) S . 这也是唯一正确的分解方式: 质心 C 换成其它相对刚体固定的点都是错的. 刚体对质心的转动惯量 I_C 是常数. 从而两者可以表示为:

$$L = \mathbf{e}_z \cdot [m\mathbf{r}_C(t) \times \mathbf{v}_C(t)] , \quad S = I_C\omega(t) , \quad J = L + S$$

- 其它动轴.除非变换参考系,否则一般极少被考虑.

平面平行运动的刚体的自由度为 3, 所以只需要 3 个方程就足以求解其运动,这三个方程通常取做质心运动定律和绕质心的转动定律(rotation law):

$$\sum_i F_{xi} = ma_{Cx}$$

$$\sum_i F_{xi} = ma_{Cx}$$

$$\sum_i M_i = I_C\beta$$

3.3 空间刚体运动 *

我们来定义空间刚体的转动惯量,首当其冲的一点是,沿不同方向的轴的转动惯量是不一样的. 那么我们改变一下定义. 把转动惯量视作从角速度到角动量的一个映射,为了简洁,把映射直接写成一个点乘,后面就知道其合理性:

$$\mathbf{I} : \omega \rightarrow \mathbf{L} = \mathbf{I} \cdot \omega$$

那么这个映射一定是线性的.这是因为如果角速度为 $\omega = \lambda_1\omega_1 + \lambda_2\omega_2$. 那么根据第一节给出的角动量计算方法:

$$\mathbf{L} = \int \mathbf{r} \times (\omega \times \mathbf{r}) dm$$

得到 $\mathbf{I} \cdot \omega = \lambda_1\mathbf{I}_1 + \lambda_2\mathbf{I}_2$

那么它就一定可以写作一个矩阵:

$$\mathbf{I} = \begin{bmatrix} I_x & I_{xy} & I_{xz} \\ I_{yx} & I_y & I_{yz} \\ I_{zx} & I_{zy} & I_z \end{bmatrix}$$

背后具有这样的原理:如果我们让角速度为 $(1, 0, 0)$, 把此时角动量计算出来写作 (I_x, I_{xy}, I_{xz}) . 同理如果角速度为 $(0, 1, 0)$ 时的角动量为 I_{yx}, I_y, I_{yz} , 角速度为 $(0, 0, 1)$ 时的角动量为 I_{zx}, I_{zy}, I_z . 那么任意角速度情况都可以视作以上三种情况的线性叠加:

$$(\omega_x, \omega_y, \omega_z) = \omega_x(1, 0, 0) + \omega_y(0, 1, 0) + \omega_z(0, 0, 1)$$

所以相应的角动量也是直接叠加的:

$$\mathbf{L} = \mathbf{I} \cdot \boldsymbol{\omega} : \begin{cases} L_x = I_x\omega_x + I_{xy}\omega_y + I_{xz}\omega_z \\ L_y = I_{yx}\omega_x + I_y\omega_y + I_{yz}\omega_z \\ L_z = I_{zx}\omega_x + I_{zy}\omega_y + I_z\omega_z \end{cases}$$

那么把上式写成一个矩阵的形式是再合适不过的了:

$$\begin{bmatrix} L_x \\ L_y \\ L_z \end{bmatrix} = \begin{bmatrix} I_x & I_{xy} & I_{xz} \\ I_{yx} & I_y & I_{yz} \\ I_{zx} & I_{zy} & I_z \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}$$

所以是不是说在这个意义下刚体的转动惯量 \mathbf{I} 就是一个矩阵? 这依然有不妥的地方. 因为如果换一个坐标系, 这个矩阵就会变化, 但是从角速度到角动量的映射是不会发生变化的. 我们有一种更严谨的定义方法, 是把转动惯量形式上写成如下的张量(tensor) 形式:

$$\mathbf{I} = I_x e_x e_x + I_y e_y e_y + I_z e_z e_z + I_{xy} e_x e_y + I_{xz} e_x e_z + I_{yx} e_y e_x + I_{yz} e_y e_z + I_{zx} e_z e_x = I_{zy} e_z e_y$$

其中 $e_i e_j$ 是一种新定义的并矢(dyadic) 运算的结果, 这种运算把两个矢量自由地并在一起, 满足按照分量的分配律, 但是不满足交换律, 即形成的并矢, 即一个张量的九个基, 对称的两个基是不被认为相等的:

$$e_i e_j \neq e_j e_i$$

张量因为代表了物理量之间的映射而具有了物理意义. 但是如果觉得这一层意义不够明显, 我们可以把张量代表的映射表示出来, 这一步是通过张量与矢量的点乘操作体现出来的. 事实上我们定义这个点乘一样地符合分配律, 而基之间的点乘符合:

$$(e_i e_j) \cdot e_k = e_i (e_j \cdot e_k)$$

即如果 $j = k$ 则就是 e_i , 如果不等于, 那就是零. 这就不难验证, 的确能把之前的矩阵式, 按照张量的方式理解为:

$$\mathbf{L} = \mathbf{I} \cdot \boldsymbol{\omega}$$

最后, 如果把之前的角动量定义式按照三重矢积公式, 按分量展开可以得到:

$$\mathbf{L} = \int \mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r}) dm = \int \left(\sum_i r_i^2 \sum_j \omega_j e_j - \sum_i r_i e_i \sum_j r_j \omega_j \right) dm$$

利用之前定义的二级矩:

$$\mathbf{L} = (M_{xx} + M_{yy} + M_{zz}) \sum_i \omega_i e_i - \sum_{ij} M_{ij} \omega_j e_i$$

写成分量形式为:

$$\begin{cases} L_x = (M_{yy} + M_{zz})\omega_x + M_{xy}\omega_y + M_{xz}\omega_z \\ L_y = M_{yx}\omega_x + (M_{xx} + M_{zz})\omega_y + M_{yz}\omega_z \\ L_z = M_{zx}\omega_x + M_{zy}\omega_y + (M_{xx} + M_{yy})\omega_z \end{cases}$$

对比之前的惯量张量, 我们发现了实际上惯量张量的非对角线分量其实就是对应的惯量积, 从而还具有对称性:

$$I_{ij} = I_{ji} = M_{ij}$$

而对角线分量却不是惯量积. 它是真正对这些轴的转动惯量, 如 x 轴:

$$I_x = M_{yy} + M_{zz}$$

数学上可以证明, 对任何点, 存在特殊的三条正交轴 $x - y - z$, 称为主轴(principal axes), 可以使得惯量积都为零, 只剩下三个主转动惯量:

$$\begin{bmatrix} L_x \\ L_y \\ L_z \end{bmatrix} = \begin{bmatrix} I_x & 0 & 0 \\ 0 & I_y & 0 \\ 0 & 0 & I_z \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}$$

即:

$$\mathbf{L} = (I_x\omega_x, I_y\omega_y, I_z\omega_z)$$

第四章 弹性体

4.1 弹性体的物理描述

所谓弹性体就是完全弹性(elasticity)的物体. 弹性描述的是使物体发生形变的力撤除以后物体可以回到静息状态的属性. 弹性力学研究的对象与范围就是弹性体的力学性质. 一般来说, 固体主要具有弹性而液体主要具有黏性, 若是研究中间的状态, 非牛顿流体(non-newtonian fluid) 和塑性固体(plastic solid), 那就是黏弹性力学(rheology) 要研究的对象了. 典型的黏弹性过程受力不是简单地正比于位移而是与速度, 与历史相关. 因此而可以发生永久的不可恢复的变形.

正因为如此, 完整描述弹性体的运动学时, 不得不额外留心所有点的实际位移. 在流体时也许速度更需要注意. 所以我们写出一个初始 $t = 0$ 位置矢量为 \mathbf{R} 的点, 经过 t 时间到达位置为 \mathbf{r} 处, 也就是我们要定义一个 $3D \times 1D$ 到 $3D$ 的映射:

$$\mathbf{r} = \mathbf{r}(\mathbf{R}, t)$$

不失普遍性地, 我们考虑如何刻画在 $\mathbf{R} = \mathbf{0}$ 的形变. 我们需要研究在 $\mathbf{R} = \mathbf{0}$ 的附近 $d\mathbf{R} = dX\mathbf{e}_x + dY\mathbf{e}_y + dZ\mathbf{e}_z$ 处的位移 $\boldsymbol{\delta} = \mathbf{r} - \mathbf{R}$ 与中心的位移去比较. 数学上有以下泰勒展式:

$$\boldsymbol{\delta}(d\mathbf{R}, t) = \boldsymbol{\delta}(\mathbf{0}, t) + d\mathbf{R} \cdot \nabla \boldsymbol{\delta}$$

上式中 $\nabla \boldsymbol{\delta}$ 是一个有九个分量的张量, 张量这一概念上一章介绍过, 它是九个分量的三行三列式的组合, 现在它的作用是可以与之前的矢量 $d\mathbf{R}$ 点乘把它线性地映射为另一个矢量:

$$\begin{aligned} \nabla \boldsymbol{\delta} &= \sum_{i,j} \frac{\partial \delta_j}{\partial X_i} \mathbf{e}_i \mathbf{e}_j \\ \nabla \boldsymbol{\delta} : \sum_i dX_i \mathbf{e}_i &\rightarrow \sum_j d\delta_j \mathbf{e}_j = \sum_j \left(\sum_i \frac{\partial \delta_j}{\partial X_i} dX_i \right) \mathbf{e}_j \\ \nabla \boldsymbol{\delta} : & \begin{bmatrix} \frac{\partial \delta_x}{\partial X} & \frac{\partial \delta_y}{\partial X} & \frac{\partial \delta_z}{\partial X} \\ \frac{\partial \delta_x}{\partial Y} & \frac{\partial \delta_y}{\partial Y} & \frac{\partial \delta_z}{\partial Y} \\ \frac{\partial \delta_x}{\partial Z} & \frac{\partial \delta_y}{\partial Z} & \frac{\partial \delta_z}{\partial Z} \end{bmatrix} \end{aligned}$$

不难发现第二个式子是不证自明的. 所以实际上刻画形变的包含于 $\nabla \boldsymbol{\delta}$ 这个张量. 但是并不是完全取决于它, 考虑像刚体这样的不能变形的物体, 上一章介绍过, 旋转依然是可能的. 不妨设刚体不仅随 $\mathbf{R} = \mathbf{0}$ 的点发生了 $\boldsymbol{\delta}(\mathbf{0}, t)$ 式的平动, 也要发生一个 $d\phi$ 的小角度转动. 在这里我们让转动的角度足够小以至于可以做小角近似. 这样就可以把刚体式的位移的以上张量写成:

$$\boldsymbol{\delta}(d\mathbf{R}, t) = \boldsymbol{\delta}(\mathbf{0}, t) + d\phi \times d\mathbf{R}$$

$$\nabla \boldsymbol{\delta} : \begin{bmatrix} 0 & d\phi_z & -d\phi_y \\ -d\phi_z & 0 & d\phi_x \\ d\phi_y & -d\phi_x & 0 \end{bmatrix}$$

这个矩阵是一个反对称矩阵,从而我们得出一个结论:一个固体在某点位移对应的 $\nabla\delta$ 如果是反对称的,则不产生任何形变,仅仅是局部整体发生了平移和旋转.

但是有一个简单的定理.任何一个方矩阵 $[M_{ij}]$ 都能被唯一地分解为对称矩阵和反对称矩阵.分别称作原来矩阵的对称部分(symmetric component) 和反对称部分(anti-symmetric component).用矩阵的转置可以很简单的得到这个结果:

$$[M_{ij}] = [S_{ij}] + [A_{ij}]$$

$$[S_{ij}] = \frac{[M_{ij}] + {}^t[M_{ij}]}{2}, \quad [A_{ij}] = \frac{[M_{ij}] - {}^t[M_{ij}]}{2}$$

那么问题就很简单了,之前那个矩阵的对称部分就是描述形变的部分.这个部分被叫做应变张量(strain tensor),以后用 ε 来表示¹:

$$\varepsilon = \frac{1}{2}(\nabla\delta + {}^t\nabla\delta)$$

$$\varepsilon : \begin{bmatrix} \frac{\partial\delta_x}{\partial X} & \frac{1}{2}\frac{\partial\delta_y}{\partial X} + \frac{1}{2}\frac{\partial\delta_x}{\partial Y} & \frac{1}{2}\frac{\partial\delta_z}{\partial Y} + \frac{1}{2}\frac{\partial\delta_y}{\partial Z} \\ \frac{1}{2}\frac{\partial\delta_y}{\partial X} + \frac{1}{2}\frac{\partial\delta_x}{\partial Y} & \frac{\partial\delta_y}{\partial Y} & \frac{1}{2}\frac{\partial\delta_x}{\partial Z} + \frac{1}{2}\frac{\partial\delta_z}{\partial X} \\ \frac{1}{2}\frac{\partial\delta_z}{\partial Y} + \frac{1}{2}\frac{\partial\delta_y}{\partial Z} & \frac{1}{2}\frac{\partial\delta_x}{\partial Z} + \frac{1}{2}\frac{\partial\delta_z}{\partial X} & \frac{\partial\delta_z}{\partial Z} \end{bmatrix} = \begin{bmatrix} \varepsilon_x & \theta_{xy}/2 & \theta_{xz}/2 \\ \theta_{xy}/2 & \varepsilon_y & \theta_{yz}/2 \\ \theta_{xz}/2 & \theta_{yz}/2 & \varepsilon_z \end{bmatrix}$$

其中三个对角元素 ε 称作正应变(normal strain).而非对角的元素 θ 称作剪应变(shear strain).如图4.1所示,在一种非常简单的模型中,大气中的棒被在沿棒方向施加一个拉力而伸长,但宽方向很自然地会产生些许的收缩.那么根据在棒里取出不同的微元形式,其变形方向也会有所改变.红色部分沿 x 方向就发生了明显的正应变.这是因为合适地平移,旋转其变形后的微元对齐形变前的微元(虚线)后,明显发现在 x 方向长度变大了.如果初始长度为 X ,之后在微元范围内端点的 $x = X + \delta$.于是根据上面的定义,正应变其实就是:

$$\varepsilon = \frac{\delta}{X} = \frac{x - X}{X}$$

而蓝色部分是个平行四边形,将底边与变形前的底边对齐以后,我们发现 $x - y$ 平面上顶角不再是 90° ,这其实就标志着剪应变.如果这个角度减小了 α ,那么之后的 $x = X + Y \tan \alpha$, $y = Y$.这就说明 $\delta_x = Y \tan \alpha$, $\delta_y = 0$.从而:

$$\theta_{xy} = \frac{\partial\delta_y}{\partial X} + \frac{\partial\delta_x}{\partial Y} = \tan \alpha \approx \alpha$$

可见这个角度变化其实就是剪应变,它与边的对齐方式无关,如果 x, y 轴夹角变小便是正的.正应变,剪应变都是无量纲的物理量.

接下来需要考虑弹性体的内部受力情况.令人惊讶的是,它也必然由一个对称张量描述.首先我们意识到弹性力内部的受力具有以下特征:A. 是空间点的函数,不同点处可以不同,但每一点应当有一个受力情况,它就是内部的应力(stress).B. 不是一个矢量.显然指出弹性体中的一点,并不能马上对应出来这个点处的某个受力的情况,因为我们考虑的是内力而不是外力,但是点处的受力不可能有明确的施力物体与受力物体.那么其实还需要在这一点处找到一个面矢量 $d\mathbf{S}$,才能说是谁对谁的力.我们这样就发现了,其实指定一个应力,本质上就是在每一点指定一个面矢量 $d\mathbf{S}$ 到相互作用力 $d\mathbf{F}$ 的映射.C. 显然,这样的映射应当为线性映射².

¹本书印刷体张量都是与矢量一致的粗体.手写时,为了区分,可以把张量写作带异型箭头的形式 $\overset{\leftrightarrow}{T}$ 或者直接用自由指标的分量代指构成的整体 T_{ij} .

²证明留给读者做思考,需要用到极限微元受力分析.

这样就几乎已经说明, 应力其实就是一个张量. 因为从一个矢量 $d\mathbf{S}$ 到另一个矢量 $d\mathbf{F}$ 的线性映射的数学模型其实就是张量, 它存在 $3 \times 3 = 9$ 个分量. 我们进一步指定, $d\mathbf{F}$ 的含义 $d\mathbf{S}$ 指向的那一侧的体元对这一侧的体元通过 $d\mathbf{S}$ 施加的相互作用力. 这个张量就是:

$$\left\{ \begin{array}{l} dF_x = \sigma_{xx}dS_x + \sigma_{xy}dS_y + \sigma_{xz}dS_z \\ dF_y = \sigma_{yx}dS_x + \sigma_{yy}dS_y + \sigma_{yz}dS_z \\ dF_z = \sigma_{zx}dS_x + \sigma_{zy}dS_y + \sigma_{zz}dS_z \end{array} \right.$$

$$\boldsymbol{\sigma} : d\mathbf{S} \rightarrow d\mathbf{F}$$

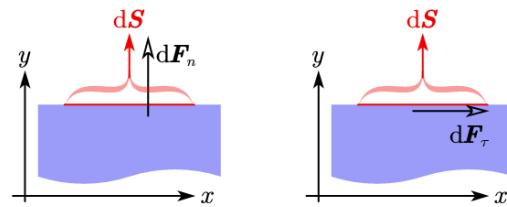


图 4.2: 正应力与剪应力

$$\boldsymbol{\sigma} : \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix}$$

这个张量就叫做**应力张量**(stress tensor). 通过极限微元的受力分析, 可以证明这个张量还必须是对称的³. 也就是说, 我们可以写成以下形式:

$$\boldsymbol{\sigma} : \begin{bmatrix} \sigma_x & \tau_{xy} & \tau_{zx} \\ \tau_{xy} & \sigma_y & \tau_{yz} \\ \tau_{zx} & \tau_{yz} & \sigma_z \end{bmatrix}$$

同样的, 参考图4.2, 我们可以发现, 对角线元素 σ 代表**正应力**(normal stress) 而非对角线元素 τ 代表**剪应变**(shear stress). 如果把面元取为 $d\mathbf{S} = dS_y e_y$, 考虑在平面上的受力就会产生两个分量:

$$d\mathbf{F}_n = \sigma_y dS_y e_y, \quad d\mathbf{F}_\tau = \tau_{xy} dS_y e_x$$

前者就是垂直于面的以拉力为正的正应力, 后者就是平行于面方向的剪应力. 而应力本身都是和以往学过的压强量纲一致, 国际单位是帕斯卡 Pa:

$$\sigma = \frac{dF_n}{dS}, \quad \tau = \frac{dF_\tau}{dS}$$

引入两个张量以后, 剩下的就是构造两者之间的因果关系: 应力是如何造成应变的. 在纯粹的弹性理论中, 我们可以假设应力张量到应变张量的映射再一次是一个逐点线性的映射. 这样的结果是数学上造成了每一点需要引入一个四阶的张量来描述这样的映射. 三维情况下四阶的张量一共会造成 $3^4 = 81$ 个独立分量. 由于应力应变张量本身具有对称性, 故其实只需要 6^2 个独立分量. 再由于体系的非耗散性的要求⁴, 其独立分量数最终减少到 21 个. 这也是高度非对称的介质的弹性系数中独立的量的个数. 然而, 如果介质是完全各向同性的, 也就是说沿所有三维空间中任意方向的长度与角度的拉伸与压缩都是相同困难的情况下, 张量的对称性理论可以告诉我们, 独立的弹性系数只会剩下两个. 具体来说, ϵ 与 σ 之间的关系必然会变成以下不依赖于坐标系选取的形式:

$$\boldsymbol{\sigma} = 2\mu \cdot \boldsymbol{\epsilon} + \lambda \text{Tr}(\boldsymbol{\epsilon}) \cdot \mathbf{I}$$

³也留给读者自己完成.

⁴导致了某种形式互易定理.

而按照这样的方式选取的弹性系数 λ 和 μ 被叫做**拉梅参数**(Lamé parameters). 式中 Tr 代表取迹操作, 而 I 是单位张量. 带入之前的两个张量的写法上式实际上意味着:

$$\sigma_x = 2\mu\varepsilon_x + \lambda(\varepsilon_x + \varepsilon_y + \varepsilon_z)$$

$$\sigma_y = 2\mu\varepsilon_y + \lambda(\varepsilon_x + \varepsilon_y + \varepsilon_z)$$

$$\sigma_z = 2\mu\varepsilon_z + \lambda(\varepsilon_x + \varepsilon_y + \varepsilon_z)$$

$$\tau_{ij} = \mu\theta_{ij}$$

至少我们能从上式发现, 剪应力和正应力是分别独立地导致剪应变和正应变的, 尤其是剪应力, 它在三个方向甚至都是互相独立地, 而这个比例系数就还被叫做**剪变模量**(shear modulus), 用 G 来表示. 即拉梅系数 $\mu \equiv G$. 这个在切向成立的结论称作**横向胡克定律**(transverse Hooke's law):

$$\frac{F_\tau}{S} = G \frac{\delta}{y}$$

而前三个式子对应的正应力正应变之间的关系比较复杂. 首先我们把三式相加可以得到:

$$\sigma_x + \sigma_y + \sigma_z = (2\mu + 3\lambda)(\varepsilon_x + \varepsilon_y + \varepsilon_z)$$

我们意识到, 任何物体在大气环境下实际上就受到周围分子不断撞击产生的大气压力而体积收缩, 实际上就是三个方向应力相等于压强 $\sigma_x = \sigma_y = \sigma_z = p$, 而且三个应变 ε_i 就等价于线压缩率, 应当是体压缩率的三分之一的情况, 这样我们得到:

$$p = \left(\lambda + \frac{2}{3}\mu \right) \frac{\Delta V}{V}$$

这个系数就叫做**体弹性模量**(bulk modulus):

$$K = \lambda + \frac{2}{3}\mu$$

一般来说, 弹性体预先在大体系中被压缩, 在这个基础上, 线性地叠加上由于其他形式的应力 σ 导致的新应变 ε .

还有一种至关重要的变形. 如果我们在一根弹性棒的 x 方向施加应力 σ , 但是 $y - z$ 方向不施加任何的力 $\sigma_y = \sigma_z = 0$, 扣除原来大气造成的应变, 通过解方程得到 $\varepsilon = \varepsilon_x$ 和 $\varepsilon' = \varepsilon_y = \varepsilon_z$ 的值:

$$\begin{cases} \sigma = 2\mu\varepsilon + \lambda(\varepsilon + 2\varepsilon') \\ 0 = 2\mu\varepsilon' + \lambda(\varepsilon + 2\varepsilon') \end{cases} \Rightarrow \begin{cases} \varepsilon = \frac{\lambda + \mu}{\mu(3\lambda + 2\mu)}\sigma \\ \varepsilon' = -\frac{\lambda}{2(\lambda + \mu)}\varepsilon \end{cases}$$

第一个式子翻过来写就是描述拉力与棒伸长之间的**纵向胡克定律**(longitudinal Hooke's law). 相关的弹性系数称作**杨氏模量**(Young's modulus), 即 $E = \mu(3\lambda + 2\mu)/(\lambda + \mu)$. 而第二个式子反映了如果棒在拉力方向伸长, 在垂直方向必然缩短的现实. 缩短率与伸长率的比值称为**泊松比**(Poisson's ratio), 即 $\nu = \lambda/2(\lambda + \mu)$:

$$\frac{F_n}{S} = E \frac{\delta}{x}, \quad \frac{\delta'}{y} = \nu \frac{\delta}{x}$$

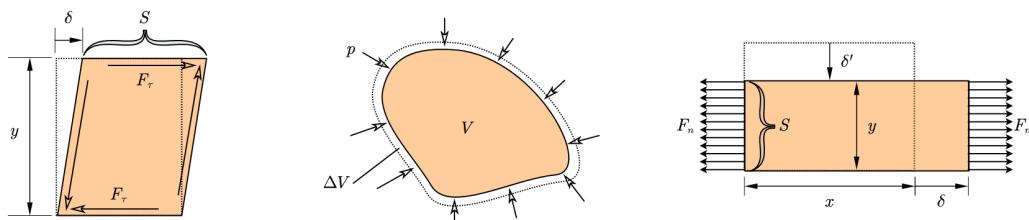


图 4.3: 剪变模量, 体弹性模量与杨氏模量

三个弹性模量的单位也是应力的单位, 即帕斯卡, 而典型材料的弹性模量在 10^{10}Pa 的数量级.

4.2 弹性棒, 弹性绳, 弹性膜与弹性体

在弹性的语境下,除了体状的大块弹性介质的研究,一些边缘化的状态:比如弹性体的表面,或者本来就在空间上受到限制的情况,包括本节介绍的弹性棒,绳,膜等,都是值得研究,且具有相似行为的体系.这种相似性体现在具有同一类的动力学方程和能量结构上.并最终导致了下一节介绍的弹性波的结果.

4.2.1 弹性棒

考虑一根弹性棒^{4.6},忽略其在垂直棒方向的运动带来的动力学效应.那么棒就被简化为一维的模型.在变量上,我们选取棒未形变时的坐标 x 和时间 t 为自变量,而沿正方向向前的位移 ξ 为因变量,即 $\xi(x, t)$.此时如 (a) 图,在棒上 x 处找一个固连在棒上的面元,随 x 改变的不仅仅是 ξ ,其实还有通过这一个面元相互拉扯的应力 σ .于是我们就来到了 (b) 中关于应力的计算.同时参考 (c) 图,一个体积元 dx 的左右侧原坐标 x 与 $x + dx$,但是它们的位移是不一样的,前者是 $\xi(x)$,后者则是 $\xi(x + dx) = \xi + d\xi$,即:

$$d\xi = \frac{\partial \xi}{\partial x} dx$$

那么这一段微元长度就变为 $dx + d\xi$.根据胡克定律,由于伸长产生的应力就是:

$$\sigma(x, t) = E \frac{d\xi}{dx} = E \frac{\partial \xi}{\partial x}$$

最后 (c) 图再取出 dx 的微元,其质量密度为 ρ ,截面积为 S ,那么其质量和两段受力的合力为:

$$dm = \rho S dx$$

$$dF = d\sigma S = ES \frac{\partial^2 \xi}{\partial x^2} dx$$

根据牛顿第二定律 $dF = dm \cdot a$, a 则是位移 u 对时间的二阶导数.这就给出了以下方程:

$$\frac{\partial^2 \xi}{\partial t^2} = \frac{E}{\rho} \frac{\partial^2 \xi}{\partial x^2}$$

这样的方程有一个名字,这也是我们第一次遇到这样的方程,称为**波动方程**(wave equation).一维波动方程的一般形式为:

$$\frac{\partial^2 \xi}{\partial t^2} = v^2 \frac{\partial^2 \xi}{\partial x^2}$$

其中 v 必须是一个常数数学上才叫波动方程, v 称为**波速**(wave velocity).而弹性棒对应的波速就是:

$$v = \sqrt{\frac{E}{\rho}}$$

波动方程和波速,正如其名,在下一节我们去解它之后就会发现就是产生的弹性波的特征.

下一个问题是考虑棒振动的能量.动能是比较简单的,我们把微元的动能做积分就是总动能:

$$T = \int_0^L \frac{1}{2} dm \left(\frac{\partial \xi}{\partial t} \right)^2$$

但我们尤其关心能量的**定域化**(localization) 问题.就是说能量是否可以表示为逐点的能量密度的体积分.就动能来看这也是成立的:

$$T = \int_0^L \mathcal{T} \cdot S dx \Rightarrow \mathcal{T} = \frac{1}{2} \rho \left(\frac{\partial \xi}{\partial t} \right)^2$$

现在我们来思考势能的定义. 显然棒不均匀伸长的情况相对于以往棒均匀伸长是两种完全不同的情况, 棒所储存的势能不仅要重新推导, 甚至需要在推导中确立其存在性. 事实上如果某一段 dx 内应力为拉力 $\sigma > 0$, 那么如果在这个力下这一段伸长 $d\xi(t+dt) - d\xi(t) = \frac{\partial}{\partial t} \left(\frac{\partial \xi}{\partial x} dx \right) dt$, 内部应力就会做负功, 我们只需要证明这个功对应的功率 (除以 dt) 等于对应体积内的势能的增加, 即:

$$ES \frac{\partial \xi}{\partial x} \cdot \frac{\partial^2 u}{\partial x \partial t} dx = \frac{d}{dt} (\mathcal{V} S dx)$$

这样就得到了:

$$\mathcal{V} = \frac{1}{2} E \left(\frac{\partial \xi}{\partial x} \right)^2, \quad V = \int_0^L \mathcal{V} \cdot S dx$$

总能量总是要守恒的, 对于棒的能量的计算, 现在就可以普遍地写为棒上**能量密度**(energy density) 的积分, 这个总能量密度就是动能密度 \mathcal{T} 和势能密度 \mathcal{V} 的和:

$$w = \frac{1}{2} \rho \left(\frac{\partial \xi}{\partial t} \right)^2 + \frac{1}{2} E \left(\frac{\partial \xi}{\partial x} \right)^2$$

4.2.2 弹性绳

考虑一条质量线密度为 λ , 预先被拉紧且拉力为 T 的细绳. 由于形变小所以沿绳方向的拉力几乎是个常数. 那么考虑这个绳在垂直 x 方向发生位移 $\xi(x, t), \eta(x, t)$. 我们立马就发现了这个问题不同于上一个问题之处在于因变量变多了. 上一个弹性棒中的运动模式是**纵波**(longitudinal wave). 而这里要建立的是垂直传播方向的**横波**(transverse wave) 模型. 横波不同于纵波的一大本质区别在于它有**偏振**(polarization). 我们把 ξ 和 η 两个方向叫做其偏振方向.

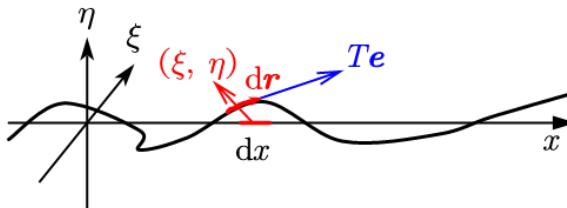


图 4.5: 弹性绳的建模

在建模上, 两者具有相似性. 同样是找到 dx 微元两端作用的力. 现在它大小几乎就是 T , 只是方向变成了 e , 它是绳端的单位切向量:

$$dr = (dx, d\xi, d\eta) \Rightarrow e = \frac{(dx, d\xi, d\eta)}{\sqrt{dx^2 + d\xi^2 + d\eta^2}}$$

由于在 ξ, η 方向的位移产生的斜率是小量, 上式近似为:

$$e = \left(1, \frac{d\xi}{dx}, \frac{d\eta}{dx} \right) = \left(1, \frac{\partial \xi}{\partial x}, \frac{\partial \eta}{\partial x} \right)$$

那么拉力 Te 的主要效果也就是在 $T_\xi = T \frac{\partial \xi}{\partial x}$ 和 $T_\eta = T \frac{\partial \eta}{\partial x}$ 两个方向上, 在微元段 dx 两侧的力的差产生了在 ξ, η 方向的加速度. 那么就得到了波动方程:

$$\begin{aligned} \frac{\partial^2 \xi}{\partial t^2} &= \frac{T}{\lambda} \frac{\partial^2 \xi}{\partial x^2} \\ \frac{\partial^2 \eta}{\partial t^2} &= \frac{T}{\lambda} \frac{\partial^2 \eta}{\partial x^2} \end{aligned}$$

可以发现, 波动方程的形式相对弹性棒没有本质的变化, 只不过从一个方程变成了两个独立量要满足的波动方程. 以及波速公式变成了:

$$v = \sqrt{\frac{T}{\lambda}}$$

能量的表现与推导方法也是类似的. 唯一的区别是现在的密度指的是线密度. 可以给出:

$$\mathcal{T} = \frac{1}{2}\lambda \left(\frac{\partial \xi}{\partial t} \right)^2 + \frac{1}{2}\lambda \left(\frac{\partial \eta}{\partial t} \right)^2$$

$$\mathcal{V} = \frac{1}{2}T \left(\frac{\partial \xi}{\partial x} \right)^2 + \frac{1}{2}T \left(\frac{\partial \eta}{\partial x} \right)^2$$

还有一点值得指出. 就是在这样的情况下势能还有一种极其简单的等效计算方法, 就是表面能(surface energy) 的计算方法, 体系的势能实际上就是简单地正比于变形之后的总绳长的:

$$V = T \int |\mathrm{d}\mathbf{r}| = T \int \sqrt{\mathrm{d}x^2 + \mathrm{d}\xi^2 + \mathrm{d}\eta^2} = T \int \sqrt{1 + \left(\frac{\partial \xi}{\partial x} \right)^2 + \left(\frac{\partial \eta}{\partial x} \right)^2} \mathrm{d}x$$

事实上, 对上式进行小量近似就可以发现:

$$V = T \int \left[1 + \frac{1}{2} \left(\frac{\partial \xi}{\partial x} \right)^2 + \frac{1}{2} \left(\frac{\partial \eta}{\partial x} \right)^2 \right] \mathrm{d}x = V_0 + \int \mathcal{V} \mathrm{d}x$$

从而不变拉力绳的势能实际上单纯地正比于绳子被拉长到了多长. 形成单位长度的绳子需要的能量是完全一致的.

4.2.3 弹性膜

弹性膜的情况则与弹性绳的情况互补: 这一次是自变量变成了二维的 x, y , 当然还有时间 t 一共三个. 倒是因变量只剩下垂直于面方向的位移 ξ . 即 $\xi(x, y, t)$. 面的惯性由质量面密度 μ 来描述. 而面的弹性, 类似于弹性绳, 我们也给面拉紧造成各向同性的内部张力. 这种拉紧是均匀的, 所以才造成均匀的面密度. 而同时也会对面上线元 $\mathrm{d}s$ 造成彼此之间垂直于线元的, 平行于面方向的拉力 $\mathrm{d}F = \sigma \mathrm{d}s$. 其中 σ 称作张力系数(tension coefficient). 那么如果在 $x - y$ 平面上选取右旋的闭合曲线圈, 中间围着一块膜. 在边界上一个典型的线段元和当点处的法向量就是:

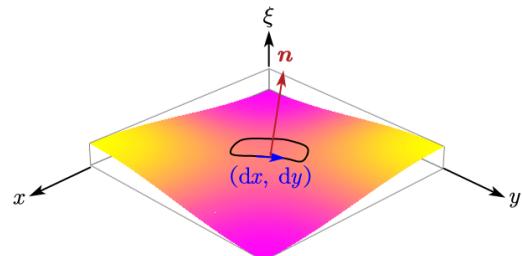


图 4.6: 弹性膜的建模

$$\mathrm{d}\mathbf{r} = (dx, dy, d\xi) = \left(dx, dy, \frac{\partial \xi}{\partial x} dx + \frac{\partial \xi}{\partial y} dy \right), \quad \mathbf{n} = \left(-\frac{\partial \xi}{\partial x}, -\frac{\partial \xi}{\partial y}, 1 \right)$$

那么通过这一段弧元, 周围膜对这一段膜施加的力, 忽略二阶小量就是:

$$\mathrm{d}\mathbf{F} = -\sigma \mathbf{n} \times \mathrm{d}\mathbf{r} = \sigma \left(dy, -dx, \frac{\partial \xi}{\partial x} dy - \frac{\partial \xi}{\partial y} dx \right)$$

足以发现, 如果考虑曲线中的膜受到的合力, 显然 $x - y$ 方向是平衡的. 合力在 ξ 方向, 大小可以根据数学上的格林公式⁵得到:

$$F_\xi = \sigma \oint \left(\frac{\partial \xi}{\partial x} dy - \frac{\partial \xi}{\partial y} dx \right) = \iint \sigma \left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} \right) \mathrm{d}x \mathrm{d}y$$

从中我们可以发现, 单位面积上给出的内力就是:

$$f = \frac{\mathrm{d}F_\xi}{\mathrm{d}x \mathrm{d}y} = \sigma \left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} \right)$$

⁵ 格林公式(Green formula) 为, 对任何平面区域 D 与其边界右旋闭合曲线 ∂D :

$$\oint_{\partial D} P(x, y) \mathrm{d}x + Q(x, y) \mathrm{d}y = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathrm{d}x \mathrm{d}y$$

那么根据牛顿第二定律, 这就是让膜产生 ξ 方向加速度的力. 化简就得到了膜对应的二维波动方程:

$$\frac{\partial^2 \xi}{\partial t^2} = \frac{\sigma}{\mu} \left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} \right)$$

对应波速就是:

$$v = \sqrt{\frac{\sigma}{\mu}}$$

同理, 对应动能和势能的面密度为:

$$\begin{aligned} \mathcal{T} &= \frac{1}{2} \mu \left(\frac{\partial \xi}{\partial t} \right)^2 \\ \mathcal{V} &= \frac{1}{2} \sigma \left(\frac{\partial \xi}{\partial x} \right)^2 + \frac{1}{2} \sigma \left(\frac{\partial \xi}{\partial y} \right)^2 \end{aligned}$$

同样的, 势能可以重新表述为表面能:

$$V = \sigma \int |dS| V_0 + \int \mathcal{V} dx dy$$

4.2.4 弹性体 *

终于我们可以讨论自变量有 $x - y - z$ 三维空间和 t 一维时间, 因变量, 即弹性体的位移矢量 δ 也有三个维度的最复杂的情况. 本章第一节的讨论总结下来就是两个式子:

$$\boldsymbol{\varepsilon} = \frac{1}{2} (\nabla \boldsymbol{\delta} + {}^t \nabla \boldsymbol{\delta})$$

$$\boldsymbol{\sigma} = 2\mu \cdot \boldsymbol{\varepsilon} + \lambda \text{Tr}(\boldsymbol{\varepsilon}) \cdot \mathbf{I}$$

它给出了位移的一阶倒数是如何产生力的. 另外一部分物理公式, 就是计算单位体积受到的合力, 最后联立牛顿第二定律, 我们就有希望给出波动方程.

计算合力需要对一个体积区域 D 的边界闭合外向曲面 ∂D 上受到的外力进行积分:

$$\mathbf{F} = \oint_{\partial D} dS \cdot \boldsymbol{\sigma}$$

这其实可以使用数学上广义上的高斯定理化为体积分:

$$\mathbf{F} = \iiint_D \nabla \cdot \boldsymbol{\sigma} dx dy dz \Rightarrow \mathbf{f} = \frac{d\mathbf{F}}{dx dy dz} = \nabla \cdot \boldsymbol{\sigma}$$

最后再联立牛顿第二定律, 我们就化简出来第三个方程:

$$\frac{\partial^2 \boldsymbol{\delta}}{\partial t^2} = \frac{1}{\rho} \nabla \cdot \boldsymbol{\sigma}$$

可以得到, 如果把 $\boldsymbol{\delta}$ 记做 (ξ, η, ζ) , 那么对应的方程就称作**纳维-柯西方程**(Navier-Cauchy equation). 完全展开的形式为:

$$\begin{aligned} \frac{\partial^2 \xi}{\partial t^2} &= \frac{\lambda + 2\mu}{\rho} \frac{\partial^2 \xi}{\partial x^2} + \frac{\lambda + \mu}{\rho} \left(\frac{\partial^2 \eta}{\partial x \partial y} + \frac{\partial^2 \zeta}{\partial x \partial z} \right) \\ \frac{\partial^2 \eta}{\partial t^2} &= \frac{\lambda + 2\mu}{\rho} \frac{\partial^2 \xi}{\partial y^2} + \frac{\lambda + \mu}{\rho} \left(\frac{\partial^2 \zeta}{\partial y \partial z} + \frac{\partial^2 \xi}{\partial y \partial x} \right) \\ \frac{\partial^2 \zeta}{\partial t^2} &= \frac{\lambda + 2\mu}{\rho} \frac{\partial^2 \xi}{\partial z^2} + \frac{\lambda + \mu}{\rho} \left(\frac{\partial^2 \xi}{\partial z \partial x} + \frac{\partial^2 \eta}{\partial z \partial y} \right) \end{aligned}$$

这样写是不利于我们分析其求解的. 通过数学化简, 我们暂时先把波动方程写成如下形式:

$$\rho \frac{\partial^2 \boldsymbol{\delta}}{\partial t^2} = (\lambda + 2\mu) \nabla (\nabla \cdot \boldsymbol{\delta}) - \mu \nabla \times (\nabla \times \boldsymbol{\delta})$$

其中需要用到矢量微分恒等式:

$$\nabla(\nabla \cdot \boldsymbol{\delta}) - \nabla \times (\nabla \times \boldsymbol{\delta}) = \nabla^2 \boldsymbol{\delta}$$

最后, 棒中的能量守恒, 动能密度是简单的. 而势能密度在只有正应变或剪应变情况下容易证明, 它就是应力和产生的应变乘积的一半. 普遍情况下只需要做叠加即可:

$$\mathcal{T} = \frac{1}{2}\rho \left(\frac{\partial \boldsymbol{\delta}}{\partial t} \right)^2 = \frac{1}{2}\rho \left(\frac{\partial \xi}{\partial t} \right)^2 + \frac{1}{2}\rho \left(\frac{\partial \eta}{\partial t} \right)^2 + \frac{1}{2}\rho \left(\frac{\partial \zeta}{\partial t} \right)^2$$

$$\mathcal{V} = \frac{1}{2}(\sigma_x \varepsilon_x + \sigma_y \varepsilon_y + \sigma_z \varepsilon_z + \tau_{xy} \theta_{xy} + \tau_{yz} \theta_{yz} + \tau_{zx} \theta_{zx}) = \frac{1}{2}\lambda(\nabla \cdot \boldsymbol{\delta})^2 + \frac{1}{2}\mu(\nabla \boldsymbol{\delta} : \nabla \boldsymbol{\delta} + {}^t \nabla \boldsymbol{\delta} : \nabla \boldsymbol{\delta})$$

式中 “:” 符号表示张量之间的双点乘运算. 具体来说:

$$\left(\sum_{ij} T_{ij} e_i e_j \right) : \left(\sum_{ij} S_{ij} e_i e_j \right) = \text{Tr}(\boldsymbol{T} \cdot \boldsymbol{S}) = \sum_{ij} T_{ij} S_{ji}$$

4.3 弹性波

波动方程的求解问题其实是非常简单的. 它其实是普遍的偏微分方程(partial differential equation, PDE)求解问题中最简单的一类. 不同于常微分方程, 一个多元函数, 如 $\xi(x, t)$, 在自变量取某值时不仅仅可以求函数值和对 x 的各阶导数, 还可以求对 t 的各阶导数, 甚至还可以求 x 和 t 的各阶混合导数. 对这些值提出限制, 就构成了偏微分方程:

$$F \left(x, t, \frac{\partial^{m+n} \xi}{\partial x^m \partial t^n} \right) = F \left(x, t, \xi, \frac{\partial \xi}{\partial x}, \frac{\partial \xi}{\partial t}, \frac{\partial^2 \xi}{\partial x^2}, \frac{\partial^2 \xi}{\partial x \partial t}, \dots \right) = 0$$

与常微分方程类似的一点是, 它一般也存在着大量的解. 尤其是线性偏微分方程的特殊情况, 其解也是满足常微分方程的叠加原理, 通特解原理的:

$$\sum_{i,j} c_{ij}(x, t) \frac{\partial^{m+n} \xi}{\partial x^m \partial t^n} = f(x, t)$$

其中最简单的波动方程其实就是二阶的齐次的常系数的线性偏微分方程:

$$f(x, t) = 0, c_{00} = c_{01} = c_{10} = 0, c_{20} = -v^2, c_{11} = 0, c_{02} = 1, c_{ij} = 0 \ (i > 2 \text{ or } j > 2)$$

$$\Rightarrow \frac{\partial^2 \xi}{\partial t^2} = v^2 \frac{\partial^2 \xi}{\partial x^2}$$

如何找到以上方程的所有解(通解)? 我们介绍两种常见的思路:

4.3.1 分离变量法

猜测原方程具有如下形式的特解:

$$\xi(x, t) = A(x)u(t)$$

代入原方程, 两边同时除以 $A(x)u(t)$, 化为纯粹的分离变量的常微分方程形式:

$$\frac{\frac{d^2 u}{dt^2}}{u} = v^2 \frac{\frac{d^2 A}{dx^2}}{A}$$

左侧必然是不含 x 的那一类函数. 而右侧又必须是不含 t 的那一类函数. 两者还是同一个函数. 所以这个函数 x, t 都不含有, 那便是常数. 最后再注意到随时间变化的部分在简谐函数下恰好能符合常数的要求:

$$u(t) = e^{i\omega t} \Rightarrow \frac{\frac{d^2 u}{dt^2}}{u} = -\omega^2$$

最后求解对 A 的限制, 得到:

$$A(x) = A e^{\mp i k x} \quad , \quad \frac{\omega}{k} = v$$

从而我们就得到了通解的一组基:

$$\left\{ \xi(x, t) = A e^{i(\omega t - kx)}, \xi(x, t) = A e^{i(\omega t + kx)} \middle| \frac{\omega}{k} = v \right\}$$

根据傅里叶理论. 上述解的叠加其实就是所有可能在一维下符合波动方程的解. 这样一组基本的解就是右行波和左行波解. 注意到这样的波有一点不符合实际情况: 它完全没有色散, 所有的频率的波传播速度都完全一样, 说明我们的建模过于简单.

4.3.2 变量代换法

如果我们尝试变量代换:

$$x_+ = x + vt, x_- = x - vt$$

那么容易发现, 原来的方程居然变成了:

$$\frac{\partial^2 \xi}{\partial x_+ \partial x_-} = 0$$

这样, 只需要 ξ 仅仅是其中一个变量的函数, 对另一个函数的偏导数就是零, 就能符合以上的方程. 即任意的 $f(x_+)$ 和 $g(x_-)$ 都是以上方程的解. 由于叠加原理, 我们就能肯定地找到以下特解:

$$\xi(x, t) = g(x - vt) + f(x + vt)$$

这其实也构成了通解, 它甚至直接就是任意通解的形式, 不仅是可以叠加形成通解的一组基:

$$\{\xi(x, t) = g(x - vt) + f(x + vt) | \forall f \forall g\}$$

其中任意的 $g(x - vt)$ 就代表初始时刻波形 $g(x)$ 任意的一列右行波. 而 f 就是左行波. 即任何一维的波动总可以分解为任意波形的左行波和右行波的叠加. 上一节的任意频率的各种简谐波又可以叠加形成任意波形, 实际上就是傅里叶理论的重要内容.

4.3.3 多维情况

如果是类似于弹性绳那样的波动方程:

$$\frac{\partial^2 \xi}{\partial t^2} = v^2 \frac{\partial^2 \xi}{\partial x^2} \quad , \quad \frac{\partial^2 \eta}{\partial t^2} = v^2 \frac{\partial^2 \eta}{\partial x^2}$$

那么这个一维传播的二维振动的基础行波解就可以表达为:

$$(\xi, \eta) = (A, B) e^{i(\omega t \pm kx)} \quad , \quad \frac{\omega}{k} = v$$

其中偏振矢量 (A, B) 可以是任何可能性, 如果是 $(0, 1)$ 就是 ξ 方向的线偏振, 如果是 $(1, 0)$ 就是 η 方向的线偏振. 故这个振动还可以按偏振模式分解为两个. 注意到 A, B 都是复数还可以带有相位. 如果两个方向复振幅的模相等而相位差 90° , 那么这个偏振模式就叫做圆偏振. 即, 相对简单一维振动, 它还多了偏振这一个自由度.

如果是类似于弹性膜那样的波动方程:

$$\frac{\partial^2 \xi}{\partial t^2} = v^2 \left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} \right)$$

那么这个二维传播的一维振动的基础行波解就被表达为:

$$\xi = A e^{i(\omega t - k_x x - k_y y)} , \quad \frac{\omega}{\sqrt{k_x^2 + k_y^2}} = v$$

那么这个波就同时具有了以下自由度: 频率 (决定 ω 和 k 以及引申的周期, 波长等等), 传播方向 (决定 k_x 和 k_y 的比例和正负), 强度 (决定 A 的模), 以及相位 (决定 A 的幅角).

最后考虑弹性体中的波动方程, 即纳维-柯西方程:

$$\rho \frac{\partial^2 \boldsymbol{\delta}}{\partial t^2} = (\lambda + 2\mu) \nabla (\nabla \cdot \boldsymbol{\delta}) - \mu \nabla \times (\nabla \times \boldsymbol{\delta})$$

这个方程看上去很难直接求解. 但是实际上数学上有一重要定理为我们提供了重要的线索. 就是斯托克斯-亥姆霍兹定理 (Stokes-Helmholtz's theorem) 发现, 任何三维空间中的矢量 $\boldsymbol{\delta}(\mathbf{r})$ 都可以分解为一个有旋无散场 $\boldsymbol{\delta}_S$ 与有散无旋场 $\boldsymbol{\delta}_P$ 的叠加:

$$\boldsymbol{\delta} = \boldsymbol{\delta}_S + \boldsymbol{\delta}_P \quad : \quad \nabla \cdot \boldsymbol{\delta}_S = 0, \quad \nabla \times \boldsymbol{\delta}_P = \mathbf{0}$$

基于这一点, 我们可以考虑纳维-柯西方程在仅有有旋无散的振动和有散无旋的振动情况下的各自的解, 只要存在, 叠加在一起就是通解了. 再次结合式子:

$$\nabla(\nabla \cdot \boldsymbol{\delta}) - \nabla \times (\nabla \times \boldsymbol{\delta}) = \nabla^2 \boldsymbol{\delta}$$

首先考虑有旋无散的振动. 那么纳维-柯西方程简化为:

$$\rho \frac{\partial^2 \boldsymbol{\delta}_S}{\partial t^2} = -\mu \nabla \times (\nabla \times \boldsymbol{\delta}_S)$$

但是旋即借助前式发现:

$$-\nabla \times (\nabla \times \boldsymbol{\delta}_S) = \nabla^2 \boldsymbol{\delta}_S$$

这就相当于给出了:

$$\frac{\partial^2 \boldsymbol{\delta}_S}{\partial t^2} = v_S^2 \nabla^2 \boldsymbol{\delta}_S = v_S^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \boldsymbol{\delta}_S , \quad v_S = \sqrt{\frac{\mu}{\rho}}$$

这样就可以求解了, 其通解的基为:

$$\boldsymbol{\delta}_S = \mathbf{A}_S e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})} , \quad \frac{\omega}{|\mathbf{k}|} = v_S$$

最后验证它是否符合散度为零的条件, 发现必须要满足:

$$\mathbf{k} \cdot \mathbf{A}_S = 0 \Rightarrow \mathbf{k} \perp \mathbf{A}_S$$

即这样的波其实是一种横波. 它构成的要素有频率, 传播方向, 强度, 偏振, 以及相位. 这样的弹性体波称作剪波 (shear wave). 它发生的形变就是剪应变, 其波速恰好用的就是剪切模量 $\mu = G$:

$$v_S = \sqrt{\frac{G}{\rho}}$$

再考虑有散无旋的振动. 那么纳维-柯西方程简化为:

$$\rho \frac{\partial^2 \boldsymbol{\delta}_P}{\partial t^2} = (\lambda + 2\mu) \nabla (\nabla \cdot \boldsymbol{\delta}_P)$$

但是旋即借助前式发现:

$$\nabla(\nabla \cdot \boldsymbol{\delta}_P) = \nabla^2 \boldsymbol{\delta}_P$$

这就相当于给出了：

$$\frac{\partial^2 \boldsymbol{\delta}_P}{\partial t^2} = v_P^2 \nabla^2 \boldsymbol{\delta}_P = v_P^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \boldsymbol{\delta}_P , \quad v_P = \sqrt{\frac{\lambda + 2\mu}{\rho}}$$

这样就可以求解了，其通解的基为：

$$\boldsymbol{\delta}_P = \mathbf{A}_P e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})} , \quad \frac{\omega}{|\mathbf{k}|} = v_P$$

最后验证它是否符合旋度为零的条件，发现必须要满足：

$$\mathbf{k} \times \mathbf{A}_S = 0 \Rightarrow \mathbf{k} // \mathbf{A}_S$$

即这样的波其实是一种纵波。它构成的要素有频率、传播方向、强度，以及相位。这样的弹性体波称作压波(pressure wave)。它发生的形变就是正应变，注意其波速用的既不是弹性棒情况的杨氏模量 E ，也不是体弹性模量 K ，与后者关系更近，实际上由第一节的知识：

$$v_P = \sqrt{\frac{K + 4G/3}{\rho}}$$

最终我们发现，作为纵波的压波波速 v_P 总是要比作为横波的剪波 v_S 快一些。故在地震波场合下，总是把前者叫做首波(primary wave)，后者叫做次波(secondary wave)。两种命名方法都是导致纵波简记为 P 波，横波简记为 S 波的原因。也经常地，我们会把特定方向传播，有特定频率(横纵波速不同会导致 k 不同)的横波两种偏振，加上纵波称作三种不同的偏振，波从振源传播到特定点，三种偏振需要的时间是不同的。纵波总是先到，横波后到。

地球内部结构虽然目前只能从现象去反过来建立模型，但比较令人相信的一点是内部存在分层的现象，在地表以下约 2891km 处存在从固态的地幔到高压“液态”的地核的古登堡不连续性(Gutenberg discontinuity)。作为类似于固态的高压液态，在建立模型上十分重要的一点是，其剪变模量 G 几乎降低到了 0。这给出了两个结果：一是通过剪变的弹性得以传播和存在的剪波无法存在，故地震波在地幔-地核分界面上会发生 S 波的全反射⁶。二是压波就彻底变成了由体弹性模量决定的波速公式，这一个公式也是正常气体、液体中的体波的波速公式：

$$v_P = \sqrt{\frac{K}{\rho}}$$

尤其是在理想气体中，比如空气中传播的声波，其波速就可以把气体视作理想气体，而气团的压缩与膨胀都被视作绝热的(因为振动频率高气体完全无法达到热学平衡)，于是：

$$\begin{aligned} pV^\gamma &= C \Rightarrow \frac{dp}{p} + \gamma \frac{dV}{V} = 0 \\ \Rightarrow K &= -\frac{dp}{dV/V} = \gamma p \end{aligned}$$

带入声速，结合密度公式 $\rho = \mu p / RT$ ，得到气体中的声速公式：

$$v = \sqrt{\frac{\gamma RT}{\mu}}$$

这个计算早在 1687 年就由牛顿完成，但是受限于当时几乎等于空白的热学、力学理论的限制与同样几乎是空白的实验测量数据，牛顿给出的公式中没有出现绝热指数 γ ，导致它计算出的室温下空气中声速只有 300m/s。比正确值 340m/s 小了不少。这个误差后来由拉普拉斯解释与修正。

4.3.4 再论格波

⁶少量 S 波有可能转化为 P 波继续向地核传播

第五章 流体

5.1 流体的物理描述

我们暂时只研究流体的动力学特征. 规避掉任何有关热学的问题.

首先考察流体的运动学. 流体已经不再具有恢复任何形变的能力, 故引入流体相对初始时刻的位移已经不是最合适的方法了. 如果依然这么处理, 把流体质元的位置写为初始位置和时间的函数, 这个描述方法称作拉格朗日法(Lagrange's method). 即:

$$\mathbf{r} = \mathbf{f}(\mathbf{R}, t)$$

这样我们自然地可以得到一条曲线和它的方程, 便是选定某个 $\mathbf{R} = \mathbf{R}_0$, 于是位置 \mathbf{r} 单独成为某个时间的函数. 这代表某个质元的实际运动过程和轨迹. 这个轨迹称作迹线(path line).

但是, 跟常见的做法是, 我们不再关心每一个时刻各个空间位置的质元的初始位置在哪儿, 这没有太大的意义. 值得关心的当下流体在做何种流动. 即每一个点处的流速 $\mathbf{v}(\mathbf{r}, t)$ 是如何的. 其实就是上面函数的对时间偏导数:

$$\mathbf{v}(\mathbf{r}, t) = \frac{\partial \mathbf{f}}{\partial t} \Big|_{\mathbf{R}: \mathbf{f}(\mathbf{R}, t) = \mathbf{r}}$$

用流速场 $\mathbf{v}(\mathbf{r}, t)$ 来描述流体的方法称作欧拉法(Euler's method). 而欧拉法的抽象函数也可以可视化: 我们找到这样的空间曲线, 数学上称作流速场的积分曲线: 它在每一点的切线方向与某一时刻 t_0 下该点流速相同, 或者说它是以下微分方程的解:

$$\frac{dx}{v_x|_{t=t_0}} = \frac{dy}{v_y|_{t=t_0}} = \frac{dz}{v_z|_{t=t_0}}$$

这样的曲线就叫流线(streamline), 一组流线还可以形成流管(streamtube) 的概念. 流线与迹线显然不一定重

合, 因为流线只需要给出某一时刻的流速场就可以画出来, 但迹线显然是要追踪不同时刻一个质元的运动, 如果下一时刻的流速场变了, 流体质元就会沿另一套流线去运动了. 但即使迹线不再直观, 使用欧拉法的优越性与完备性后面就会逐渐发现.

讨论流体的运动时有两个常见的特殊情况与条件可以使问题变得更简单:

一是, 常见的液体, 常常被建模为不可压缩流体(incompressible fluid). 它有两种等价的表述方式. 一方面由于体积不可被压缩, 质量又守恒. 其密度便是个常数:

$$\rho = \text{Const.}$$

又或者关注流速场的性质就会发现, 这个场只能是无散的, 否则就会因为流体流向局部体积元造成内部质量增加:

$$\nabla \cdot \mathbf{v} = 0$$

正因为这一点, 液体作为流体的计算更简单, 其结果也更简单: 动力学的流动与热学往往是解耦合的, 流体的温度分布在这个模型性不会影响到流体的运动, 反过来流体的流动却必然对传热有影响. 比如搅动冷

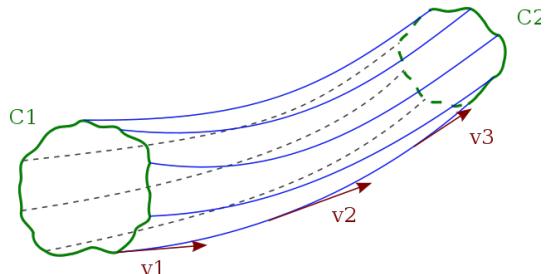


图 5.1: 流线与流管

热不均匀的水,水的运动几乎由搅动的方式决定而与温度分布有关,但水的温度变化显然受到搅动的强烈影响.但是在非常常见的对流传热(heat transfer by convection)问题中,温度变化导致的密度变化产生的动力学效果是不可或缺的一环,那种场合就绝对不能把流体视作不可压缩流体.典型的情况,液体比如水锅烧水时的强对流情况,探究洋流的产生等等,气体则几乎都需要视作可压缩的流体.

可压缩的流体也存在质量的连续性方程(continuity equation).我们可以把可变的密度 ρ 和流速 v 相乘作为流密度矢量 j ,那么:

$$\nabla \cdot (\rho v) + \frac{\partial \rho}{\partial t} = 0$$

再注意到, $v \cdot \nabla$ 这个算符其实就是在跟随运动的流体一起运动,但是时间不变化,求物理量导数的算符,称为随体导数(material derivative).它与在局部求时间的偏导算符 $\partial/\partial t$,称作局部导数(local derivative)之和就是跟随流体质元一起运动,还考虑时间的流逝情况下的完整导数,即全导数(total derivative):

$$\frac{D}{Dt} = v \cdot \nabla + \frac{\partial}{\partial t}$$

通过这一点我们就把反映物质转移与质量守恒的连续性方程写成了以下更具有解释性的形式:

$$\rho \nabla \cdot v + \frac{D\rho}{Dt} = 0$$

这意味着,流体速度场存在散度就会导致流体质元随运动发生密度变化.从而也把不可压缩流体满足的两个条件联系在了一起.

第二个特殊情况与条件是,如果流体的流速场不随着时间变化而仅仅是位置的函数:

$$v = v(r)$$

这样的流动就称为定常流动(steady flow).否则称作非定常流动(unsteady flow).定常流动的意义需要在本章过程中慢慢体会.目前已经能够发现一点:只有在定常流动情况下,流线与迹线才是彻底重合的.

下面类似于弹性体,我们把流体在某个代表点,不妨就设在原点 0 ,附近的 r 处的流速场进行泰勒展开,保留到一阶:

$$v(r) = v(0) + r \cdot \nabla v$$

同理,把张量 ∇v 分解为对称部分和反对称部分:

$$\nabla v = \epsilon + \omega$$

$$\epsilon = \frac{\nabla v + {}^t \nabla v}{2}, \quad \omega = \frac{\nabla v - {}^t \nabla v}{2}$$

这样流体的运动就被分解为三个部分:

$$v(r) = v(0) + r \cdot \omega + r \cdot \epsilon$$

最后一项尤其值得注意,前两项表示流体跟随中心整体的平动和绕中心的转动,质元与质元之间本质上是没有相对运动与变形的.第三项就表示会不可避免地产生质元与质元相对摩擦的情况.它由对称张量 ϵ 描述.与弹性体的应变张量相似却又不同,这里不是位移而是速度作为变量去求导.称作应变率张量(rate-of-strain tensor):

$$\epsilon = \begin{bmatrix} \frac{\partial v_x}{\partial x} & \frac{1}{2} \frac{\partial v_y}{\partial x} + \frac{1}{2} \frac{\partial v_x}{\partial y} & \frac{1}{2} \frac{\partial v_z}{\partial y} + \frac{1}{2} \frac{\partial v_y}{\partial z} \\ \frac{1}{2} \frac{\partial y}{\partial x} + \frac{1}{2} \frac{\partial v_x}{\partial y} & \frac{\partial v_y}{\partial y} & \frac{1}{2} \frac{\partial v_x}{\partial z} + \frac{1}{2} \frac{\partial v_z}{\partial x} \\ \frac{1}{2} \frac{\partial v_z}{\partial y} + \frac{1}{2} \frac{\partial v_y}{\partial z} & \frac{1}{2} \frac{\partial v_x}{\partial z} + \frac{1}{2} \frac{\partial v_z}{\partial x} & \frac{\partial v_z}{\partial z} \end{bmatrix}$$

最后还需要描述流体中的力.用应力张量来描述流体中的受力情况是合适地.但是我们发现问题可以进一步简化.首先我们考虑一种更简单的情况:流体静力学(hydrostatics).

我们知道,在流体中没有静摩擦的可能性.流体中的摩擦,即湿摩擦(wet friction),只可能是静摩擦.这是因为流体没有保持初始形态的能力,它只有一定的保持体积的能力,故对于不改变体积的剪应变,顶多只能阻碍,但不能恢复.故一个流体在剪应力下应当马上运动起来,不可能处于静止状态.故流体静力学必然只能处于正应力下,而且沿不同方向的正应力还必须相等,否则在不同坐标系下就有可能产生剪应力.这实际上就证明了我们熟悉的流体静力学内应力由标量压强(pressure)来描述的观点,注意正应力以拉力为正,压强对应负的正应力:

$$\sigma = -pI$$

那么计算一个体积区域 V 边界 ∂V 上的压力合力,可以由高斯定理得到:

$$\mathbf{F} = \oint_{\partial V} dS \cdot \sigma = - \oint_{\partial V} pdS = \int_V \nabla \cdot \sigma dV = - \int_V \nabla p dV$$

这样就计算出来了内应力造成的体积力:

$$\frac{d\mathbf{F}}{dV} = -\nabla p$$

例如在重力场中的水 $\nabla p = \rho g$ 为常矢量,那么对没入水中的体积造成的浮力就是:

$$\mathbf{F} = - \int_V \nabla p dV = - \int_V \rho g dV = -\rho g V$$

这就是古希腊时期就被发现的著名的阿基米德原理(Archimedes' principle),今天看来它等价于高斯定律和尚未证明但正确的 $\nabla p = \rho g$.这个式子其实就是平衡方程.设想处于平衡状态的静力学流体还受到外力场,体积力为 $f(r)$.那么为了平衡就需要满足:

$$\nabla p = f$$

马上就可以发现,这个式子一方面作为压强分布的计算公式,同时,也由于左侧是一个标量场的梯度,在闭合路径上的积分必然给出零的结果,它又等于右侧,我们等于证明了:只有在外保守力场下的流体系统才有可能处于平衡状态.

那么,当流体开始运动起来,此时继续用压强描述流体内部应力就不再合适了,此时与上一章弹性体相同的是,我们可以引入对称的应力张量来修正原来的压强对应的单位张量:

$$\sigma = -p_0 I + \begin{bmatrix} \sigma_x & \tau_{xy} & \tau_{zx} \\ \tau_{xy} & \sigma_y & \tau_{yz} \\ \tau_{zx} & \tau_{yz} & \sigma_z \end{bmatrix}$$

但仍然有一点值得注意,如果 $\sigma_x, \sigma_y, \sigma_z$ 同时增加一个量,这和改变了 p_0 没有本质区别,所以我们希望压强 p_0 反映了正应力的平均效应,故调整前后的配比使得后面的张量称为无迹的张量 τ ,称为黏滞应力张量(viscous stress tensor)或偏应力张量(deviatoric stress tensor):

$$\tau = \begin{bmatrix} \tau_x & \tau_{xy} & \tau_{zx} \\ \tau_{xy} & \tau_y & \tau_{yz} \\ \tau_{zx} & \tau_{yz} & \tau_z \end{bmatrix} \quad (\tau_x + \tau_y + \tau_z = 0)$$

针对黏滞应力张量是否会产生,即流体有没有黏性(viscosity).我们又可以找到一种特殊的情况:无黏性流体(inviscid fluid),实际上经常被定义为理想流体(ideal fluid).此时流体即使在动,其应力也只有压强.反过来就是需要考虑黏滞的黏性流体(viscous fluid).

5.2 定常流动动力学

本节讨论范围仅限于无黏性流体。前半部分结论对于普遍的可压缩、非定常流体都是完全正确的。

若想为流体找到一个动力学方程，其实我们几乎已经在上一节完成了一大半。由于没有黏滞，上一节计算的内力的合力结果为：

$$\frac{d\mathbf{F}}{dV} = -\nabla p$$

再加上外体积力 \mathbf{f} ，现在要做的只不过是把上一节的平衡方程改成质元的牛顿定律，注意其中的加速度，应该用跟随质元一起运动而计算的全导数：

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p + \mathbf{f}$$

这就已经是能充分地解决问题的，流体需要符合的动力学方程。它被称作**欧拉方程**(Euler equation)。

另一个高频使用的方程是，如果我们再加上定常流动和不可压缩流体的限制条件，这就会导致著名的**贝努力方程**(Bernoulli equation)。暂时我们只加上定常流动这一个条件

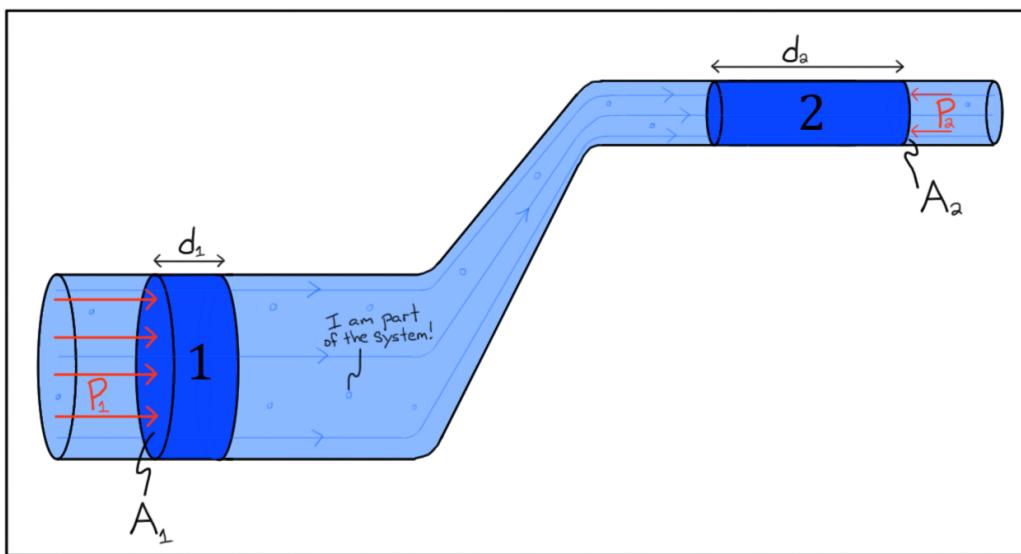


图 5.2: 贝努力方程

我们找到一根流线，并找到流线上的两个点 1, 2，在流线上垂直于流速方向把流线扩充为流管。那么在两处的面积 A 和密度 ρ 就要满足连续性方程，这对应着**流量**(flux) 的不变性：

$$Q = \rho_1 v_1 A_1 = \rho_2 v_2 A_2$$

将外体积力写作质量势 φ 的梯度 $\mathbf{f} = -\rho \nabla \varphi$ 。写出定常流动下的欧拉方程：

$$\mathbf{v} \cdot \nabla \mathbf{v} = -\frac{\nabla p}{\rho} - \nabla \varphi$$

如果计算以下量的梯度：

$$\nabla \left(\frac{v^2}{2} + \frac{p}{\rho} + \varphi \right) = \mathbf{v} \times (\nabla \times \mathbf{v}) + \mathbf{v} \cdot \nabla \mathbf{v} + \frac{\nabla p}{\rho} - \frac{p}{\rho^2} \nabla \rho + \nabla \varphi$$

借助欧拉方程，上式三项互相抵消，仍然剩下两项。但是现在我们在考察左侧标量在同一条流线上的表现，两侧点乘 \mathbf{e}_v ，右侧又乘项由于垂直也消失：

$$\frac{d}{ds} \left(\frac{v^2}{2} + \frac{p}{\rho} + \varphi \right) = -\frac{p}{\rho^2} \frac{dp}{ds}$$

这说明我们得不到一个沿着流管守恒的量. 事实上, $-\frac{1}{\rho^2} \frac{d\rho}{ds}$ 项其实就是 $\frac{d}{ds} \frac{1}{\rho}$, 即单位质量的物体具有体积沿着流线的变化率, 那么它再乘以压强, 就表示由于体积增加导致的对外界做功而减小的内部能量¹:

$$-\frac{p}{\rho^2} \frac{d\rho}{ds} = -\frac{du}{ds}$$

这样就得到了可压缩流体的贝努力方程:

$$u + \frac{v^2}{2} + \frac{p}{\rho} + \varphi = \text{Const.}$$

通过这一点可以发现, 可压缩流体的内能往往是与动力学耦合在一起的, 而内能又往往影响温度. 所以内能的改变又会导致热学与动力学的耦合, 问题就变得很复杂.

但是如果流体不可压缩, 那么上式等号右侧 ρ 就是个常数, 从而直接左侧就具有零的导数. 那么我们就得到了不可压缩流体的贝努力方程, 往往写作:

$$\frac{\rho v^2}{2} + p + \rho \varphi = \text{Const.}$$

5.3 黏滞流体动力学

描述黏滞流体的最标准的模型是牛顿流体(Newtonian fluid) 模型. 核心的定律称作牛顿黏滞定律(Newtonian law of viscosity). 如图5.3. 如果流体速度场是完全单向的, 且其大小的改变 dv 也只在垂直流速方向移动 dl 时发生, 那么就定义流速梯度为:

$$2\varepsilon = \frac{dv}{dl}$$

而由此产生的切应力就是:

$$\tau = \frac{dF}{dS}$$

牛顿定律指出, 两者应当成正比, 比例系数就是黏滞系数(viscosity), 常简称黏度 η . 其单位是 Pa · s:

$$\tau = \eta \cdot 2\varepsilon$$

其实这里的黏度是动力学黏度(dynamic viscosity) 的简称. 实用中常常遇到需要把它除以介质密度的情况, 所以还有运动学黏度(kinematic viscosity) 的定义, 其单位是 m²/s, 与扩散系数单位是一样的:

$$\nu = \frac{\eta}{\rho}$$

显然, 当我们考虑把牛顿黏滞定律进行推广时, 第一个值得注意的现象就是, 首先如果把流速方向记做 x 方向而变化方向记做 y 方向, 那么一是应力 τ 其实就是作用于 $x-y$ 平面上的剪应力, 为了满足流体的力矩平衡, 不仅对 x 方向的分界面会产生 x 方向的黏滞摩擦力, 对 y 方向的分界面也同时会造成 y 方向的力以抵消这个效应, 在上图中, 总是左侧的液体对右侧的液体给向下的力, 右侧则给左侧向上的力.

第二点是, 显然下式不具有对称性:

$$\tau_{xy} = \eta \frac{\partial v_x}{\partial y}$$

如果是 v_y 在随 x 产生改变, 上式就正确了, 此时同样会产生剪应力. 所以我们猜想如下式子是正确的:

$$\tau_{xy} = \eta \left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right)$$

¹还需要绝热条件, 这样内能改变就只和做功有关.

事实上这也是牛顿黏滞定律唯一可能正确的推广形式. 单凭借这个结论我们就能得到很多有意义的结果. 这里直接不加证明地介绍其中的两个. 一是圆柱管道内的液体流动的哈根-泊萧叶方程(Hagen-Poiseuille equation). 如果管道半径 R , 体积流量 Q , 那么在长为 l 的一段上形成流动需要的压强差为:

$$\Delta p = \frac{8\eta Ql}{\pi R^4}$$

第二个式子是球形物体在黏滞流体中的阻力公式. 著名的斯托克斯定律(Stokes law). 如果球体半径为 R , 速率为 v , 那么阻力大小为:

$$F = 6\pi\eta rv$$

普遍的情形下, 之前介绍过的黏滞应力张量与应变率张量之间的关系, 根据对称性的原理, 与上一章类似地, 就写作:

$$\tau = 2\eta \cdot \varepsilon + \lambda \text{Tr}(\varepsilon) \cdot I$$

只不过在这里, 由于 τ 的无迹性, λ 没有别的选则, 只能取 $-2\eta/3$. 这样就得到了黏滞流体的内力的决定性的方程. 最后, 类似于欧拉方程的方式进行推导, 我们最终得到了一个方程:

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho \mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p + \eta \nabla^2 \mathbf{v} + \frac{1}{3} \eta \nabla (\nabla \cdot \mathbf{v}) + \mathbf{f}$$

这个方程就是著名的纳维-斯托克斯方程(Navier-Stokes equation). $\mathbf{v} \cdot \nabla \mathbf{v}$ 项使得这个方程成为非线性的偏微分方程. 它的求解是如此之难, 以至于在 2000 年它成了 7 个千禧年难题(Millennium Prize problems)之一, 而且仅仅是为了解决最最初步的问题: 解的存在性与光滑性. 第一个解决它的科学家将获得一百万美元的奖金, 但目前这个问题的解决依然是遥遥无期的.

5.4 * 流体中的波

5.5 波的色散

第六章 热力学第一定律

6.1 热力学第零定律

热力学系统的状态可以分为平衡态(equilibrium state) 和非平衡态(non-equilibrium state). 平衡态这样一种状态, 首先它是一种宏观性质都不随着时间演化的状态---稳定态¹, 但稳定态不一定是平衡态, 如两端温度保持不变后稳定导热的杆, 如稳定喷气的火箭, 如不断气化液化工作循环制冷的空调工作物质. 原因是它们没有符合以下三个平衡条件:

热学平衡 系统各部分温度严格相等.

力学平衡 系统内力(如压强)与外力达到平衡, 如重力场下的等温大气模型.

化学平衡 系统各部分热运动扩散的趋势相互平衡, 如饱和糖水与糖块共存.

可以发现, 既使系统状态不均匀, 如重力场等温大气模型的各处压强和饱和糖-糖水模型中的两相糖分子数密度, 系统也是处于平衡态的. 而如果这些平衡条件没有达成, 系统内部将发生各种宏观可见的输运过程(transport phenomena). 如温度不均导致的热传导, 速度与压强不均导致的黏滞, 及浓度不均导致的扩散. 不平衡的力产生了不平衡的流, 相应的研究是非平衡态物理研究的对象. 这种情况下既使系统随时间稳定, 也不是平衡态.

上文中已多次提及温度(temperature)这一物理概念. 今天的我们完全可以用统计力学的思考方式来理解它: 诚然, 温度是分子热运动剧烈程度的量化. 如理想气体的温度定义为:

$$T = \frac{2\bar{\varepsilon}_k}{3k}$$

其中 $\bar{\varepsilon}_k$ 表示分子平均平动能, 而 k 表示玻尔兹曼常数. 也就不难理解, 为什么温度不均将导致传热, 因为分子平均动能大的子系统 A 与分子平均动能小的子系统 B 发生热接触时, 平均效应是 A 中的分子通过相互作用要把能量给 B. 这叫热传递(heat transfer).

然而, 在统计力学还没有诞生前, 甚至在很长的一段时间内人们对热与温度的能量本质并没有明确的认识, 热质说(caloric theory)大行其道, 人们把热量导致的温度升高视为与功导致的能量升高为完全独立的两个维度. 但正是在这样的环境内, 诞生了基于经验规律的物态方程, 发明了测温, 量热的各种实用仪器, 乃至甚至产生了公理化的热力学体系. 热力学第零定律即给出了温度可定义的依据.

热力学第零定律 任意两个各自达到热平衡的系统若发生热接触可以直接形成热平衡, 则说这两个系统具有一致的温度. 等温具有反身性, 交换性与传递性.

热力学第零定律的补充 如果两个各自达到热平衡的系统若发生热接触一者传递给另一者热量, 则说前者比后者温度高.

¹注意, 一是从微观上看, 组成系统的微观粒子仍然处于无休无止的运动中, 其微观坐标与速度在发生几急剧的变化, 但平衡态只对宏观性质定义. 二是统计力学可以证明, 处于平衡态的系统总是由于内在随机性而产生涨落, 有时甚至不可忽略, 如临界乳光现象. 这说明我们的平衡态模型只能是一种近似, 实际情况往往还有很多需要考虑的因素.

这一定律确定了温度作为系统的**态函数**(state function), 它能像一个数那样可以进行大小的比较. 有两点值得注意:

一是热力学第零定律定义了温度, 但它没有量化一个**温标**(temperature scale). 如何量化一个温标呢? 从各种系统平衡态的性质出发可以以某一具有特点的状态量作为温度评价的标度. 如理想气体温标, 黑体辐射谱温标和噪声温标等等. 而我们发现, 温度高低的比较似乎与其传热的难易程度有关, 是否可以利用这一点自然而然地定义普适温标呢? 热力学第二定律回答了这个问题, 答案是肯定的.

二是热力学第二定律与热力学第零定律的内在联系. 作为热力学第零定律的补充, 其实它正是热力学第二定律的一部分. 热力学第二定律中得到了很好的量化定义了热力学温标以完善温度这一概念. 其实, 这两个定律的背后是简单的统计规律, 若我们采取一种统计力学的方法, 我们的理论基础便是统计公设, 而热力学定理反而成为我们需要去证明和解释的对象了.

6.2 热力学第一定律

热力学第一定律的研究对象是热力学过程, 是既定系统连续发生的这样或那样的真实状态改变. 而热力学第一定律的微观本质是普遍的能量转化与守恒原理. 它写为:

$$dU = dW + dQ$$

U 为系统内能, 表示系统内部能量的总和. 等式左边为内部能量增量, 右边为外界输入系统内部的能量, 其中 dW 为外界对系统内部的做功, 而 dQ 为外界通过热接触, 以微观碰撞交换能量的形式, 产生的平均宏观能量输送的大小. 注意, U 前面的 d , 表示无穷小变化量, 但 W 与 Q 前面的 d , 仅仅表示无穷小量(微元), 不是某个热力学态函数的微分. 这也是两个微分符号记法不相同的原因. 这两种不同的热力学量可以称为**状态量**(state quantity) 与**过程量**(process quantity).

在具体考察上述三个量的具体范围前, 明确一下过程发生的具体形式是十分有必要的.

6.2.1 准静态与非准静态过程

准静态过程(quasi-static process): 过程进行地速度无限缓慢, 以至于每一时刻系统都几乎处于平衡态. 如等温传热的过程; 保持外界压强始终等于气缸压强地压缩气体的过程; 亦或是缓慢地冷却热溶液, 使得溶质析出的过程, 都是准静态过程. 相反, 高温物体向低温物体放热的过程, 气体向真空自由膨胀的过程, 把糖丢进纯水里溶解的过程或两种不同气体混合的过程, 就都是非准静态过程.

过程的准静态与非准静态有以下两个特征:

一是准静态由于每时每刻都是平衡态, 所以可以由平衡态的均匀状态参量去描写, 从而得出各状态参量随时间的变化关系, 消去时间可得**状态方程**(process equation). 更特别地, 在过程中所发生的内能变化, 做功, 吸热全都可以用状态方程去确定, 比如以 p , V 作为状态参量的理想气体:

$$dU = \frac{\partial U}{\partial p} dp + \frac{\partial U}{\partial V} dV ; \quad dW = -pdV ; \quad dQ = dU - dW$$

注意在上式中最后一式不是说 dQ 是 dU 与 dW 的结果, 在热一的公式中, 等式右边的做正负功和吸放热永远是原因, 等式左边内能的增减永远是结果. 不过这是一个很常用的计算式, 图像也还算直观: 气体吸收的热量用来增加内能和对外做功.

而非准静态过程则不能由状态方程描写, 也没有以上三个方程的成立, 我们在更晚些时候展开讨论.

二是准静态过程由于每时每刻都达到了平衡, 在平衡的状态下发生热量, 功或者粒子的交换, 所以是可逆过程². 而由于非准静态过程没有达到各种平衡, 故自发产生的各种过程都是不可逆的.

我们接下来仔细考虑内能, 功与热量三个物理概念.

²严格地说, 还要求过程没有耗散, 耗散是一类功变热的行为, 它使得有序的功对应的能量转化为无序的热对应的能量.

6.2.2 内能

内能(internal energy), 就是系统内部的能量和. 它由许多可能的成分构成, 如:

分子平动能 三个自由度;

分子转动动能 线性分子只有两个自由度³; 非线性分子则有三个;

分子振动动能 ⁴ 分子间键长, 键角偏离平衡位置的集体振动模式, 其自由度数与上两者之和为 $3n, n$ 为分子含原子数;

分子振动势能 由于分子的振动, 其势能也会比其平衡位置更大, 近似地与分子振动动能相等;

键能 即电子运动的动能势能之和 (全部能量);

静能 这里指以上运动的动能, 势能全部去除后, 由于更加深层次内部相对运动造成的静能量, 如核能, 基本粒子的自能等等.

有两点值得注意:

一是在内能具体的计算中, 对于我们不关心的, 过程中不发生改变的部分我们将其置零. 如双原子分子理想气体的问题中研究的内能不会去包括两原子间化学键的能量, 除非发生化学反应. 而在常温下其振动自由度往往难以被激发, 故其等效自由度约为 5. 自然界对于尺度的放缩并不具有对称性, 所以很多能量实际上在更微小的层次内封装为不变的整体, 这使得我们只需要研究我们关心的尺度上的问题, 即**有效理论**(effective theory) 的思想.

二是有一些形式的能量是否属于内能的一部分视问题的不同而可以做不同的处理. 这体现为功能原理上:

$$dW_{\text{conservative}} = -dE_{\text{potential}}$$

即, 在热力学第一定律的计算中, 我们不去计算保守力部分做的功. 而是把它对应的势能视为系统内能的一部分:

$$dU = dW_{\text{others}} + dW_{\text{conservative}} + dQ \implies dU_{\text{total}} = dU + dE_{\text{potential}} = dW_{\text{others}} + dQ$$

最经典的例子是那些热力学系统中每个粒子都具有外场中的相互作用势能的情况. 如气体的重力势能与电介质的静电能, 就常常当做系统内能的一部分, 然而有些情况下又使用不含这些能量的系统内能更加方便. 再比如在恒定的外压强下的气体系统, 外界对气体做功可以用体积功来表示:

$$W = \int_V^{V+\Delta V} (-p_0 dV) = -p_0 \Delta V$$

从中可以看出, 气体具有体积能 $E_{\text{volume}} = p_0 V$, 可以定义新的总内能为:

$$H_0 = U + p_0 V$$

如果气体与外界压强平衡⁵, 则上式中的压强就是气体压强 p , 以上量称为气体的**焓**(enthalpy):

$$H = U + pV$$

³分子对自转轴的转动不会带来系统的改变, 这与全同粒子的量子本性有关系.

⁴注意由于量子效应, 室温下振动能量可以认为不被激发.

⁵如系统在大气压下发生一个化学反应等等.

6.2.3 功

功(work) 有两个特点:一, 功是一种能量转移, 考虑功能原理, 对一个系统做正功使得系统动能增加, 而这是系统内能的一部分, 从而能量传入系统. 二, 功总是与力对应. 热力学中为了方便我们引入所谓**广义力**(generalized forces) 与**广义坐标**(generalized coordinates) 来方便表示我们的系统. 如理想气体广义坐标如果取做体积 $X = V$, 则广义力就是外压强的相反数 $Y = -p_0$, 它们满足外界对研究的系统做功为:

$$dW = Y dX = -p_0 dV$$

这被叫做**体积功**(volume work). 与其他形式的功对应的广义力与广义坐标并在一张表中供参考:

| 功形式 | 广义力 | 广义坐标 |
|-------|----------|------|
| 体积功 | $-p$ | V |
| 表面张力功 | σ | S |
| 电路功 | U | Q |
| 极化功 | E | P |
| 磁化功 | H | M |

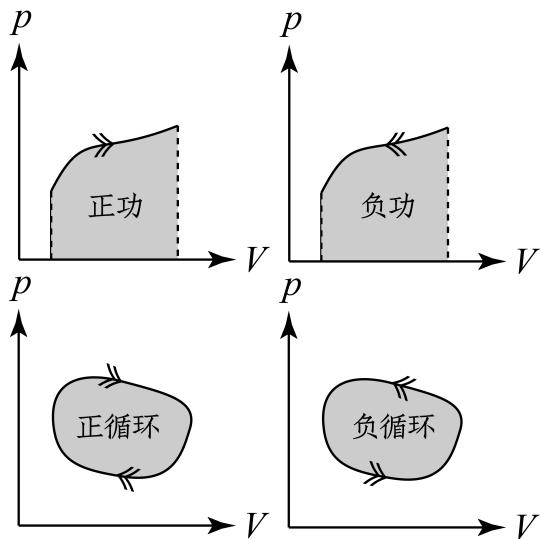


图 6.1: 准静态过程的做功

对于体积功, 在准静态过程中可以由 $p - V$ 图下的面积计算. 在循环过程中则等于过程曲线所包含的面积. 注意其正负号与过程方向之间的关系. 但对于非准静态过程. 做功一定要由外界参数 (活塞加速度, 活塞外气压等) 计算而不是用内部气体的过程方程. 这是因为即使形式地写出内部气体所发生的过程的过程方程. 由于气体发生膨胀处的压强不等于过程方程中的平均压强, 对做功的计算是没有意义的.

6.2.4 热量

做功与**传热**(heat transfer) 是截然不同的两种向系统注入能量的方式. 做功依赖于广义坐标这一参数的改变. 本质上与每一个微观粒子的能量对于广义坐标的依赖有关. 然而热量则完全不依赖于外在的参数. 它与系统粒子关于不同能量的状态的占据数**配分**(partition) 有关. 也就是与温度息息相关. 温度高的物体, 占据在高能量状态上的粒子较多, 低能量状态的粒子较少, 与低温物体发生热接触时, 单位时间内高能粒子碰撞低能粒子传递能量的可能性更高. 其统计平均效应就是高温物体向低温物体放热. 准静态的**等温传热**又是怎样一个过程呢? 从以上分析可知, 产生温度梯度是有热传导⁶的必要条件, 没有了温度差热量的交换也就无从谈起了. 从这个意义上说, 实际发生的过程其实都是非准静态过程, 准静态过程仅仅是一种实际过程的一种近似. 或是我们虚设的一种模型. 如在等温传热模型中, 我们设想只让一个物体的高能粒子与另一个物体的低能粒子发生热接触而实现热传递, 这无疑会破坏热力学的统计规律⁷, 但有助于我们理解这个过程的发生机理.

⁶注意, 广义的热量传递还包括热对流与热辐射.
⁷或是借助于所谓的麦克斯韦妖, 引入信息熵.

6.2.5 耗散

还有一类向系统输送能量的方式,表面上看是功,效果上却是热.这被我们称为**耗散**(dissipation).俗称**功变热**.例如常见的摩擦生热过程,非弹性碰撞的过程,激光泵浦的过程.都是典型的功变热的例子.我们看一个十分浅显的例子.涡轮叶片搅动气缸中的空气使其温度升高.在这个过程中,涡轮无疑对气体做功,但这个做功并不与系统广义坐标的变化相关.它的特点是仅仅作用在系统的一个子系统上.再由系统倾向于热平衡的趋势把得来的能量平分到系统各个自由度上.从而造成了与吸热完全类似的结果:广义坐标没有变化,但系统从外界获得了能量.

究其原因,本质上是因为系统具有了所谓的**耗散结构**(dissipative structure).它不把系统看成一个平衡的整体,而是有着复杂内部非线性相互作用的对象.故输入的功不一定改变体系的广义坐标,但像加热一个系统一般,增加了系统的内能.

耗散结构的特点之一是外力驱动系统远离平衡态.在非弹性的碰撞过程中固体内部的低频弹性波的成分被突然增强,而激光发射的过程中大量粒子被输送到亚稳态而实现布居反转,这些都破坏了原来热平衡下的布局.需要结合动力学与统计方法来讨论这些现象.

特点之二是开放条件下的自组织现象.以普利高津(Proggine)为首的研究者把生命现象,连同研究过程中的很多奇特现象看成是一种耗散结构的结果.这些过程伴随着系统与外界连续不断的能量,粒子与信息的交换,使得微小的涨落可以在具有正反馈的功能结构中被放大,使得体系的对称性被破坏,形成有组织的自发序.

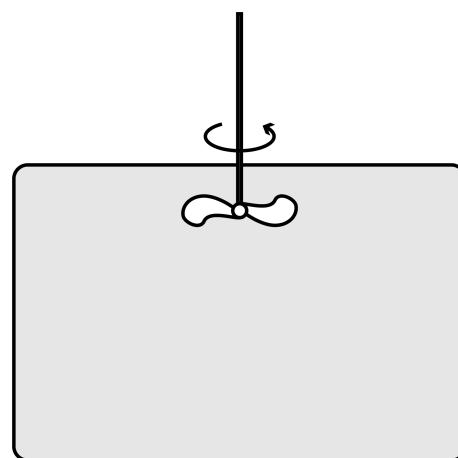


图 6.2: 搅动空气发热



图 6.3: 自组织现象: Bénard 对流与斑马身上的条纹

6.3 理想气体

6.3.1 理想气体的定义

满足以下三个条件的气体为理想气体:

1. 分子间作用力可以忽略;
2. 分子体积可以忽略;
3. 分子按自由度平分能量.

我们讨论一下这三个条件. 首先第一个条件的意思, 旨在把理想气体间的作用力当成短程力(short-range force) 处理. 仅仅当两个分子间距离进入力程范围才可以当做可以发生相互作用(碰撞或散射). 在这个意义上, 分子体积实际上也可以处理为半径为力程大小的钢球模型. 我们忽略这一个体积, 也就是说所有分子可以自由活动的体积就等于容器体积 V . 且分子间相互作用力为零.

然而这两点合在一起将导致一个严重的问题: 那便是热平衡不可以达到. 若每个分子不可以通过相互作用或刚性的碰撞与别的分子交换能量与动量. 那从根本上就不可以达到热平衡, 统计方法也就无济于事. 故分子体积足够小, 小到可以忽略其对热力学性质的影响, 但又足够大, 大到在我们关心的时间空间尺度内体系能很快地达到热平衡. 这才是对这两点的正确理解方式.

我们设分子的速率分布函数为 $f(v)$, 意思是每取 N 个分子, 速率介于 v 到 $v + dv$ 的分子数平均为:

$$dN = Nf(v)dv$$

从而分子的平均速率与方均速率为:

$$\bar{v} = \int_0^\infty v f(v)dv ; \quad \bar{v^2} = \int_0^\infty v^2 f(v)dv$$

分子质量为 m , 直径为 D , 数密度为 n . 从中我们得出分子间平均距离 d :

$$d \sim \frac{1}{\sqrt[3]{n}}$$

可以想见, 分子间平均距离 $d \sim 10\text{nm}$ (大气压下) 越大, 分子直径 $D \sim 1\text{nm}$ 越小, 气体越能够近似为理想气体. 同时其相隔两次碰撞的平均时间间隔 τ 与空间间隔 λ 也会越大. 如何计算这两个量呢? 由于分子的速度符合麦克斯韦分布律(参见第五章), 故两分子间的平均相对速率为 $\sqrt{2}\bar{v}$. 将待研究的粒子处理为钢球, 速率恒定为 $\sqrt{2}\bar{v}$, 其半径为原分子直径 D . 而其他分子处理为静止的质点, 半径为零. 这样钢球碰到质点, 即相当于原来的两个分子发生碰撞. 而时间 t 内钢球扫过的体积为:

$$V(t) = \pi D^2 \cdot \sqrt{2}\bar{v}t$$

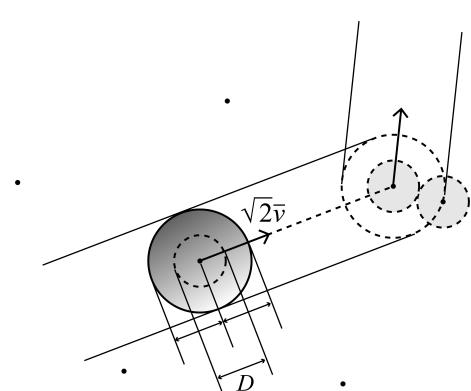


图 6.4: 曲折柱体计算

发生碰撞次数 c :

$$c(t) = nV(t)$$

当 $c = 1$ 时, 对应的时间即为平均自由时间 τ , 又称为驰豫时间(relaxation time), 而这一段时间内粒子走过的距离即为平均自由程 λ (mean free path):

$$\tau = \frac{1}{\sqrt{2}\pi n D^2 \bar{v}} ; \quad \lambda = \frac{1}{\sqrt{2}\pi n D^2}$$

即有:

$$\lambda \sim \frac{d^3}{D^2} \sim 100d \sim 1\mu\text{m}$$

以上数据仅仅在大气压下有效,注意平均自由程与分子平均速度无关,仅仅是一个几何相关量.而它与 n 成反比,也就是与 p 成反比,如果压强减小到约 1Pa(而温度不变),则平均自由程 λ 增大 5 个量级,也就是分米量级.

现在我们做一个总结:我们取了一个分子间相互作用力与分子的大小可以忽略的模型来研究理想气体,它对应的条件是:

$$D \ll d ; \quad \lambda \sim \frac{d^3}{D^2} \ll L$$

第一个条件是保证分子体积可以忽略,第二个条件是在研究的问题尺度内保证分子有足够的空间来实现动量能量交换以达到热平衡.

现在让我们专心研究第三个条件.能量按自由度均分是麦克斯韦分布的一个独特结果.详情也可以参见第五章.现在我们指出,由于系统中的各个自由度,如 x, y, z 方向的平动,分子的转动,振动的能量具有相似的形式⁸,而且可以通过相互作用(碰撞)交换能量,总能量守恒.在这样的条件下,统计规律告诉我们每一个分子每一个自由度上都分得相同的平均能量值.这个平均能量值⁹即代表了体系的温度:

$$\bar{\varepsilon} = \frac{1}{2}kT$$

体系自由度越多,分子数越多,在一定温度下储能就越多:

$$U = N \cdot \frac{f}{2}kT$$

其中 f 是系统能量自由度数¹⁰.它被视为与温度无关的常量¹¹.对于上式,我们再引入摩尔(mole)这一化学计量单位来表示巨大的原子数 N :

$$\frac{N}{N_A} = \nu ; \quad N_A = 6.022140857 \times 10^{23}/\text{mol}$$

同时,普适气体常量被定义为:

$$R = N_A k = 8.314 \text{J/mol} \cdot \text{K}$$

这样气体的内能即为:

$$U = \frac{f}{2}\nu RT$$

6.3.2 理想气体物态方程

我们发现, $dp_1 = f(v)dv$ 表示的含义为任意一个分子的速率落在 $v \sim v + dv$ 区间内的概率.然而对于这样的分子,它速度方向在空间中的取向是任意的,各向同性的.所以建立极轴后,速度与极轴夹角在 $\theta \sim \theta + d\theta$ 范围内的概率又为:

$$dp_2 = \frac{d\Omega}{4\pi} = \frac{2\pi \sin \theta d\theta}{4\pi} = \frac{\sin \theta d\theta}{2}$$

对于这样的分子,它在容器中的任何位置都是有可能且等概率的,所以出现在体积微元 dV 中的概率还有 dp_3 ,联合概率为 $dp = dp_1 dp_2 dp_3$.

⁸指谐振子形式,详见第五章.

⁹及其对应的按能量高低的配分.

¹⁰注意其与动力学自由度的区别,振动的势能也对应能量自由度,但其动力学自由度与振动动能重合.

¹¹然而实验发现这些不是常量,而是随着温度升高越来越大,存在一些基本不变的平台,思考这意味着什么.

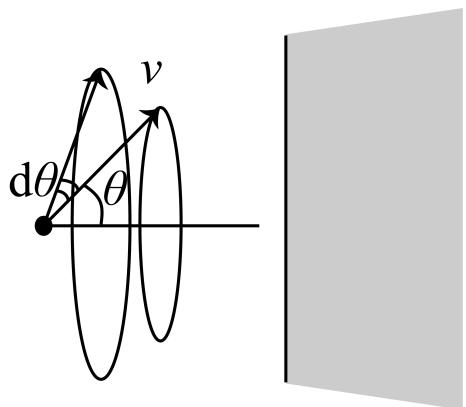


图 6.5: 粒子速度方向分布

我们接下来考虑气体的压强.在这之前我们了解两个概念是十分有帮助的:处于平衡态的系统可以看成互相达到平衡的各部分子系统(subsystem)直接组成.这些部分的所有强度量(intensive property)都是完全一致的,如理想气体的压强,温度,粒子数密度,摩尔体积,摩尔内能等等.而子系统的广延量(extensive property)大小之和即为总系统的对应量的大小,如分子数,摩尔数,体积,内能,焓等等.我们可以这么理解:强度量描述体系的内禀属性,而广延量仅仅反应全同子系统的简单堆积.如一杯盐水,确定了我们关心的强度量——温度和浓度之后,我们就可以回答“怎样的盐水”这一个问题.而接下来只需要再回答“有多少”这一个简单的问题,就能算出所有的广延量来.基于这一点理解,我们可以发现以下两个性质是成立的:

保持所有强度量不变,一个广延量变化 λ 倍,所有广延量随之变化 λ 倍.

两个广延量之比为强度量,典型的如质量 M 比体积 V 为密度 ρ ,体积 V 比摩尔数 ν 为摩尔体积 v 等.

理想气体作为极其简单的系统,它的独立强度量仅仅只有两个.一是系统的温度 T .它决定了分子的平均热运动剧烈程度,也就是平均能量大小.另一个就是分子数密度 n ,它反映单位体积里有多少个全同单元,是一种“浓度”.其他所有强度量都是这两个量的导出.比如压强 p ,用分子动理论的方法:

$$\begin{aligned} p &= \int_0^\infty \int_0^{\frac{\pi}{2}} n dS v \cos \theta dt \cdot f(v) dv \cdot \frac{\sin \theta d\theta}{2} \cdot 2mv \cos \theta / dS dt \\ &= nm \cdot \int_0^\infty v^2 f(v) dv \cdot \int_0^{\frac{\pi}{2}} \sin \theta \cos^2 \theta d\theta \\ &= \frac{1}{3} nm v^2 \\ &= \frac{2}{3} n \bar{\varepsilon}_k \\ &= nkT \end{aligned}$$

可见对容器壁的压强正比于温度和分子数密度,这一点的物理图像还是很直观的.这个压强被我们称之为外压强(outer pressure).它是由于分子碰撞容器壁引起.还有所谓的内压强(inner pressure)的概念,它由于分子间的碰撞引起.理想气体的内压强与外压强相等¹².但如果气体足够稀薄,以至于容器尺度与分子平均自由程可以比拟.此时内压强的概念也失去了意义,这意味着气体完全不可以用动力学方法处理,压强的梯度不会导致气体团的加速度,而是以下面所介绍的泻流的方式进行.

上文我们推出来的关系式被称作(强度量之间的)物态方程(equation of state),它有很多变式,如:

$$pV = NkT ; \quad pV = \nu RT ; \quad pv = RT ; \quad p = \frac{\rho RT}{\mu}$$

上面最后一式中 μ 表示摩尔质量,稍作变形还可以得到密度公式:

$$\rho = \frac{\mu p}{RT}$$

用分子动理论方法还可以得到另一个十分重要的量——泻流数(effusion number).如果在容器壁上开

¹²然而范德瓦尔斯气体则不然,详见第四章.

一个小孔. 则速率在在 v 到 $v + dv$ 区间内的分子, 单位面积单位时间通过小孔的分子数为:

$$\begin{aligned} d\Gamma &= \int_0^{\frac{\pi}{2}} n dS v \cos \theta dt \cdot f(v) dv \cdot \frac{\sin \theta d\theta}{2} / dS dt \\ &= \frac{n}{2} \cdot \int_0^{\frac{\pi}{2}} \sin \theta \cos \theta d\theta \\ &= \frac{1}{4} n v f(v) dv \\ &= \frac{1}{4} dn v \end{aligned}$$

可见速率为 v 的那一部分分子, 泻流数就是 $d\Gamma = \frac{1}{4} dn v$. 那么对于所有分子, 总泻流数为:

$$\Gamma = \frac{1}{4} n \bar{v}$$

泻流的条件是开的孔的大小足够小, 以至于分子在穿过孔时不会受到别的分子的碰撞. 也就是其平均自由程 λ 显著地大于孔的直径 L . 相反地, 如果 $\lambda \ll L$, 那么就变成了一种压强差驱动气体集团运动的动力学模型, 我们在下一节阐述.

最后, 理想气体的内能则可以根据之前关于温度定义的讨论, 直接写出:

$$U = \nu C_{mV} T$$

其中 C_{mV} 被称为定容摩尔热容(molar heat capacity at constant volume). 其意义之后将了解. 现在只需知道 $C_{mV} = \frac{f}{2} R$.

6.3.3 混合理想气体

考虑几种理想气体的混合, 我们也需要忽略几种组分分子间的相互作用, 尤其是应该排除几种分子间发生化学反应的情况¹³. 此时, 几种气体由各自的数密度 n_i 和热平衡下的公共温度 T 描述. 另一种描述方法是给出总分子数密度 n 与各组分分子所占的比例(强度量) x_i :

$$n = \sum n_i ; \quad x_i = \frac{n_i}{n}$$

称为摩尔分数(mole fraction), 注意 x_i 间要满足的归一关系:

$$\sum x_i = 1$$

不同组分间分子纵然有碰撞, 但这对于这一部分分子碰撞容器壁产生的压强大小没有影响. 也就是说, 压强等于各组分以 n_i 单独存在时对容器壁产生的分压 p_i . 这叫做道尔顿分压定律(Dalton's law of partial pressures):

$$p = \sum p_i ; \quad p_i = n_i kT = x_i p ; \quad p = nkT$$

值得注意, 每个分子的质量 m_i 或能量自由度数 i 并不影响压强的计算. 压强单方面地决定于分子数密度与温度(分子平均平动动能. 分子质量改变的是整个体系的摩尔质量 μ , 取一定体积的气体:

$$M = \nu \mu = \sum M_i = \sum \nu_i \mu_i \Rightarrow \mu = \sum x_i \mu_i$$

其中出现的摩尔数 ν_i 之比即为分子数密度 n_i 之比, 而摩尔质量 μ_i 与分子质量的换算关系也很基础, 请读者一定要熟练使用:

$$\nu_i : \nu_j = \frac{n_i V}{N_A} : \frac{n_j V}{N_A} = n_i : n_j ; \quad \mu_i = N_A m_i$$

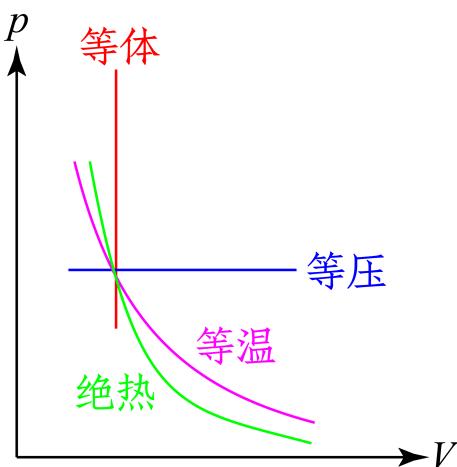
能量自由度数影响的是摩尔内能 u :

$$u = \nu \overline{C_{mV}} T = \sum u_i = \sum \nu_i C_{mVi} T \Rightarrow \overline{C_{mV}} = \sum x_i C_{mVi} = \sum \frac{x_i f_i}{2} R$$

上面我们定义了平均摩尔定容热容 $\overline{C_{mV}}$. 平均的含义一定是按分子数 n_i (摩尔数 ν_i , 比例 x_i) 平均.

¹³如混合NH₃与Cl₂.

6.3.4 理想气体的过程



所谓**过程**(process) 是状态的集合. 我们考虑的范围内, 过程的初态和末态都需要是均匀的平衡态. 中间态如果也可以视为平衡态, 就叫做**准静态过程**(quasi-static process), 若中间态蕴含着非平衡, 则称为**非准静态过程**(non-quasi-static process). 下面列举常见的过程:

等体过程 (isochoric process)

作为等体过程体积功为零 $W = 0$, 只能以从外界吸放热 Q (或是如图6.2所示的功变热形式.). 系统吸热导致温度 T 升高, 定义系统**热容**(heat capacity) 为:

$$C = \frac{dQ}{dT}$$

而压强 p 将升高, 定义**压强系数**(pressure coefficient) 为:

$$\beta := \frac{1}{p} \frac{dp}{dT}$$

根据理想气体的物态方程与内能公式我们算出¹⁴:

$$C_V = \left(\frac{dQ}{dT} \right)_V = \nu C_{mV} ; \quad \beta_V = \frac{1}{p} \left(\frac{\partial p}{\partial T} \right)_V = \frac{1}{T}$$

准静态的等压过程 (isobaric process)

若是控制压强 p 而不是体积 V 不变, 区别在于气体从外界吸收的热量中一部分要用来对外做功:

$$dU = dQ - pdV$$

热容按同样的方法定义, 分别称为等体热容和等压热容¹⁵. 而**热膨胀系数**(thermal expansion coefficient) 定义为:

$$\alpha := \frac{1}{V} \frac{dV}{dT}$$

注意这实际上是**体膨胀系数**(volumetric thermal expansion), 由近似方法可知, 近似为线膨胀系数 α_L 的三倍. 对于理想气体等压过程, 同理可得:

$$C_p = \left(\frac{dQ}{dT} \right)_p = \nu(C_{mV} + R) = \nu C_{mp} ; \quad \alpha = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_p = \frac{1}{T}$$

准静态的等温过程 (isothermal process)

系统与恒温大热库接触, 吸热膨胀, 压强减小; 放热收缩, 压强增大. 这时 pV 为过程不变量. 我们定义**压缩系数**(compression coefficient):

$$\kappa := -\frac{1}{V} \frac{dV}{dp}$$

注意若是在固体情形, 以上量被称为**体模量**(bulk modulus). 对于理想气体, 等温压缩系数为:

$$\kappa_T = -\frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_T = \frac{1}{p}$$

¹⁴热力学的角标表示这个量不变, 而这里热容的表达式下一章学了熵的概念后可以写:

$$C = T \frac{\partial S}{\partial T}$$

¹⁵中文太灵活, “等体” 也有人说成“定体”, “等容”, “定容”. 统一术语十分困难.

准静态的绝热过程 (adiabatic process)

绝热过程中 $dQ = 0$, 对于理想气体:

$$\begin{aligned} dU + pdV &= 0 \\ \nu C_{mV} dT + \frac{\nu RT}{V} dV &= 0 \\ C_{mV} \frac{dT}{T} + R \frac{dV}{V} &= 0 \end{aligned}$$

惯例上定义绝热指数(adiabatic index) γ 为:

$$\gamma := \frac{C_{mp}}{C_{mV}} = 1 + \frac{R}{C_{mV}} = \frac{f+2}{f}$$

则可以由 γ 和 R 表示两个摩尔热容:

$$C_{mV} = \frac{1}{\gamma - 1} R ; \quad C_{mp} = \frac{\gamma}{\gamma - 1} R$$

从而我们从绝热过程的热力学第一定律方程积分得:

$$TV^{\gamma-1} = \text{Const.}$$

也可以写为:

$$pV^\gamma = \text{Const.} ; \quad \frac{p^{\gamma-1}}{T^\gamma} = \text{Const.}$$

前一式可以用于在 $p-V$ 图6.6上确定过程曲线. 后一式则注重关心强度量之间的变化关系而不关心广延性质. 在绝热过程中 (用角标 S 表示绝热, S 其实表示熵, 具体原因见下一章.), 热膨胀系数, 压强系数与压缩系数分别为:

$$\alpha_S = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_S = -\frac{1}{\gamma-1} \frac{1}{T} ; \quad \beta_S = \frac{1}{p} \left(\frac{\partial p}{\partial T} \right)_S = \frac{\gamma}{\gamma-1} \frac{1}{T} ; \quad \kappa_S = -\frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_S = \frac{1}{\gamma} \frac{1}{p}$$

准静态的多方过程 (polytropic process)

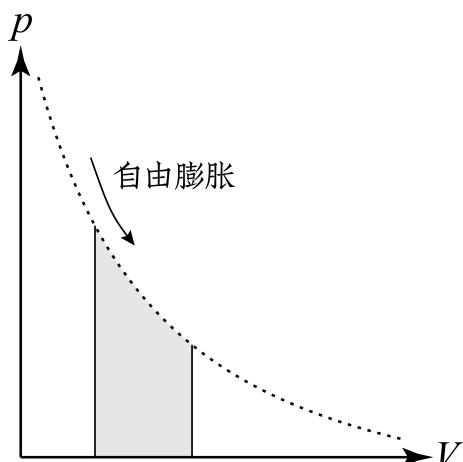
多方过程可以由过程方程 $pV^n = \text{Const.}$ 描述, 多方指数 $1 < n < \gamma$. 实际过程若发生缓慢, 可以近似处理为某种形式的多段多方过程. 多方过程的热容为:

$$C_{mn} = \left(\frac{1}{\gamma-1} - \frac{1}{n-1} \right) R$$

而其他响应函数 (指上面定义的 α, β 与 κ) 的形式与绝热过程算出来的类似, 只需要把 γ 改为 n . 多方过程在特殊的参数条件下即变成以上过程:

| 绝热指数 n | 摩尔热容 C_m | 描述 |
|------------------|--|--------------------|
| $n < 0$ | $\left(\frac{1}{\gamma-1} + \frac{1}{1-n} \right) R$ | 吸收大量热量, 压强体积温度同时增加 |
| $n = 0$ | C_{mp} | 等压过程 |
| $0 < n < 1$ | $\left(\frac{1}{\gamma-1} + \frac{1}{1-n} \right) R$ | 相比吸热等温膨胀, 温度有升高 |
| $n = 1$ | ∞ | 等温过程 |
| $1 < n < \gamma$ | $-\left(\frac{1}{n-1} - \frac{1}{\gamma-1} \right) R$ | 膨胀降温时要从外界吸热 |
| $n = \gamma$ | 0 | 绝热过程 |
| $n > \gamma$ | $\left(\frac{1}{\gamma-1} - \frac{1}{n-1} \right) R$ | 压缩气体同时给气体热量的过程 |
| $n = \infty$ | C_{mV} | 等温过程 |

自由膨胀过程

图 6.7: $p - V$ 图上画自由膨胀

理想气体向真空的自由膨胀过程(free expansion process)是一类典型的非准静态过程. 在这样一个过程中, 由于外压为零, 外界对气体做功为零. 而系统绝热, 从而从外界吸热也为零. 故由于热力学第一定律, 内能不变. 再根据理想气体内能直接依赖于温度的性质, 我们发现: 理想气体的自由膨胀是一个非准静态的等温膨胀.

从而在 $p - V$ 图上, 我们画一根等温线, 它是否代表向真空的自由膨胀过程? 我们需要画成虚线, 线上每一个点都表示平衡态, 而从一个点到另一个相邻的点中发生的过程则没有达到力学平衡, 从而不能视为准静态过程. 这一根虚线能够代表气体在自由膨胀过程中的状态, 因为即使不均匀, 整体的内能和可以确定出平均温度, 整体的体积和平均温度又可以确定出平均压强. 如果在某个时刻自由膨胀结束, 体系达到平衡态后其状态就落在了这根等温线上. 只不过, 这根线下面积此时也没有了实际意义, 因为气体对外做功应该是零.

向固定压强膨胀

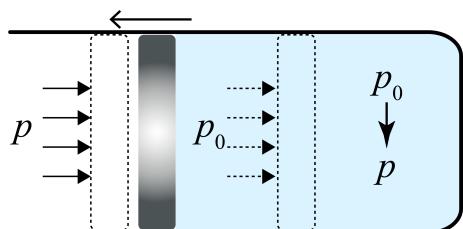


图 6.8: 向固定压强膨胀

考虑原平衡态压强为 p_0 的气体, 突然令外压变小为 p , 气体将发生某种膨胀, 如果膨胀稍过度, 其压强小于外界压强 p 时, 又会被压缩, 这样来回振动, 直到达新的平衡态. 如果这个过程系统绝热, 则系统对外界做功应该写为:

$$W = p(V - V_0)$$

这一个模型可以这样理解: 它保证了外界压强 p 的严格不变性. 内部气体由于发生了急剧的体积变化, 最靠近活塞的部分气体压强由于急剧的膨胀实际上就会等于外界气压 p . 这是忽略活塞质量(轻活塞)的情形. 若是考虑活塞的质量, 这个结论也还是成立, 因为内外气体对活塞做功之和转化为活塞动能. 而前后平衡态活塞实际上动能都是零. 故内部气体对外做功可以由外界固定压强部分气体的体积功来计算. 这就是上式. 那么同样, 由热力学第一定律和理想气体的性质, 我们有:

$$\frac{V}{V_0} = 1 + \frac{p_0 - p}{\gamma p} ; \quad \frac{T}{T_0} = \frac{1}{\gamma} + \left(1 - \frac{1}{\gamma}\right) \frac{p}{p_0}$$

如果我们在初态压强 p_0 与末态压强 p 之间搭设很多的压强平台 p_i , 使得外压是缓慢地从 p_0 逐级减小到 p , 那么用极限的方法容易证明(请读者自己完成), 过程恰好变为准静态的绝热过程:

$$pV^\gamma = p_0V_0^\gamma ; \quad \frac{p^{\gamma-1}}{T^\gamma} = \frac{p_0^{\gamma-1}}{T_0^\gamma}$$

此时屡次的非平衡态将无限趋于平衡态, 而非准静态过程也就无限趋于准静态过程了.

节流过程 (throttling)

节流过程亦是对很多复杂系统过程的抽象,如冷凝机内部工作物质所发生的循环的定常流动. 我们从中提取出一个模型, 如下图, 气体在内壁光滑的管道内做定常流动, 通过一多孔塞后压强减小, 气体膨胀. 整个过程绝热. 我们取一个固定的气团 (图中红加蓝) 作为研究对象. 经过一定时间 t 后左侧与右侧体积变化分别为 V_1 与 V_2 , 那么气体的流量被定义为:

$$Q = \frac{m_i}{t} = \frac{\rho_i V_i}{t} = \rho_i v_i A$$

由于是定常流动, 左右两侧流量必须相等, 这是质量守恒的要求, 而体积流量 $v_i A$ 可以不相等, 因为气体可以被压缩.

下面我们考虑气体的热力学第一定律, 由于 Q 小, 我们忽略气体的整体动能 (这一点在下一节我们将做详细讨论).

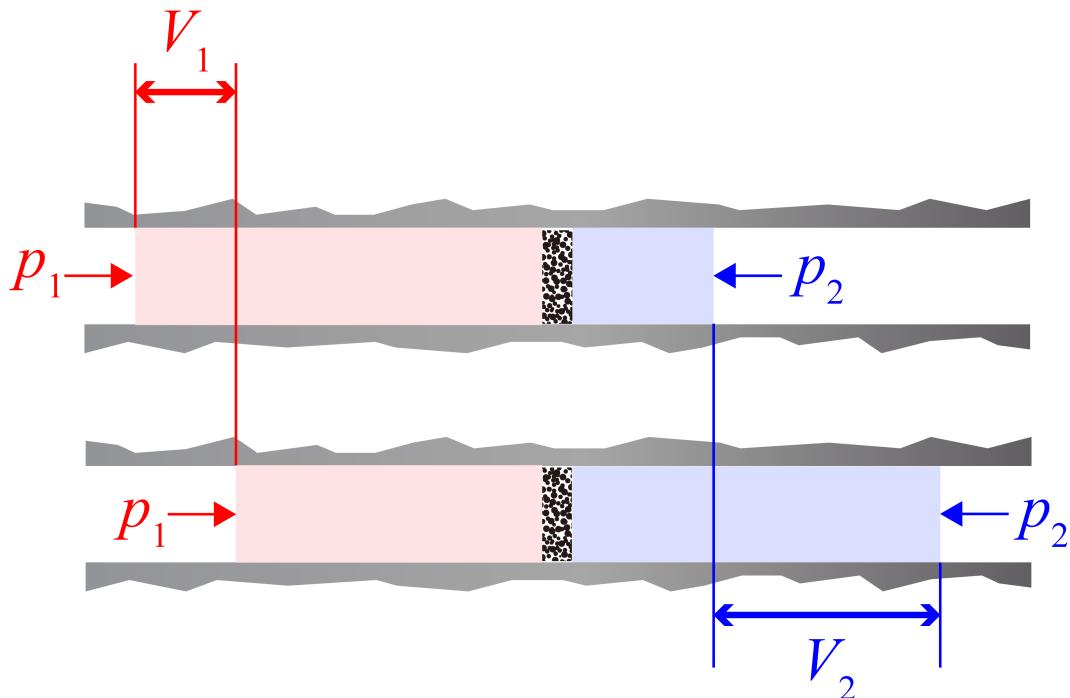


图 6.9: 多孔塞制节流阀

对于红 + 蓝的整体, 热一定律表述为:

$$p_1 V_1 - p_2 V_2 = U_2 - U_1$$

即:

$$U_1 + p_1 V_1 = U_2 + p_2 V_2 ; \quad H_1 = H_2$$

也就是说, 节流过程前后焓不变的过程. 若流过的气体是理想气体, 那么

$$H = \nu C_{mV} T + pV = \nu(C_{mV} + R)T = \nu C_{mp} T$$

可见这也是一个温度不变的过程. 但是吸热, 做功都为零, 它并不是一个达到平衡的准静态过程.

6.4 开放系统的理想气体

开放的气体系统一般有两个特点,一是其性质一般被描述为:在某处 r 的气体性质怎样.而描述该气体性质的只能是各种强度量.如压强 $p(r)$,温度 $T(r)$ 等等.二是气体往往处于动态的物质,能量的交换与流动中.这就意味着我们分析其物理规律时必须单独取出一个气团(parcel of air) 作为研究对象.研究周围气体对它可能造成的微观扩散,黏滞,传热过程,研究周围气体对气团压力造成的体积功与加速过程.下面我们由简入难地分析这些问题.

6.4.1 静态平衡问题——重力场中的大气

对于静态平衡的热力学体系,气团首先要符合的是静力学平衡条件:

$$\nabla p + f = 0$$

其中 f 代表体积力,我们考虑重力场系统,则为:

$$\nabla p + \rho g = 0$$

而密度反过来又同时和压强,温度有关:

$$\rho = \frac{\mu p}{RT}$$

于是,压强的分布纯粹由温度分布决定,而温度分布则看似成为了一个纯粹的热学问题,而且在一定范围内粗略地研究我们可以简单地假设温度随高度均匀地变化 $T(z) = T(0) - \Gamma z$.然而这个问题并不简单,大体上如果讨论限于地球表面对流层的大气,其基本图像可以视为由太阳照射地表,地表升温而加热大气导致的.而对于不同的气候,地理环境,时间与天气条件这个温度随高度递减的方式都十分地不相同,我们也还是定义环境温度递减率(environmental lapse rate) 来线性表示温度的降低:

$$\Gamma = -\frac{dT}{dz} ; \quad T(z) = T(0) - \Gamma z$$

了解这一点后便可计算出其压强分布:

$$p(z) = p(0) \cdot [1 - \frac{\Gamma z}{T(0)}]^{\frac{\mu g}{R}} = p(0) \cdot [\frac{T(z)}{T(0)}]^{\frac{\mu g}{R}}$$

当环境递减率取某些不同的值时大气的性质是不同的,我们考虑其对流性质.考虑一个干的气团,让它在垂直方向上发生微扰,认为力学平衡总是很快地达到——气团压强始终等于外大气压.但热力学平衡总是来不及达到——既使气团温度与外大气不相等,它与外界的热量交换可以忽略,从而可以发现,气团做绝热碰撞:

$$p(z) = p(0) \cdot [\frac{T'(z)}{T(0)}]^{\frac{\gamma}{\gamma-1}}$$

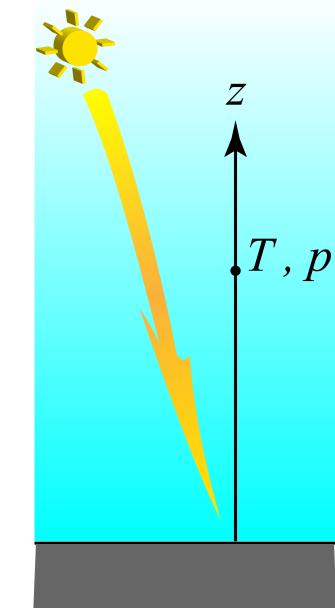


图 6.10: 地球大气

比对气团温度 $T'(z)$ 与外大气温度 $T(z)$,我们发现有一临界的环境温度递减率(dry adiabatic lapse rate),它表征干燥的气团发生微扰后温度,密度恰好与周围大气相等的情况:

$$\Gamma_d = \frac{\gamma - 1}{\gamma} \frac{\mu g}{R} = \frac{\mu g}{C_{mp}} = 9.8^{\circ}\text{C}/\text{km}$$

若 $\Gamma > \Gamma_d$,则环境递减率过高,向上微扰的气团温度高于环境温度,所受浮力偏大,将偏离平衡位置,产生热对流,成云致雨.反之, $\Gamma < \Gamma_d$ 时浮力偏小,气团受到回复力而稳定在平衡位置处,对应着稳定无对流的大气.对于地球大気体系,有对流与无对流的情况都有时发生,其平均递减率就在临界绝热递减率附近.

对于临界大气为何温度随高度线性递减,而递减率又与定压热容有直接的关联,有一种较为直接的证明方法.其思想为虚功原理结合下面经常使用的流管分析法.我们让从 $z = 0$ 到 $z = h$ 的截面积为 A 的气柱在顶部与底部的压力作用下向上发生虚位移,由于是临界平衡,故位移以后恰巧每一小块气柱 Adz 都与新的位置的环境大气性质全同,也就是与之前在这儿的那块气柱性质一样.整体上看,相当于把 $z = 0$ 处的一块质量为 δm 的气团位移到了 $z = h$ 处.写出热力学第一定律,要注意其本质为能量守恒,这里还要包括重力势能:

$$p(0)\delta V(0) - p(h)\delta V(h) = \delta U(h) - \delta U(0) + \delta mgh$$

而由焓的定义 $\delta H = \delta U + p\delta V$, 以及理想气体的焓有公式 $\delta H = \delta\nu C_{mp}T$, 我们得出:

$$\delta\nu C_{mp}[T(0) - T(h)] = \delta\nu\mu gh$$

即:

$$T(h) = T(0) - \frac{\mu g}{C_{mp}}h$$

对于实际地球对流层大气来说,还需要考虑空气中水蒸气受冷造成的液化放热现象,其临界递减率将会取决于气体含水蒸气的多少.而地球下层大气还分为对流层(troposphere) 与平流层(stratosphere),平流层含大量臭氧,太阳的紫外辐射将在传播的过程中被平流层逐渐吸收.所以平流层的上部具有较强的辐射和较高的平衡温度,而温度随高度递增,提供了一个稳定的大气环境,极难发生对流.

6.4.2 能量守恒——伯努利方程

我们把流体力学里讨论的伯努利方程(Bernoulli Equation) 推广到热力学中. 它适用条件为定常流动(steady flow) 中的同一条流线(streamline) 上. 它也是推广的热力学第一定律, 它代表能量的守恒.

不可压缩流体的伯努利方程的图像其实很清晰, 我们若不考虑流体势能的变化, 那么伯努利原理实际是在说, 两侧的压强差导致了流体的加速:

$$p_1 - p_2 = \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2$$

从而沿流线, 以下量守恒:

$$p + \frac{1}{2}\rho v^2 = \text{const.}$$

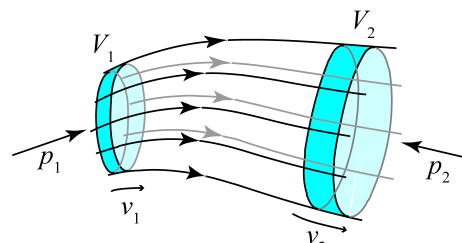


图 6.11: 气体沿一条流管流动

但若考虑可以压缩的气体, 以及考虑保守外力的做功, 是否有以上守恒量呢? 我们假设保守外力为质量力: 单位质量的流体的势能为 Φ , 那么对于一定时间间隔 t 下的流管发生的流动 $m = Qt$, 有以下能量守恒:

$$p_1 V_1 - p_2 V_2 = (U_2 + \frac{1}{2}mv_2^2 + m\Phi_2) - (U_1 + \frac{1}{2}mv_1^2 + m\Phi_1)$$

两边同时除以质量, 得守恒律:

$$u + \frac{p}{\rho} + \frac{v^2}{2} + \Phi = \text{const.}$$

其中 $u + p/\rho = u + pv = h$ 即为比焓, 而 u 为比内能, v 为比体积, “比” 表示单位质量的某物理量, 它是一种强度量. 从而以上量我们还可以写为:

$$h + \frac{v^2}{2} + \Phi = \text{const.}$$

我们还定义所谓的阻滞焓(stagnation enthalpy) 的概念. 它表示让运动的气体发生某种虚构的定常流动绝热地停下来后气体应该有的焓. 那么比阻滞焓为:

$$h' = h + \frac{v^2}{2}$$

最后这一广义的伯努利方程,就被我们写为一个广义的两项的能量守恒形式:

$$h' + \Phi = \text{const.}$$

那么对于理想气体,我们还可以定义阻滞温度:

$$T' = T + \frac{\mu v^2}{2C_{mp}} ; \quad h' = \frac{1}{\mu} C_{mp} T'$$

而对于普通不可压缩流体,我们一般不关心其热学性质,故焓可以扣除内能项 $h = p/\rho + v^2/2$,而 ρ 为常数.这样我们又可以定义阻滞压强:

$$p' = p + \frac{1}{2}\rho v^2 ; \quad h' = \frac{p'}{\rho}$$

6.4.3 动量守恒——欧拉方程

让我们考虑管道内的理想气体流动,管道的截面积在不断减小的情况 $A = A(x)$.可以想见,这会导致流速的增加,进一步导致压强的减小,对应方程为**连续性方程**(continuity equation)与伯努利方程,分别对应质量与能量守恒:

$$\rho v A = \text{const.}$$

$$\frac{1}{\mu} C_{mp} T + \frac{v^2}{2} = \text{const.}$$

以上方程还应联系物态方程理解:

$$\rho = \frac{\mu p}{RT}$$

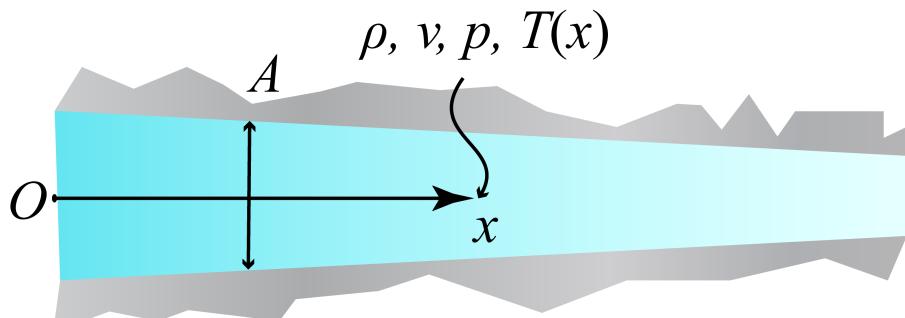


图 6.12: 截面积减小的管道内的气流

注意到我们把 A 的变化当做已知量以后,却仍设了待求解的四个物理量 $v(x), p(x), T(x), \rho(x)$,但目前却只有三个方程.那么我们漏了怎样一个方程呢?这个方程就是流体的**欧拉方程**(Euler equation),它是一个纯粹的动力学方程,不包含任何热效应.一般写作:

$$\frac{D\mathbf{v}}{Dt} = -\frac{\nabla p}{\rho} - \nabla \Phi$$

其中 D 代表**随体导数**(material derivative):

$$\frac{D}{Dt} = \frac{d}{dt} + \mathbf{v} \cdot \nabla$$

上式代表了压强梯度力与保守外力对流体的加速效应,在上述情形中没有外力所以等式右边只有一项.实际上,它代表了整个体系动量守恒.而对于上述情形,气体做定常流动,上式写为:

$$v \frac{dv}{dx} = -\frac{1}{\rho} \frac{dp}{dx}$$

这一个式子也可以由上一节的类似方法——取一定长度的气团来导出. 但切不可忽略管壁对气体在过程中作用的冲量:

$$p_1 A_1 \mathbf{e}_x \cdot \mathbf{t} - p_2 A_2 \mathbf{e}_x \cdot \mathbf{t} + \int -pdA \cdot \mathbf{t} = m\mathbf{v}_2 - m\mathbf{v}_1$$

左边恰巧是在整个表面的压强的面积分, 可以由高斯定理化为体积内的体积分:

$$\int_{x_1}^{x_2} -\frac{dp}{dx} \cdot Adx = Q(v_2 - v_1)$$

又考虑到以上积分式无法直接积出结果, 故对积分上限求导, 再考虑到 $Q = \rho v A$, 化简后即得欧拉方程:

$$\frac{dp}{dx} = -\rho v \frac{dv}{dx}$$

对于管道截面积相对于长度变化缓慢的情况¹⁶, 联立以上四个方程, 解得:

$$\begin{aligned}\frac{dv}{v} &= -\frac{dA}{A} \left(1 - \frac{\mu v^2}{\gamma RT}\right)^{-1} \\ dp &= \rho v^2 \cdot \frac{dA}{A} \left(1 - \frac{\mu v^2}{\gamma RT}\right)^{-1} \\ dT &= \frac{\mu v^2}{C_{mp}} \cdot \frac{dA}{A} \left(1 - \frac{\mu v^2}{\gamma RT}\right)^{-1}\end{aligned}$$

可见, 随着截面积 A 的减小, 流速加快, 压强和温度都要减小, 利用这个效应可以制作纯粹依靠风力制冷的“土空调”. 注意到以上结论还依赖于一个条件:

$$\frac{\mu v^2}{\gamma RT} < 1$$

这意味着分子代表的气团集体运动能量要远小于分母代表的微观热运动能量, 这在一般情况下是很容易满足的, 因为 $v \sim 10\text{m/s}$ 而热运动平均速率 $\bar{v} \simeq 500\text{m/s}$, 若对于高速运动的气体, 其发生的过程中气团速度的改变与热运动平均速度可以比拟, 实际上把气团处理为平衡的理想气体也是不妥当的.

¹⁶即, 在所关心的范围内截面积的相对变化很小:

$$\frac{\Delta A}{A} \ll 1$$

第七章 热力学第二定律

热力学第一定律固然强大,它根本上杜绝了所谓的**第一类永动机**(perpetual motion machine of the first kind)的可能性,也排除了一大类不现实的热力学过程的发生.然而由这样一个只关心整体性质,不深入到微观本质的热力学规律又能告诉我们多少呢?**炼金术**(alchemy)不违背热力学第一定律,它企图实现物质与物质之间的随意转化,**第二类永动机**(perpetual motion machine of the second kind)亦不违背热力学第一定律,它基于能量守恒,主张同一份能源的无限再生利用.对于前者,微观机理彻底地否定了它,不同的原子有不同的构成,原子核的结构十分稳定,至今无法自由操控.而对于后者,统计规律也彻底地进行了否定,它告诉我们,一个体系的过程总是有确定的方向性,在一定条件下,过程被我们分类为**自发过程**(spontaneous process)与**不自发过程**(non-spontaneous process),具有特定方向性的过程一旦发生,其时间反演的过程就将不现实.但十分神奇的一点是,不同体系的不可逆特性居然是紧密相关的,若将一个过程的不可逆性与另一个不可逆过程的不可逆性建立联系,我们就能从中同时发现统一的热力学温标与态函数熵这两个概念,从而量化过程的不可逆性,这一套宏观的规律称为热力学第二定律.它道出了热力学过程的方向性,本质是微观统计规律,限制了另一大类不现实的热力学过程的发生.

7.1 循环过程

在历史上,热力学基本规律的确立与发展和热机的使用与改进是紧密相关的.直到今天卡诺循环的引入仍然不失为介绍热力学第二定律理论体系的强有力工具.本节将介绍热机,热泵的基本模型与著名的卡诺循环.为之后引入热力学第二定律相关概念打下基础.

7.1.1 热机与热泵

上一章介绍过循环过程的概念.我们介绍了 $p - V$ 系统的正负循环.现在我们把这样的 $p - V$ 体系看成很多实际生活中一些机器的工作过程的抽象.其实际承担了状态循环变化的体系称为**工作物质**(working body).而其联动的**机械装置**(mechanical device)可以与外界交换功.最后,这样的体系还需要可以与其交换热量的对象,凡是工作物质从中吸热的被称为**热源**(heat source),而工作物质向之放热的被称为**冷库**(cold sink).这样我们就抽象出了一个理论模型,根据热力学第一定律,如果做正循环,工作物质一个循环将从高温热源吸收更多的热量,向低温冷库放更少的热量,而通过机械装置输出功,故被称为**动力循环**(power cycle),对应机器为**热机**(heat engine).而如果考虑相反的过程¹,工作物质一个循环向高温热源放热,从低温冷库吸热,机械装置要对它做功,则是一个**泵热循环**(heat pump cycle).对应机器被称为**热泵**(heat pump).

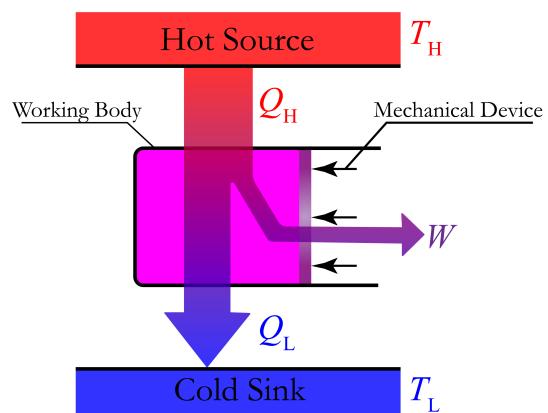


图 7.1: 热机模型

¹这样的过程不一定可逆,这里指的是那些定性现象相反的循环.



图 7.2: 俄罗斯巴拉科沃 (Balakovo) 核电站汽轮机

根据以上原理,热机主要分为以下两大类:

蒸汽机 (steam engine) 蒸汽机是一类外燃机(external combustion engines), 工作物质是在气相与液相混合的水 (水蒸气). 用外界燃烧的热量, 或者来自太阳能, 核能, 地热等的非燃烧热量加热锅炉 (boiler) 中的水, 由此产生高温高压蒸汽, 用蒸汽带动机械装置运转向外输出能量. 经过这样的装置后 (这中间一般视为节流过程), 水蒸气一般变为过饱和的冷气体或者还夹杂着水滴, 进一步导入冷凝器而液化变为液态水回收进入锅炉. 完成一个循环. 18 世纪末与大半个 19 世纪可谓“蒸汽时代”, 蒸汽机在大型机械生产工厂, 蒸汽机车, 蒸汽机轮船等交通工具中大放异彩. 虽然石油开采为小型汽车生产提供了可能, 大型交通工具和生产业动力来源之后也慢慢地几乎被内燃机与电力所取代. 但直到今天, 我们的生产生活有 90% 以上的电量都来自于汽轮机(steam turbine) 形式的热电厂. 值得一提的是其中化石能源 (煤电, 气电, 以及较为小规模的油电) 仍然占据 75% 的高比重 (中国与美国的情况大致相同), 新型能源的比重不大, 但在稳步上升中.

内燃机 (internal combustion machine) 蒸汽机的锅炉设计让燃烧能量给了流动的工作物质, 实现了工作物质不含燃料, 不与燃料混合或被燃料污染. 这有效地提高了燃料能量利用的效率. 然而却需要为工作物质准备额外的空间, 不利于把热机做小. 与之相反, 内燃机利用了燃烧室(combustion chamber) 的设计, 工作物质直接是高能量密度的燃料和氧化剂 (一般就是空气). 利用高温高压下点燃化学反应释放的巨大化学能来产生热, 导致气体反向对机械装置——活塞(piston) 或涡轮(rotor) 做功. 与外燃机工作物质发生的真循环过程不同, 内燃机发生的过程是一种开放循环. 以常规的四冲程活塞式内燃机为例, 气体经历四个冲程: 吸入(intake), 压缩(compression), 做功(power), 排出(exhaust). 整个循环从常态燃料中开始, 最后以排出气体在大气中冷却结束. 而燃气涡轮发动机(gas turbine engine) 更是针对四个阶段分别设计功能区, 使得四个阶段同时进行, 提供了极大的功率输出. 按照其机械装置的不同分为四类: 军用战斗机往往采用涡喷发动机(turbojet engine), 而民航客机往往是涡扇发动机(turbofan engine); 运输机, 巡逻机经常采用涡桨发动机(turboprop engine); 而直升机和其他陆上, 海上交通工具则通过涡杆发动机(turboshaft engine) 将动力用传动轴导出. 在这样的设计中, 燃料产生的能量用来

驱动涡轮带动整个装置大量吸入外界空气²并有效地压缩气体从而产生动力. 涡轮转动的能量也正是很多民航飞机上电力的来源.

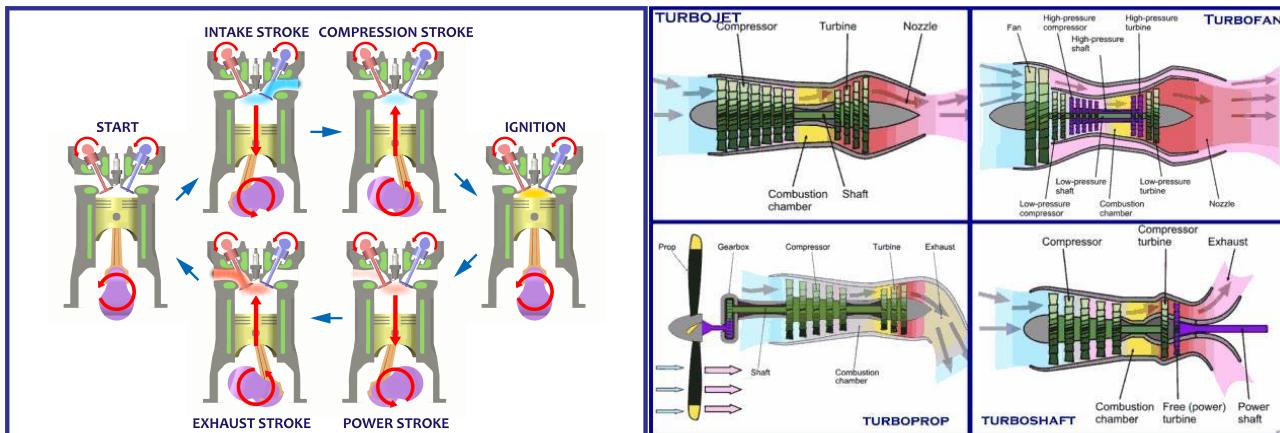


图 7.3: 左: 四冲程活塞内燃机; 右: 四类涡轮机

我们发现, 将实际的热机抽象为前图 7.1 那样的模型往往是过分地简单了. 实际问题中的工作物质往往成分在发生改变, 有物质的输入输出, 甚至无法视为一个完整的循环过程. 但理论可以帮我们分析与把握这个过程的一些主要信息. 在实际问题中热机的能效(energy efficiency) 是主要的关心因素之一, 它定义为输出的有用功 W 与燃料提供的化学能 Q_H 之比:

$$\eta = \frac{W}{Q_H}$$

而热泵也根据其设计功能分为制冷机(cooling machine), 制热机(heating machine) 与可逆机(reversible machine). 可逆的热泵同时具备制冷制热两种功能, 采用换向阀来控制气流的流向从而改变效果. 当采用制热模式时, 其效率往往比使用单独的电阻加热高 3 到 4 倍.

无论何种类型的热泵, 其构成与原理可以被抽象为典型的四步. 高温热源处更高温的气体在冷凝器(condenser) 处变为液态而放热; 膨胀阀(expansion valve) 处的节流过程使液态流体温度与压强大降而成为气液混合的状态; 低温冷库处稀薄的半液态流体在蒸发器处蒸发吸热; 最后通过压缩机(compressor) 泵入冷凝管, 压强与温度显著升高. 这样的循环称为斯特林循环(Stirling cycle).

而对于热泵的性能系数(coefficient of performance), 需要对两种不同的模式区分:

对于制冷模式, 取低温冷库处吸热比输入功率:

$$\eta = \frac{Q_L}{W}$$

对于制热模式, 取高温热源处放热比输入功率:

$$\eta = \frac{Q_H}{W}$$

两者都可能大于一.

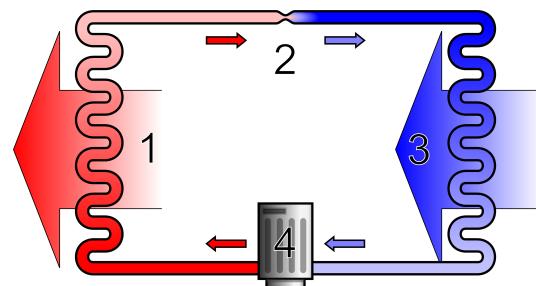


图 7.4: 热泵的普遍构成

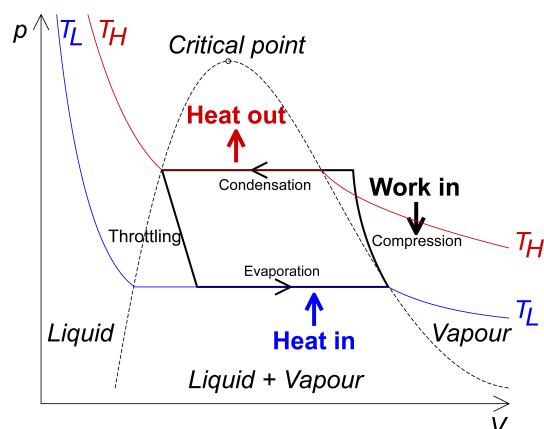


图 7.5: 热泵的工作循环

²火箭除外, 它需要自备燃料与氧化剂.

7.1.2 热机循环

从瓦特 (Watt) 改良蒸汽机开始,一代代的工程师开始致力于提高热机的效率.这其中便产生了大量实用的热机模型.直到 1824 年法国物理学家与工程师卡诺 (Carnot) 给出了惊人的卡洛热机模型与关于热机工作效率上限的论断.这实际上意味着热力学(thermodynamics) 的诞生.之后热机优化建立在严格的热力学计算的基础上.发展到今天理论早已成熟,只待技术的发展与革新.

简单的热机循环计算往往将工作物质视为理想气体在 p - V 图上的准静态过程构成的循环.下将一一列举:

勒鲁瓦循环 (Lenoir cycle)

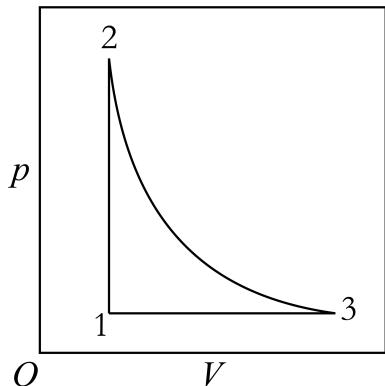


图 7.6: Lenoir cycle

1860 年法国工程师勒鲁瓦发明的勒鲁瓦机是第一个为商业设计的内燃机.其对应循环常常用来描述脉冲喷气发动机(pulse jet engine)的模型.它区别于其他热机的特点是无需压缩气体.一个循环内,理想气经历:

- i 1-2: 定容爆炸吸热
- ii 2-3: 绝热膨胀
- iii 3-1: 定压放热

定义爆炸比(explosion ratio) r_p :

$$r_p = \frac{p_2}{p_1}$$

则可以计算出其效率为:

$$\eta = 1 - \gamma \frac{T_3 - T_1}{T_2 - T_1} = 1 - \gamma \frac{r_p^{\frac{1}{\gamma}} - 1}{r_p - 1}$$

奥托循环 (Otto cycle)

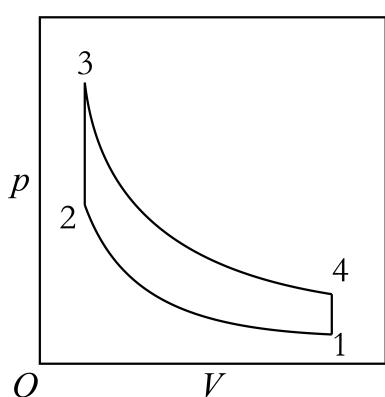


图 7.7: Otto cycle

这是最典型的四冲程型汽油机(gasoline engine, gas engine)的工作循环.特点是通过压缩气体产生高温高压易燃汽油空气混合体,再通过火花塞点燃混合体,过程可以视为瞬间发生的,从而体积不变,压强突然增大.一个循环内,理想气经历:

- i 1-2: 绝热压缩
- ii 2-3: 定容爆炸吸热
- iii 3-4: 绝热膨胀
- iv 4-1: 定容放热

定义压缩比(compression ratio) r :

$$r = \frac{V_1}{V_2}$$

则可以计算出其效率为:

$$\eta = 1 - \frac{T_1}{T_2} = 1 - \frac{T_4}{T_3} = 1 - r^{-(\gamma-1)}$$

对于空气与燃油的混合气 $\gamma \approx 1.3$,由于爆炸燃烧机制,目前的压缩比能做到 $r \approx 10$,这样看其理论效率上限 $\sim 50\%$,目前一般的汽油机效率为 $25\% \sim 35\%$.最高能到 40% ,离理论上限还尚有一段提升空间.

迪塞尔循环 (Diesel cycle)

迪塞尔机又称为**柴油机**(Diesel fuel engine). 其与汽油机设计上存在区别. 为了提高汽油机的效率, 柴油机企图增大压缩比, 故采用粗重的活塞. 并取消汽油机中点火爆炸的过程, 而让压缩冲程的末温直接使得燃料能够燃烧, 而缓慢注入燃料以使得燃烧充分. 这样, 典型的迪塞尔循环可以认为是这样四个过程:

i 1-2: 绝热压缩

ii 2-3: 定压燃烧吸热

iii 3-4: 绝热膨胀

iv 4-1: 定容放热

定义压缩比 r 与**停气比**(cut off ratio) α :

$$r = \frac{V_1}{V_2} ; \quad \alpha = \frac{V_3}{V_2}$$

则可以计算出其效率为:

$$\eta = 1 - \frac{1}{\gamma} \cdot \frac{T_4 - T_1}{T_3 - T_2} = 1 - \frac{\alpha^\gamma - 1}{\gamma r^{\gamma-1} (\alpha - 1)}$$

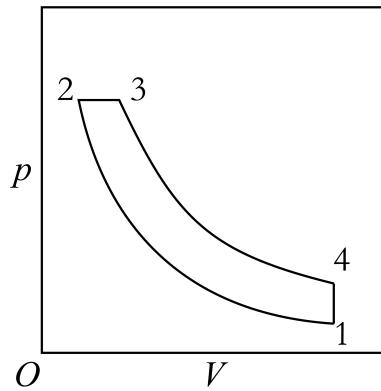


图 7.8: Diesel cycle

布莱顿循环 (Brayton cycle)

涡轮机等开放式热力学循环的一般抽象是带有两个等压过程的布莱顿循环. 可以认为是这样四个过程:

i 1-2: 绝热压缩

ii 2-3: 定压燃烧吸热

iii 3-4: 绝热膨胀

iv 4-1: 定压放热

定义**压强比**(pressure ratio) r_p :

$$r_p = \frac{p_2}{p_1}$$

则可以计算出其效率为:

$$\eta = 1 - \frac{T_1}{T_2} = 1 - r_p^{-\frac{\gamma-1}{\gamma}}$$

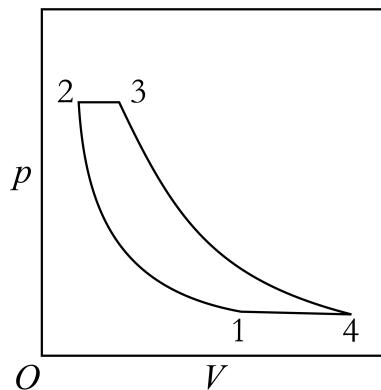


图 7.9: Brayton cycle

卡诺循环 (Carnot cycle)

卡诺循环是一种具有极其重要的理论意义的理想循环。它不能对应实际应用与生活与生产的实际热机。但是却像直线与平面之于几何理论那样,对卡洛循环与热机的很多讨论都将与热力学理论体系紧密相关。对于理想气体,卡诺循环意味着其吸放热都发生在与两个恒温热源做等温接触的过程中。

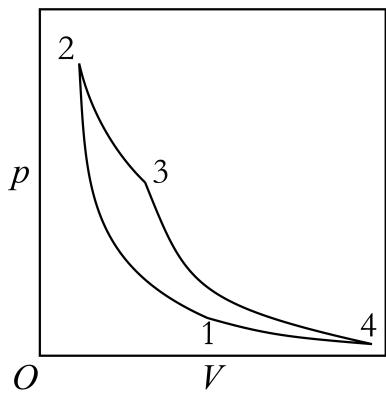


图 7.10: Carnot cycle

整个循环由两条绝热线与两条等温线围成。过程如下:

i 1-2: 绝热压缩

ii 2-3: 等温吸热

iii 3-4: 绝热膨胀

iv 4-1: 等温放热

高温热源的温度为 T_H , 低温冷库的温度为 T_L . 那么可以计算出卡诺热机的效率为:

$$\eta = 1 - \frac{T_L}{T_H}$$

7.2 理想气体的熵

理想气体的准静态过程是这一章内我们关心的问题。如何判断一个微元准静态过程中理想气体吸热还是放热? 具体来说, 我们关心在 (T, V) 点沿着 (dT, dV) 变化方向改变状态所需要吸收的热量微分 dQ 。由热一, 它应为:

$$dQ = dU + pdV = \nu C_{mV} dT + \frac{\nu RT}{V} dV$$

这可以进行如下的化简:

$$dQ = \nu C_{mV} T \cdot \left[\frac{dT}{T} + (\gamma - 1) \frac{dV}{V} \right] = \nu C_{mV} T \cdot d \ln(TV^{\gamma-1})$$

我们定义态函数——**熵**(entropy):

$$S_1 = \nu C_{mV} \ln(TV^{\gamma-1})$$

那么自然得出:

$$dQ = T dS_1$$

这给出了气体吸放热与状态变化之间的联系。一个直接的推论就是, 准静态绝热过程中 $TV^{\gamma-1}$ 是不变的(绝热方程)。或者说, 准静态绝热过程可以被称为**等熵过程**(isentropic process)。而熵作为一个态函数, 等熵线就是绝热线。

熵在定义时可以相差一个与状态无关的常数(但与摩尔数 ν 有关, 目前它不随着状态变化而变化)。我们为了方便暂时不写这个常数。但如果一开始以 (p, V) 或者 (p, T) 作为状态参量。那么自然而然写出来的符合 $dQ = T dS$ 的熵为:

$$S_2 = \nu C_{mV} \ln(pV^\gamma) \quad ; \quad S_3 = \nu C_{mV} \ln \frac{T^\gamma}{p^{\gamma-1}}$$

它们都相差一个过程不变的常数:

$$S = S_3 \quad ; \quad S_1 = S + \nu R \ln(\nu R) \quad ; \quad S_2 = S + \nu C_{mp} \ln(\nu R)$$

我们把 S_3 取为标准的熵。原因在于, 如果我们要讨论摩尔数 ν 变化的情况时。在对数里放两个强度量可以使表达式成为一个广延量——当摩尔数增大到 λ 倍而所有强度量不变时, 熵 S 也增大到 λ 倍。而其他两个 S 附加的项 $\nu R \ln(\nu R)$ 与 $\nu C_{mp} \ln(\nu R)$ 使得定义失去了这种广延性。

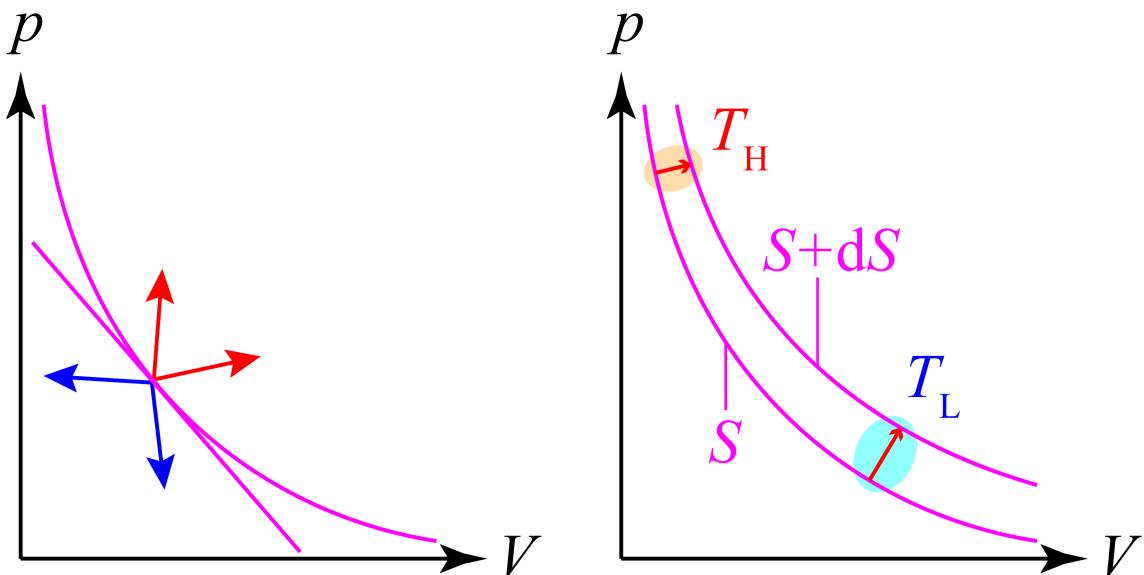


图 7.11: 左: 吸放热的元过程方向; 右: 两根等熵线间的“热量差”

我们由此得出这样的结论: 在一个准静态过程中. 沿着等熵线以下的方向熵降低 (如图7.11中蓝色箭头方向), 从而要放热. 而往上走则熵增加, 要吸热. 但注意到既使沿着切线方向走熵还是会降低 (无论往上还是往下), 只不过这时候其降低是一个二阶小量而已. 也就是说如果气体历经 $p - V$ 图上的负斜率直线过程, 一定是先吸热, 后放热. 在中间有一个熵的极值点符合:

$$k = \frac{dp}{dV} = -\gamma \frac{p}{V}$$

我们画出两条相隔很近的等熵线 (差 dS). 注意到两条线之间的“热量差”这个概念本身不够严谨, 因为热量是过程量不是状态量. 但由于两条等熵线相隔很近, 所以从一点出发到另一条绝热线上的点还在这一点附近 (微元过程, 取 $\Delta S \rightarrow 0$ 的极限). 所以可以认为温度近似不变. 尽管这样, 吸热仍然与过程发生的平均温度 T 有关:

$$Q_H = T_H dS \quad ; \quad Q_L = T_L dS$$

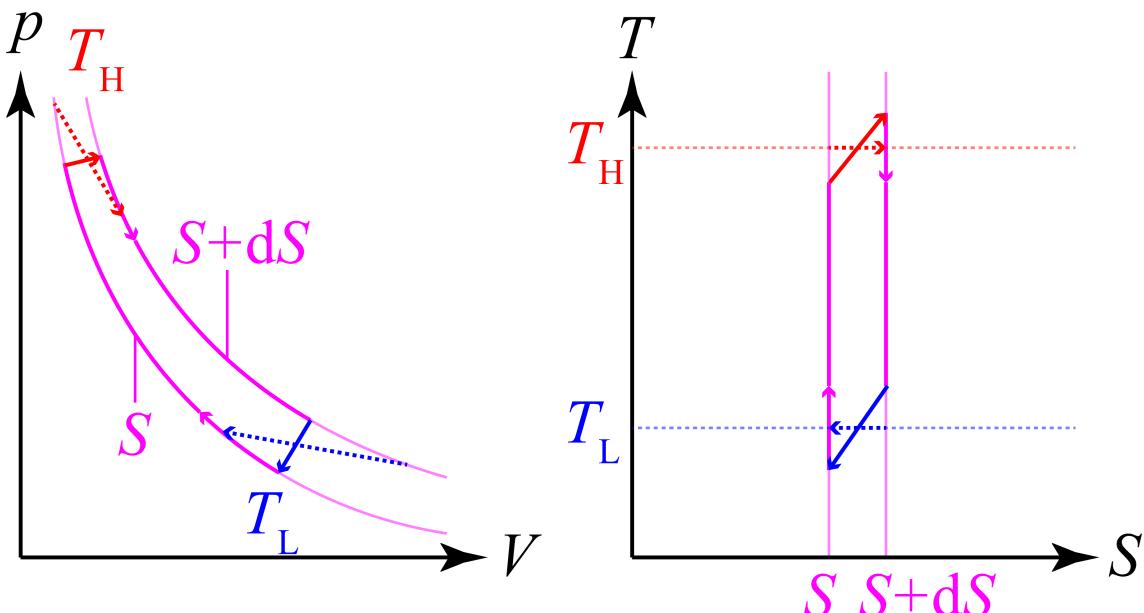


图 7.12: 示功图与示热图

由于 S 是态函数, 我们也可以选取 (T, S) 来作为体系的状态参量. 而做 $T - S$ 图表示状态与过程. 此时等温线实际上就是水平线, 而绝热线就是竖直线. 而其包围的面积实际上就是整个过程所吸收的热量:

$$Q = \oint T dS$$

而由于热一, 它一定等于 $p - V$ 图中的面积, 它表示对外做功:

$$W = \oint p dV$$

正是由于这个直观的关系, 热力学理论分析时我们经常对照两个图像. 分别叫做示功图 (work indicator diagram) 与示热图 (heat indicator diagram).

还有两点值得讨论:

第一, 以上讨论仅仅限于理想气体的准静态过程. 我们发现对于理想气体, 如果采用理想气体温标 (即分子内能恰正比于温度), 恰好可以引入独特的态函数熵来计算其吸热. 然而, 对于非准静态过程初末态之间的联系我们还一无所知, 仅仅知道热一总是成立的.

第二, 对于其他体系是否有 $dQ = T dS$? 这看上去似乎不显然. 首先我们知道热力学第零定律定义了温度是某种热平衡的系统共同具有的属性, 但我们一直使用的温度似乎是根据理想气体的分子平均动能定义出来的 (理想气体温标):

这样, 直接把这个温度用于其他体系似乎不太可能会得出相似的结论, 而另外定义新的温标又会造成物理规律的不统一性. 温度也就失去了作为互为热平衡态共同参数的性质. 然而, 需要知道上式不是无来由的, 它来源于麦克斯韦分布律——微观体系的经典动力学与统计规律的共同结果, 在这一点上其实温度的定义是普适的, 它都反映了各个自由度上热运动的剧烈程度. 所以实际上存在普适的热力学温标, 而同样的也将存在对应的普适的态函数熵. 它们都由大量微观粒子的统计规律给出. 然而十分神奇的一点是, 普适的温度与熵的存在性在历史上和逻辑上的确是可以由简单的公理得出的热力学第二定律的体系. 这正是下一节我们要说明的. 而要深刻了解温度与熵的微观本质需要深入了解统计力学方法, 我们在下一章将给出相关的论述.

$$T = \frac{2\bar{\varepsilon}_k}{3k}$$

7.3 热力学第二定律

费曼说过, 所有自然现象的最自然特征莫过于它们明显的不可逆性. 大量微观粒子构成的宏观热力学体系在种种现象上体现出不可逆的特性, 即热力学时间箭头 (time arrow of thermodynamics). 如果令时间反演, 系统的各状态逆序发生, 外界对其做功变为对外做功, 向外界放热变为从外界吸热. 那么很多现象都给出了从未被观察到的结果: 混合均匀地两种气体自发地分开, 低温冷库向高温热源自发放热, 热能单一地转化为机械能等等. 无耗散的准静态过程是理想的可逆过程 (reversible process), 仅有系统内部处处达到细致平衡, 在与外界平衡的状态下统一地改变其状态才能将其过程反演而符合物理规律. 而广泛存在的不可逆过程 (irreversible process), 一方面在于过程反演的不实际性, 还在于如果我们把系统和外界各部分视为一个孤立的体系, 那么发生的不可逆过程无法通过任何可能的过程回到初始状态. 在发生不可逆过程的时候, 体系的某些特征沿着一个方向被永久而不可恢复的改变了, 衡量这个特征的物理量就是熵, 而这种断言就是热力学第二定律. 它回答了热一所不能回答的问题, 给出了热力学过程发生的方向性与限度. 对于这些问题的讨论是热二的特色.

热力学第二定律在逻辑体系上承接热力学第零, 第一定律, 我们现在取消前面已经有良好定义的理想气体与针对它的温标的概念. 而热机模型与其效率只取决于吸放热仍然有良好的定义:

$$\eta = 1 - \frac{Q_L}{Q_H}$$

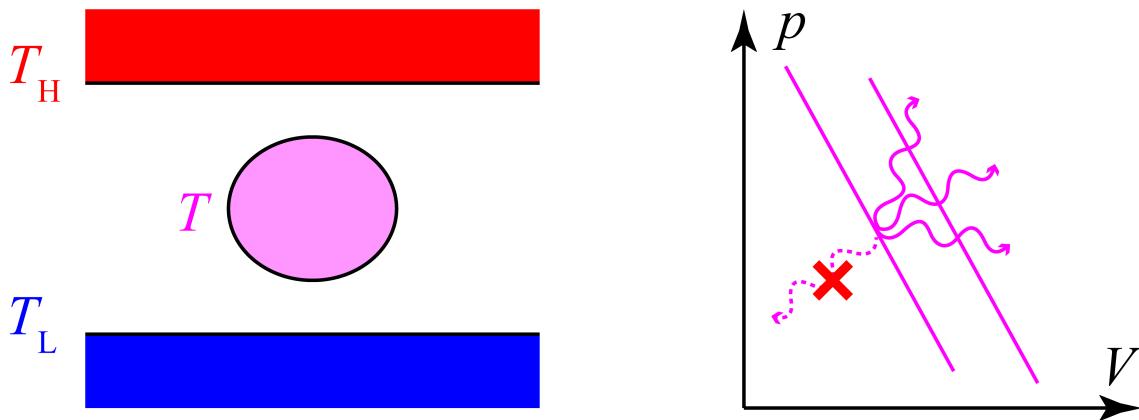


图 7.13: 熵增原理

我们讨论一个 $p - V$ 系统作为工作物质在两个恒温热源 (冷库在此被称为低温热源) 之间的工作, 各个温度暂时只是一个记号, 其值暂不确定. 为了统一起见, 我们研究系统与外界的功热交换时规定 $dU = dW + dQ$, 即以从外界吸热和外界对其做功为正. 首先, 我们孤立那个 $p - V$ 系统, 让它做绝热过程而只能与外界有功的交换. 那么准静态的绝热过程³将给出系列绝热线, 它们彼此绝不可以相交, 否则在一点处将存在多个绝热的方向. 图中为了简化而近似画成了平行直线 (这在微元的循环过程中——微小吸放热, 微小温差——是合理的处理方法). 现在考虑一个非准静态的绝热过程. 如果是绝热压缩, 那么外界给予气体的压强应该大于系统的平均压强, 这样造成的不平衡才可以使得系统自发收缩. 如果是绝热膨胀, 同理, 外界给予气体的压强应该小于系统的平均压强. 不管怎样, 都有:

$$dW \geq -pdV$$

这样造成的结果是系统的状态只能往绝热线的一侧走, 类似于热力学第零定律, 对于这个特定体系, 我们暂时可以把这些可以通过准静态绝热过程相互连接的状态称为具有共同的熵. 而更高的绝热线代表更高的熵, 以上观点给出了**熵增原理**(principle of entropy increase):

熵增原理: 孤立系统的熵只增不减, $\Delta S \geq 0$

这一点的意义不仅仅在于末态到达另一条绝热线上以后仍然孤立地自发回到初始状态, 还意味着既使借助外界——两个恒温热源, 可能出现的其他辅助工作物质的帮助, 也无法在不改变其他物体的状态 (两个热源净吸放热为零, 其他工作物质初末状态复原) 的条件下回到初始状态, 原来的功也被补偿. 这构成热力学第二定律的一种表述, 它在热力学体系里是公理, 代表了自然界的某种规律.

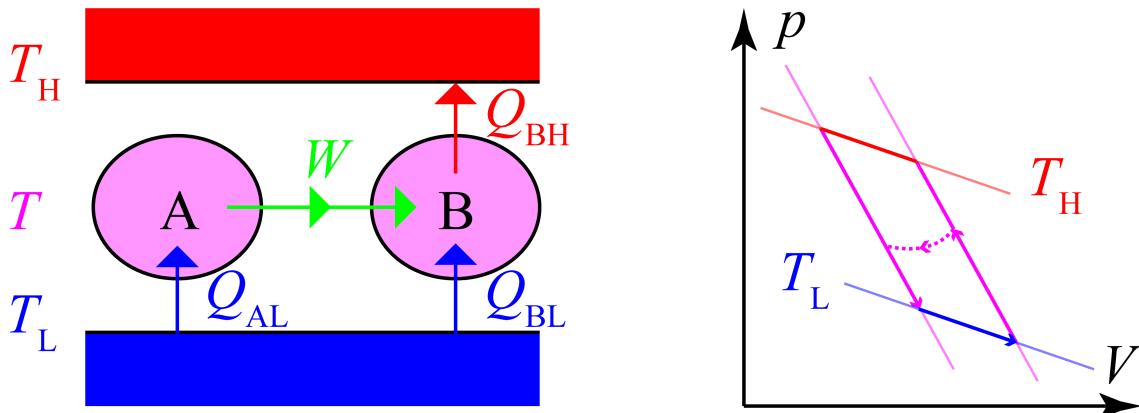


图 7.14: 三种不可逆过程的联系

³它具有极其独特的含义, 参见之后第五章的讨论.

质疑这一点会引发什么结果?如果在某个任意中间温度 T , $T_L < T < T_H$ 下可以通过某个过程让较高绝热线上的状态回到相距很近的较低绝热线上的状态.那么我们将它与两条绝热线和在低温热源上吸热膨胀的过程组成一个循环,其合效果是从低温热源吸收了热,并完全转变为对外界的做功.这也是被热力学第二定律禁止的,或者说其反过程不可逆:

$$\text{开尔文表述: 功变热不可逆, } W = -Q \geq 0$$

表述为文字,则为热机不可能在一个循环过程中从单一热源吸热转变为对外做功而不产生其他影响.

继续质疑这一点.再构造一个热泵 B,它接受来自热机 A 的功而从低温热源向高温热源泵热,其总结果是,低温热源向高温热源传了热,其他体系状态历经一个循环回到初态.同样的,它被热力学第二定律禁止:

$$\text{克劳修斯表述: 温差传热不可逆, } Q_H = -Q_L \geq 0$$

可见热力学第二定律的不同表述之间都是相互联系的.实际上如果把两个热源和工作物质视为整体,那么其实后两个表述就是整个体系的熵增原理,此时,整个循环过程来看,如果过程中时时,处处都是无耗散的准静态过程,那么熵应该是不变的,过程应该是可逆的,这就是卡诺热机模型:是高温热源与气体等温时放热 $Q_H > 0$,气体与低温热源等温时再放热 $|Q_L| < 0$.而气体对外做功为 $|W| = -Q_H - Q_L < 0$.或者是这个过程的逆过程——卡诺热泵循环.对于所有热机,热力学第二定律断言:

$$\text{卡诺定理: 卡诺热机效率最高, } \eta = 1 - \frac{-Q'_L}{Q'_H} \leq 1 - \frac{-Q_L}{Q_H}$$

证明思路是,如果有热机效率高于卡诺热机效率,意味着输出相同的功时其在两个热源的吸放热可以同时减到比卡诺热机小,那么用输出的功给一个对应的卡诺热泵,总的结果是低温热源向高温热源放热了,由于热力学第二定律,这不可以成立.

从而,卡诺定理实际上也是热力学第二定律的等价表述.这些表述都有一个特点,可逆过程对应取等号,而不等号发生在不可逆的过程中,代表过程的方向性.在等号情况下,工作于两个恒温热源间的可逆卡诺热机的吸放热比是恒定的,这给出了热力学温标的定义:

$$\frac{-Q_L}{Q_H} = \frac{T_L}{T_H}$$

只需确定一种特定系统的温度,便可以由可逆循环的吸放热比测量任意其他平衡态的温度.目前,这种特定的系统被取做水的三相共存系统⁴,温度定义为 273.16K.这就是热力学温标.而这也给出了与之共轭的熵的概念:

$$dS = \frac{dQ}{T}$$

上式对可逆过程适用.对于卡诺循环来说,如果着眼点在整个系统,那么一个循环下来工作物质的熵是复原的.而两个恒温热源的熵变根据上面温度的定义:

$$\Delta S = -\frac{Q_H}{T_H} - \frac{Q_L}{T_L} = 0$$

可见大的系统初末态在一条等熵线(实际上整个循环中每一个状态之间都彼此等熵,因为整个系统做绝热可逆过程)上.而如果我们着眼点在气体上,而且让气体做一个任意的可逆循环过程,一定有:

$$\Delta S = \oint \frac{dQ}{T} = 0$$

以上式子为克劳修斯等式(Clausius equality),同理对于不可逆过程将给出不等式,普遍的,我们对一个系统,有:

$$\Delta S = S_{\text{final}} - S_{\text{initial}} \geq \int_{\text{initial}}^{\text{final}} \frac{dQ}{T}$$

⁴2018 年预通过的新国际单位制提案指出将可能把联系微观统计规律与宏观测量的常数 k_B 作为基本常数.这样确定了温度与能量换算的线性关系,从极低温到极高温完全一致.

上式也是热力学第二定律的表述,它十分强大,其他表述都可以简单地从这个表述验证.它甚至有十分清晰的图像:右边是从外界流向系统的熵流,左右的区别在于系统内部不平衡条件下的各种流造成的熵产生,详见下一节和本章最后一节的相关讨论.对于上式,如果不等号左边为零,可以理解为功变热的不可逆性:右边积分里的 dQ 小于零将使得不等式成立,说明在过程中体系更自发的行为是接受外界对其做功,而以放热的形式形成能量守恒,当然条件是初末态等熵——一般考虑循环过程.如果不等式右边为零,给出了熵增原理——绝热体系的熵只增不减.

7.4 熵的计算

7.4.1 理想气体的熵

理想气体温标与热力学温标为何统一?一者,用理想气体温标定义的温度自动符合卡诺定理:卡诺循环效率为热力学定义值 $\eta = 1 - \frac{T_L}{T_H}$,且关键是它能使得下式成为全微分:

$$dS = \frac{dQ}{T} = \frac{dU + pdV}{T} = f(T, V)dT + g(T, V)dV$$

如果理想气体温标是 $T' = h(T) \cdot T$,那么 $\frac{dQ}{T'}$ 不可能再是一个全微分,而与热力学第二定律定义的温标不一致.

二者,热力学第二定律的体系看似复杂而充满精巧的构造,实则可以在更深的层次:微观动力学规律与统计规律的基础上进行自然的推导.此时温度,熵,热量不仅可以普适而简单地定义,而且具有清晰的物理图像,我们在第五章统计物理基础中展开说明.

所以,理想气体的熵我们可以直接使用第二节的推导,不过我们这里写作:

$$S(T, V) = \nu C_{mV} \ln T + \nu R \ln V$$

$$S(p, V) = \nu C_{mV} \ln p + \nu C_{mp} \ln V$$

$$S(p, T) = \nu C_{mp} \ln T - \nu R \ln p$$

以上三个定义不太一样,差一些取决于摩尔数 ν 的常数.而且只有最下面一个是广延量.一种常见的处理方法是把体积换为摩尔体积:

$$V_m = \frac{V}{\nu}$$

摩尔体积实际上与比体积(单位质量的体积) v ,密度 ρ ,粒子数密度 n 有如下正反比关系:

$$\rho = \frac{1}{v} = \frac{\mu}{V_m} = nm$$

这样三个式子都变成了广延量.它们还差一些与 ν 成正比,对数里面是 R 的一些广延量.这种定义的非唯一性很正常,实际过程中产生实际效果的是熵的变化量,仅有在极低温下,讨论量子体系时才需要讨论熵的绝对数值(热力学第三定律).还有一点,为了使得对数里的物理量不带单位,我们规定 $p = p_0$, $v = v_0$, $T = T_0$ 时的熵为零,最后得到统一的熵的公式:

$$S = \nu C_{mV} \ln \frac{T}{T_0} + \nu R \ln \frac{v}{v_0} = \nu C_{mV} \ln \frac{p}{p_0} + \nu C_{mp} \ln \frac{v}{v_0} = \nu C_{mp} \ln \frac{T}{T_0} - \nu R \ln \frac{p}{p_0}$$

熵是广延量,我们作此要求.说明我们可以将一个均匀的系统按某比例分成若干份,熵也会随之分成对应比例的份数.甚至,我们可以将上面的熵除以总分子数,得到每一个分子上的熵的结果:

$$\varsigma = \left(\frac{i}{2} + 1\right) k \ln T - k \ln p$$

乍一看这是荒诞的,一个分子具有能量可能还比较好理解,但是一个分子为何可以具有温度,压强,甚至熵的概念?这些问题需要等我们之后章节研究了熵的统计学和信息学含义以后才能理解.

7.4.2 固定热容固体熵

理想情况下这种固体是以温度为单参数热力学系统, 其内能为

$$U = CT$$

而吸热就是 $Q = C\Delta T$, 从而熵:

$$dS = C \frac{dT}{T}$$

从而:

$$S = C \ln \frac{T}{T_0}$$

7.4.3 大热容恒温热库

无疑, 这种情况是上一种情形的大热容极限近似, 如果温度变化 ΔT 相对于 T 很小:

$$\Delta S = C \ln \frac{T + \Delta T}{T} \simeq \frac{C\Delta T}{T} = \frac{\Delta U}{T} \simeq \Delta \frac{U}{T}$$

即此时等价的定义不会造成区别:

$$S = \frac{U}{T}$$

7.4.4 传热熵

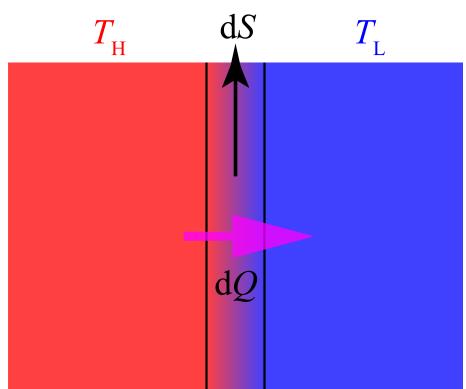


图 7.15: 传热熵产生

如果高温物体温度 T_H , 低温物体温度 T_L , 两者发生一个热接触而有热流 dQ 从高温物体流向低温物体. 这是一个非准静态过程, 它不可逆, 如何分析其熵的呢? 考虑高温物体放热 dQ , 实际上如果它是与一个温度同为 T_H 的恒温热库接触而放出这个热量, 其状态的改变是完全相同的, 不过这是一个准静态过程, 可以计算其熵变:

$$dS_H = -\frac{dQ}{T_H}$$

即, 有 $|dS_H|$ 的熵从高温物体中流了出来. 同理, 流向低温物体的熵为:

$$dS_L = \frac{dQ}{T_L}$$

整个体系熵要增加:

$$dS = dQ \left(\frac{1}{T_L} - \frac{1}{T_H} \right) > 0$$

上式被称为传热过程中的熵产生 (entropy production). 它其实也代表某种微观形式的耗散, 它发生于高温与低温物体接触时平均起来高温的动能较大的粒子动能被拉低而低温物体动能被拔高的过程中.

7.4.5 混合熵

讨论理想气体的混合, 而且混合后各种气体分子间亦无任何形式的相互作用. 命气体的第 i 组分的摩尔分数为 x_i , 混合前不同气体处于温度为 T , 压强为 p 的平衡, 不同组分占体积比为:

$$V_1 : V_2 : \cdots : V_n = x_1 : x_2 : \cdots : x_n ; \quad V_i = x_i V$$

混合后温度仍然是 T , 体积倒是共用一个 V 了. 而不同组分去对 p 进行分压:

$$p_1 : p_2 : \cdots : p_n = x_1 : x_2 : \cdots : x_n ; \quad p_i = x_i p$$

熵是广延量, 混合理想气体的熵满足吉布斯定理(Gibbs' theorem): 混合气体的熵等于所有组分的熵的和, 而每一组分在与其他组分共存的过程中, 其他组分对这个组分可能的微观状态形式并无影响(分子间无相互作用). 所以这个组分的熵就等于它单独存在时的熵. 那么过程前后的熵变其实也就是气体向真空自由膨胀的熵变(因为 T 恰好不变, 与自由膨胀特征相同), 这是一个典型的不可逆过程:

$$\Delta S_i = \nu_i R \ln \frac{V}{V_i} = -\nu R x_i \ln x_i$$

故算出混合熵(entropy of mixing) 为:

$$\Delta S = -\nu R \sum x_i \ln x_i > 0$$

我们写出末态混合均匀后熵的表达式, 注意为了保证其广延性, 我们采用 p, T, x_i 这些强度量作为其可以放到 \ln 后面的自变量:

$$\begin{aligned} S &= \sum S_i \\ &= \sum \nu_i C_{mp} \ln T - \nu_i R \ln p_i \\ &= \nu C_{mp} \ln T - \nu R \ln p - \nu R \sum x_i \ln x_i \\ &= S_0 + \Delta S \end{aligned}$$

从中混合熵的概念就十分明显了: 它代表混合气体与同样条件下的单组分气体的熵 S_0 的区别. 这很自然地引入了所谓的吉布斯佯谬(Gibbs' paradox): 一是为什么吉布斯定理不适用于单组分气体, 如果把一缸气体体积分成两份, 那么熵可以直接相加(广延性), 但不能把这缸气体的压强(粒子数)分成两份而占据相同的体积, 除非分成的两份分子有本质的不同. 二就是多组分与单组分的区别在哪里, 会造成物理量的差别.

这其实不是奇怪的事. 我们考虑一缸氢气, 它的熵应根据 S_0 计算, 用它计算也符合测量结果. 而某一天科学家们突然得知, 原来氢分子 H_2 分为两种: 核自旋平行的正氢(orthohydrogen) 和核自旋反向的仲氢(parahydrogen), 两者占比为 3 : 1. 两种分子仅仅是深入到原子核的性质略有不同, 对常温下的分子物理化学性质几乎没有影响⁵. 于是似乎要在熵公式后附加一项混合熵 $-\nu R (\frac{1}{4} \ln \frac{1}{4} + \frac{3}{4} \ln \frac{3}{4}) \approx 0.56\nu R$. 这一项如何解释? 通过第五章的学习我们将知道, 熵实际上对应着信息的多少. 在混合氢情况下我们知道了很多的信息: 每一个氢分子究竟是正氢还是仲氢. 所以熵增加了. 但加这一项有必要吗? 其实是没有必要的, 因为它作为常数, 在公式 $dQ = TdS$ 中求变化量时不出现, 从而不会出现在实际可以测量的系统发生的过程中. 然而一涉及到微观构成, 它就是必要的了, 如果两种氢之间可以相互转化, 那么如果整缸氢气都是正氢或是仲氢, 在一定时间后它将在两种氢之比为 3 : 1 处平衡, 这个过程是个不可逆过程. 这是因为正氢又分为三种可能的自旋态, 那么把这三个态与仲氢的摩尔分数记做 x_1, x_2, x_3, x_4 , 有:

$$x_1 + x_2 + x_3 + x_4 = 1$$

$$\Delta S = -\nu R \ln \sqrt[4]{x_1 x_2 x_3 x_4}$$

由算术几何平均不等式, 在 $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$ 处熵增取极大值. 在这个状态往任何一个其他状态熵必然减少而不被热力学允许. 它就是我们要求的平衡态. 所以两种氢之比为 3 : 1, 同时我们还可以看出之前我们算的熵变还不够严格, 考虑四种氢时混合熵应为

$$\Delta S = \nu R \ln 4 = 1.4\nu R$$

值得一提, 我们上面反过来通过预先在非平衡态都有良好定义的熵的极值来求平衡状态的参数条件的方法称为最大熵原理(principle of maximum entropy), 这种目前看似难以理解的方法亦会在第五章给出详细论述.

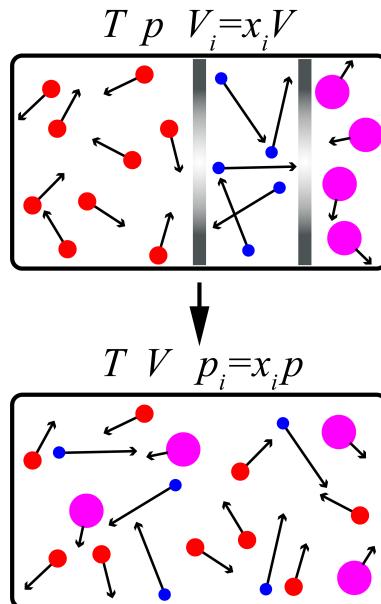


图 7.16: 混合熵

⁵然而低温下影响巨大, 这与量子效应有关

7.5 热力学函数与其特性

7.5.1 四个热力学函数与四个状态参量

在本节中我们将把态函数熵 S 视为与 p, V, T 并列的状态参量. 由热力学第一定律:

$$dU = dQ + dW$$

无论是在准静态或非准静态过程中上式都使用, 而只有在准静态过程中:

$$dQ = TdS \quad ; \quad dW = -pdV$$

于是:

$$dU = TdS - pdV$$

这个式子又是对准静态和非准静态过程都适用的了, 因为等式两边都是状态量, 它们不关心过程的性质. 这是对符合热力学规律的正确定义的各个物理量之间提出的必然要求, 被称作**热力学基本方程**(master equation). 这里, 可以理解为如果把内能写作熵 S 和体积 T 的函数, 它必须满足:

$$\left(\frac{\partial U}{\partial S}\right)_V = T ; \quad \left(\frac{\partial U}{\partial V}\right)_S = -p$$

由混合偏导数与求导次序无关的性质, 可以得到:

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V$$

而我们之前定义了焓 $H = U + pV$, 那么相应的有:

$$dH = TdS + Vdp$$

$$\left(\frac{\partial H}{\partial S}\right)_p = T ; \quad \left(\frac{\partial H}{\partial p}\right)_S = V$$

$$\left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p$$

我们再定义**亥姆霍兹自由能**(Helmholtz's free energy), 简称**自由能**(free energy) 为 $F = U - TS$, 从表达式上看, 它反映了一种能量最低与熵最高的“抗衡”. 它的相应结论为:

$$dF = -SdT - pdV$$

$$\left(\frac{\partial F}{\partial T}\right)_V = -S ; \quad \left(\frac{\partial F}{\partial V}\right)_T = -p$$

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V$$

最后还可以定义**自由焓**(free enthalpy), 它又被叫做**吉布斯自由能**(Gibbs' free energy), 定义为 $G = H - TS = U + pV - TS$, 相关性质为:

$$dG = -SdT + Vdp$$

$$\left(\frac{\partial G}{\partial T}\right)_p = -S ; \quad \left(\frac{\partial G}{\partial p}\right)_T = V$$

$$\left(\frac{\partial S}{\partial p}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_p$$

以上四个热力学基本函数的定义自然而然的给出了四个基本关系式,称为麦克斯韦关系 (Maxwell relations):

$$\left(\frac{\partial T}{\partial V}\right)_S = - \left(\frac{\partial p}{\partial S}\right)_V$$

$$\left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p$$

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V$$

$$\left(\frac{\partial S}{\partial p}\right)_T = - \left(\frac{\partial V}{\partial T}\right)_p$$

如何理解这些关系?它们实际上都来自热力学规律对从 p, V 确定 T, S 的映射的性质.如果我们画出 $p - V$ 图与 $T - S$ 图,则两个图形上的点一一对应. $p - V$ 图上的一个闭区域的面积必须和对应的 $T - S$ 图上的面积相等.这是因为其边界对应的循环过程的吸热与做功就是两个面积,由热一两者必然相等.这种性质在数学上表述为雅可比行列式 (Jacobian determinant) 必须等于一:

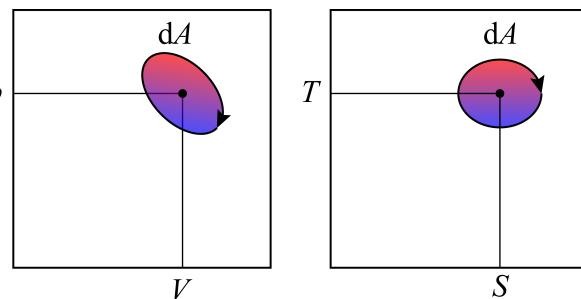


图 7.17: 循环过程的做功与吸热相等

$$\frac{\partial(T, S)}{\partial(p, V)} = \begin{vmatrix} \frac{\partial T}{\partial p} & \frac{\partial S}{\partial p} \\ \frac{\partial T}{\partial V} & \frac{\partial S}{\partial V} \end{vmatrix} = \left(\frac{\partial T}{\partial p}\right)_V \left(\frac{\partial S}{\partial V}\right)_p - \left(\frac{\partial S}{\partial p}\right)_V \left(\frac{\partial T}{\partial V}\right)_p = 1$$

而且是 +1,这代表循环的顺时针与逆时针在映射下保持不变.对于符号 $\partial(p, V)$,我们理解为点 (p, V) 周围的一小块邻域的面积大小.这样这个式子就是一个直观的微商式,而上式两个面积比之所以可以用行列式来表示,是因为在 $p - V$ 图上 (p, V) 点移动 dp ,在 $T - S$ 图上移动矢量 $(\frac{\partial T}{\partial p}, \frac{\partial S}{\partial p})dp$,若移动 dV ,则在 $T - S$ 图应移动 $(\frac{\partial T}{\partial V}, \frac{\partial S}{\partial V})dV$,那么在 $p - V$ 图上由两个移动构成的矩形面积为 $dpdV$,而 T_S 图上面积应为 $\left|(\frac{\partial T}{\partial V}, \frac{\partial S}{\partial V}) \times (\frac{\partial T}{\partial V}, \frac{\partial S}{\partial V})\right| dpdV$,根据代数知识这个叉乘就等于以上行列式.

也是基于这个认识,不难写出:

$$\left(\frac{\partial T}{\partial V}\right)_S = \frac{\partial(T, S)}{\partial(V, S)} = \frac{\partial(p, V)}{\partial(V, S)} = -\frac{\partial(V, p)}{\partial(V, S)} = -\left(\frac{\partial p}{\partial S}\right)_V$$

其他三个麦克斯韦关系也可以依据这个方法进行推导.在 $p - V$ 图和 $T - S$ 图上应该如何直观理解这个式子还请读者自己思考.

7.5.2 若干定理的证明

能态定理

对于 $p - V$ 体系,若给出状态方程,则能量关于状态参量的偏导数有以下条件必须满足:

$$\left(\frac{\partial U}{\partial V}\right)_T = T \left(\frac{\partial p}{\partial T}\right)_V - p$$

证法一：

利用 $U(S, V)$ 是特性函数(characteristic function), 用它可以表示出任意状态参量与其偏导数:

$$\begin{aligned} p &= -U_V, \quad T = U_S \\ \left(\frac{\partial p}{\partial T}\right)_V &= \left(\frac{\partial S}{\partial T}\right)_V \cdot \left(\frac{\partial p}{\partial S}\right)_V = -\left(\frac{\partial S}{\partial T}\right)_V \cdot U_{SV} \\ dT = dU_S &= U_{SS}dS + U_{SV}dV \Rightarrow \left(\frac{\partial S}{\partial T}\right)_V = \frac{1}{U_{SS}}, \left(\frac{\partial p}{\partial T}\right)_V = -\frac{U_{SV}}{U_{SS}} \end{aligned}$$

从而等式右边为:

$$T \left(\frac{\partial p}{\partial T}\right)_V - p = U_V - \frac{U_S U_{SV}}{U_{SS}}$$

而等式左边为:

$$\begin{aligned} \left(\frac{\partial U}{\partial V}\right)_T &= \left(\frac{\partial U}{\partial V}\right)_S + \left(\frac{\partial U}{\partial S}\right)_V \cdot \left(\frac{\partial S}{\partial V}\right)_T \\ dT = dU_S &= U_{SS}dS + U_{SV}dV \Rightarrow \left(\frac{\partial S}{\partial V}\right)_T = -\frac{U_{SV}}{U_{SS}} \end{aligned}$$

代入, 即有:

$$\left(\frac{\partial U}{\partial V}\right)_T = U_V - \frac{U_S U_{SV}}{U_{SS}}$$

从而两式相等.

证法二:

方程以 T, V 作为自变量, 故考虑自由能 $F = U - TS$, 它恰好以 T, V 为自变量时为特性函数, 有:

$$\left(\frac{\partial U}{\partial V}\right)_T = \left(\frac{\partial F}{\partial V}\right)_T - T \left(\frac{\partial S}{\partial V}\right)_T = -p + T \left(\frac{\partial p}{\partial T}\right)_V$$

后一步推导用到了麦克斯韦关系.

证法三:

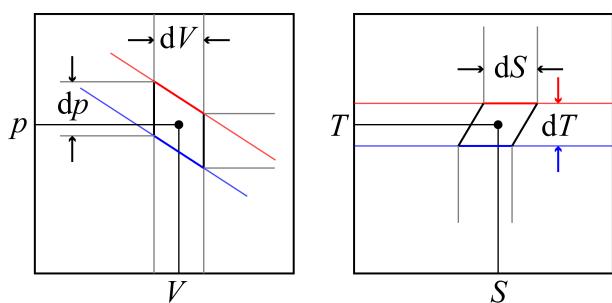


图 7.18: 构造循环过程证明能态定理

构造在两根相差 dV 的等体线和相差 dT 的等温线之间的循环过程. 这不是一个卡诺循环, 但依然近似有两个等温过程中的吸放热定义的效率为:

$$\eta = 1 - \frac{dQ_L}{dQ_H} = \frac{\delta dQ}{dQ} = \frac{dT}{T}$$

这一点从这个循环对应的 $T - S$ 图中可以看出来, 小段红线下方的面积与小段蓝线下方的面积, 由于微元过程近似为平行四边形, 两个面积的底 dS 相等, 面积差即为平行四边形的面积 $dTdS$, 而吸放热约为 TdS , 它们的比就是上式.

现在我们把注意力集中在左图中. 在等体过程中增加的压强为 $\left(\frac{\partial p}{\partial T}\right)_V dT$, 从而过程做功为:

$$dW = \left(\frac{\partial p}{\partial T}\right)_V dTdV$$

而两个等温过程中的吸热, 由热力学第一定律:

$$dQ = pdV + \left(\frac{\partial U}{\partial V}\right)_T dV$$

把两式与循环效率比对:

$$\frac{dW}{dQ} = \frac{dT}{T}$$

即得:

$$\left(\frac{\partial U}{\partial V}\right)_T = T \left(\frac{\partial p}{\partial T}\right)_V - p$$

把这个式子用于理想气体的物态方程:

$$pV = \nu RT \Rightarrow \left(\frac{\partial U}{\partial V}\right)_T = 0$$

可见理想气体的内能只和温度有关是由物态方程决定的. 而满足该物态方程的其他气体模型也需要符合这一点, 所以内能对温度的偏导数仍然符合等体热容的定义, 但热容一般是温度的函数:

$$U = \nu \int C_{mV}(T) dT$$

实际气体, 一方面在高密度, 高温度情况下其状态方程会偏离理想物态方程的形式, 另一方面, 多原子分子气体其内能-温度关系在常温-低温区一般体现为温度的单调递增函数的形式, 而在某些温度区间达到一些近似不变的平台. 这反应了气体内部自由度的量子本质. 在一种近似的模型处理中, 我们可以让 $T < T_c$ 时摩尔等体热容为 $C_{mV} = \frac{i}{2}R$, 而 $T > T_c$ 时摩尔等体热容为 $C_{mV} = \frac{f}{2}R$, 而在 $T = T_c$ 附近的微小区间内不仅热容迅速升高, 内能也突然从 $\frac{i}{2}\nu RT_c$ 上升到 $\frac{f}{2}\nu RT_c$ ⁶. 这实际上是一种相变-期间两种热容的分子可以视为共存, 详见第四章相与相变的内容.

这种情况也很容易推广到一个类似体系中: 液膜的热力学系统. 在此时热力学基本方程写作:

$$dU = TdS + \sigma dA$$

可见在这儿表面张力系数 σ 是广义力, 广义坐标是表面积 A . 故我们可以写出:

$$\left(\frac{\partial U}{\partial A}\right)_T = \sigma - T \left(\frac{\partial \sigma}{\partial T}\right)_A$$

一般 σ 是温度 T 的函数, 只不过与理想气体情况不太相同的是, σ 依赖于 T 但一般不依赖于液面表面积 A , 液膜的拉伸改变的是膜的厚度, 而表面层永远是膜的两个表面上极薄的一层, 它的性质不会改变, 从而不会改变表面张力系数 (详见下一章液体与固体的性质). 事实上, 此时我们注意到自由能的性质为:

$$dF = -SdT + \sigma dA$$

而 $\frac{\partial \sigma}{\partial A} = 0$ 意味着:

$$\left(\frac{\partial F}{\partial A}\right)_T = \sigma(T) ; \quad \left(\frac{\partial^2 F}{\partial A^2}\right)_T = 0$$

这意味着 $F = \sigma(T)A + f(T)$, 对于 $f(T)$, 显然与表面现象没有关系, 因为当面积 $A = 0$ 时这一项仍然存在, 说明不是因为表面现象引起的, 实验时与理论计算都应该扣除. 从而我们给出:

$$F = \sigma A$$

也就是说平时说的表面能在热力学下实际上是自由能. 它恰恰是表面积 A 的广延量. 而熵应该为:

$$S = -\left(\frac{\partial F}{\partial T}\right)_A = -\sigma' A$$

⁶事实上由于需要吸热而温度不变, 这个过程热容是发散的.

最后再代入到 $U = F + TS$, 得:

$$U = (\sigma - T\sigma')A$$

内能与面积的比例系数不是 σ , 但由 σ 决定, 意料之外, 情理之中. 这意味着等温地拉伸液膜是需要吸放热的, 从微观机理上来看, 熵必须大于零, 从而 A 增加熵应该也要增加. 从以上熵的表达式来看必然有:

$$\sigma' < 0$$

从而我们得出了温度升高表面张力系数减小的结论. 微观上看, 温度升高导致分子从液相通过表面层到气相更简单, 表面层变薄, 所以表面能减小. 从而拉伸液膜是要吸热的. 而如果面积不变, 而温度升高, 我们得到等面积热容为

$$C_A = -T\sigma''A$$

对于一般的有机液体有经验公式:

$$\sigma = \sigma_0 \left(1 - \frac{T}{T_C}\right)^n$$

一般 $n \simeq \frac{11}{9}$, 从而我们发现 $C_A < 0$ ⁷. 这没什么奇怪的, 要知道液膜的表面层是依附于内部的液体而存在的. 所以既使增加 A 对内部液体没有影响, 升温绝对是对两侧的液, 气相产生影响的. 这个热容将被包含在正的较大的液体气体热容中. 它的减小可以理解为温度升高时表面层变薄而直接减少内能的缘故. 另外, 在 $T = T_C$ 处 σ 减小到零, 内能与自由能也都减小到了零. 这也是预料之中的. 这一点就是液相与气相共存的临界点. 第四章相与相变将仔细研究这个现象.

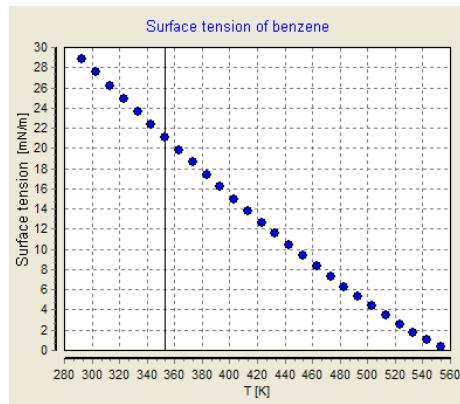


图 7.19: 苯的 $\sigma \sim T$ 关系

物性函数间关系

除了两个热容 C_{mp}, C_{mV} , 实验中容易测得的量一般只能是 p, V, T . 从而连带着可以测的还有在上一章理想气体的过程这一节中介绍的各种**物性函数**(material property function): 膨胀系数 α , 压强系数 β 和压缩系数 κ . 不同的过程中这些量间有关系:

$$\alpha_X = -p\beta_X\kappa_X$$

$$\alpha_p = p\beta_V\kappa_T$$

$$\gamma = \frac{C_p}{C_V} = \frac{\kappa_T}{\kappa_S}$$

$$C_p - C_V = pVT\alpha_p\beta_V$$

其中下标 X 表示一个任意的过程.

证明:

二自由变量的 $p-V$ 系统引入一个过程方程后变为一自由变量, 所有变化的微元互相之间都成正比, 即:

$$dp = \lambda_p dx, dV = \lambda_V dx, dT = \lambda_T dx$$

其中 x 作为过程参数反应过程发生的程度. 从而三个 X 过程的响应函数分别为:

$$\alpha_X = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_X = \frac{\lambda_V}{V\lambda_T}$$

⁷热力学预言简单的 $p-V$ 系统如果热容小于零将是不稳定的——传热将使得局部高温更高, 低温更低.

$$\beta_X = \frac{1}{p} \left(\frac{\partial p}{\partial T} \right)_X = \frac{\lambda_p}{p \lambda_T}$$

$$\kappa_X = -\frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_X = -\frac{\lambda_V}{V \lambda_p}$$

从而得到待证第一式. 而第二个式子的区别在于它不依赖于某个过程, 而是直接从物态方程中得出:

$$F(p, V, T) = 0 \Rightarrow F_p dp + F_V dV + F_T dT = 0$$

$$\Rightarrow \alpha_p = -\frac{F_T}{VF_V}, \beta_V = -\frac{F_T}{pF_p}, \kappa_T = \frac{F_p}{VF_V}$$

即可证明第二式. 这两个过程都仅仅是数学恒等式, 不依赖于任何体系的物理性质. 而第三个式子则需要热二的相关推论来推导. 首先注意到一下关于两个热容的定义是与之前说法等价的:

$$C_V = \left(\frac{\partial U}{\partial T} \right)_V = T \left(\frac{\partial S}{\partial T} \right)_V; \quad C_p = \left(\frac{\partial H}{\partial T} \right)_p = T \left(\frac{\partial S}{\partial T} \right)_p$$

从而利用雅可比行列式的方法:

$$\frac{C_p}{C_V} = \frac{\left(\frac{\partial S}{\partial T} \right)_p}{\left(\frac{\partial S}{\partial T} \right)_V} = \frac{\partial(S, p)}{\partial(T, p)} \cdot \frac{\partial(T, V)}{\partial(S, V)} = \frac{\partial(T, V)}{\partial(T, p)} \cdot \frac{\partial(S, p)}{\partial(S, V)} = \frac{\left(\frac{\partial V}{\partial p} \right)_T}{\left(\frac{\partial V}{\partial p} \right)_S} = \frac{\kappa_T}{\kappa_S}$$

这就是第三式, 而最后一式则考虑 (V, T) 为自变量:

$$C_p - C_V = T \left(\frac{\partial S}{\partial T} \right)_p - T \left(\frac{\partial S}{\partial T} \right)_V = T \left(\frac{\partial S}{\partial V} \right)_T \left(\frac{\partial V}{\partial T} \right)_p = T \left(\frac{\partial p}{\partial T} \right)_V \left(\frac{\partial V}{\partial T} \right)_p = pVT\alpha_p\beta_V$$

期间用了一次麦克斯韦关系. 至此四个关系式都得证.

吉布斯-亥姆霍兹方程

吉布斯-亥姆霍兹方程 (Gibbs-Helmholtz equation) 用来确定化学反应中的吸放热 (实际上是焓变) 与吉布斯自由能的变化间的关系, 而吉布斯函数又和化学反应发生的方向, 化学平衡的位置紧密相关. 这个关系写作:

$$\left(\frac{\partial}{\partial T} \right)_p \frac{G}{T} = -\frac{H}{T^2}$$

证明:

实际上:

$$H = G + TS = G - T \left(\frac{\partial G}{\partial T} \right)_p = -T^2 [G \left(\frac{\partial}{\partial T} \right) \frac{1}{T} + \frac{1}{T} \left(\frac{\partial}{\partial T} \right) G] = -T^2 \left(\frac{\partial}{\partial T} \right) \frac{G}{T}$$

这就是上式. 相似的关系我们还可以写出:

$$\left(\frac{\partial}{\partial T} \right)_V \frac{F}{T} = -\frac{U}{T^2}$$

$$\left(\frac{\partial}{\partial V} \right)_T \frac{F}{V} = -\frac{G}{V^2}$$

$$\left(\frac{\partial}{\partial p} \right)_T \frac{G}{p} = -\frac{F}{p^2}$$

7.5.3 自由能的含义

我们先讨论亥姆霍兹自由能,对于双参数描述的简单系统,此时热力学基本关系牢固地成立,不管过程是否准静态:

$$dU = TdS - pdV$$

但对于非双参数的,比如我们考虑两种气体的相互扩散,初态与末态似乎其表现出来的性质是类似的.比如标准状态下1mol正氢与1mol仲氢混合均匀.如果视为一个 $p-V$ 系,整个过程前后等温,等压,等体,但熵增.这类问题上式就不再是一个等式.究其本质,在与内部发生的各种过程的各种不平衡与耗散的因素造成的自发熵产生.体系总是受迫地远离平衡态,而有自发地从非平衡态回到平衡态.为了简化,我们假设系统与外界的热,功交换仍然是准静态且无耗散的⁸,熵产生仅仅发生在体系内部.此时上式应该写为:

$$\Delta U \leq T\Delta S - p\Delta V$$

这是因为外界对系统做功仍然为 $\Delta W = -p\Delta V$,但根据克劳修斯不等式系统从外界吸热满足 $\Delta Q \leq T\Delta S$.而最终 $\Delta U = \Delta Q + \Delta W$ 的缘故.注意到克劳修斯不等式的本质就是系统与外界一起看成一个整体的熵增原理.它是普适的,对任意复杂体系都适用的统计规律.上式也通常写成

$$\Delta S \geq \frac{\Delta U}{T} + \frac{p\Delta V}{T}$$

上式也与熵增原理兼容,等体,绝热的体系自然可以认为是孤立的,它内能不变,从而上式给出了 $\Delta S > 0$.但上式实际上比熵增原理更强.因为即使如果对于 V 增大的过程,即使不可逆,上式对理想气体也应该严格成立,但对于一般多内部参量的 $p-V$ 系统上式则可以取不等号.上式还给出了平衡态的条件:我们可以在不改变体系总内能和总体积的情况下分配各部分所占的内能和体积,如果可以通过某种分配方式让熵增加,则体系会自发地往这个方向演化.直至熵达到它的极大值,对应热力学平衡态.这叫做**最大熵原理**(principle of maximum entropy).

从内能的相似表达式中可以得到完全一致的结论,不过此时我们要保持熵 S 与体积 V 不变.而结论是对应的**热力学势**(thermodynamic potential)——内能取最小.对应的原理为**最小能原理**(principle of minimum energy).从内能出发的这个结论有些不足之处:熵理论上计算公式复杂,实际操作上又因为无法直接控制其不变而没有意义,且难以用微观统计观点诠释的熵的概念解释.所以仅有较为好的热力学理论价值.

亥姆霍兹自由能的优势就体现在这儿了,由于 $F = U - TS$,在等温的条件下我们写出:

$$\Delta F \leq -S\Delta T - p\Delta V$$

从而这个关系式对应的平衡判据为——在等温等体情况下自由能必须极小.这个式子实际上经常意味着纯粹的机械平衡.例如我们还记得液膜的自由能就是 $F = \sigma A$,而后一项 $-p\Delta V$ 理解为 ΔW ,这个原理就是:在一定温度下固定边界(外界不做功)的液膜的面积要收缩到最小.利用**变分法**(calculus of variation)我们可以把这个最小曲面确定下来.我们将在下一章介绍液体表面张力时介绍相关概念.

如果放宽条件,允许系统在不变的温度下对外做功 $\Delta W' = -\Delta W = p\Delta V$,那么两边同时积分给出:

$$W' \leq F_{\text{initial}} - F_{\text{final}}$$

等号成立时,这就是个机械能守恒,自由能 F 具有势能的含义.而不等号的存在意味着热力学系统的不可逆性.上式又被称为**最大功原理**(principle of maximum work).虽然名称与上面两个原理相似,但这个原理不是用作一个平衡判据,而是用于实际过程,工程上往往把整个系统与外界和合自由能称为**火用**⁹(exergy).

⁸为了表述的简化我们让体系的每处都处于共同温度 T 而每个与外界交换的体积功都发生在共同压强 p 处,只要我们愿意对于较复杂的系统我们随时都可以把下面式子写为:

$$\Delta U \leq \sum_i T_i \Delta S_i - \sum_j p_j \Delta V_j$$

⁹笔者电脑字库里没有这个字 orz.

接着我们讨论吉布斯自由能 G . 同样, 吉布斯自由能的能量最小判据在等温等压的条件下被写作:

$$\Delta G \leq -S\Delta T + V\Delta p$$

也就是说, 在等温, 等压的过程中吉布斯自由能 G 总是取到它的极小值. 而等温等压的过程又常常是各种化学反应发生的初末态要符合的条件. 此式在判断化学反应的发生方向上极为有用. 在各种燃烧的模型中, 我们发现气体发生化学反应体积膨胀对外做的功不全由体积功 $p_0\Delta V$ 表示, 这是因为把外界压强(初末态平衡压强)与过程中气体压强划等号的不合理. 此时还会有其他形式的机械功输出, 要么转化为气体的动能, 要么利用起来推动活塞对外做功. 这一部分功称为非体积功 W'' , 由于 $W' = W'' + p\Delta V$, 我们通过上面的亥姆霍兹自由能的最大功原理写出等温等压下的吉布斯自由能的最大功原理:

$$W'' \leq G_{\text{initial}} - G_{\text{final}}$$

7.5.4 化学势

最后我们再考虑一下粒子数可变的系统. 对于理想气体类似的仅由 $p - V$ 决定的系统在考虑粒子数变化后将有三个独立变化的状态参量. 我们注意到粒子数 N 不变时:

$$dG = -SdT + Vdp$$

而显然 G, S, V 恰好都是广延量, 而作为自变量的 p, T 是两个强度量. 这就是说, 如果我们引入“每个粒子携带的吉布斯自由能” $\mu = \frac{G}{N}$, 同样地, 我们不假思索的把所有广延量平分到每个粒子上, 用它们的小写表示, 如 s, v, u, h, f 等等. 研究吉布斯自由能便可以发现:

$$d\mu = -sdT + vdp$$

$$dG = d(N\mu) = -SdT + Vdp + \mu dN$$

而如果研究内能便可以发现:

$$du = Tds - pdv$$

$$dU = d(Nu) = NTds - Npdv + udN = TdNs - pd(Nv) + (u - Ts + pv)dN = TdS - pdV + \mu dN$$

另外两个热力学势是类似的:

$$dF = -SdT - pdV + \mu dN$$

$$dH = TdS + Vdp + \mu dN$$

从中可以看出 μ 具有某种特殊的含义, 我们把这个量叫做**化学势**(chemical potential). 它是在准静态的等温等压过程中向体系添加一个粒子外界需要做的非体积功. 它与相之间的平衡与转化问题息息相关. 如一个在等温等压条件下自发进行的化学反应过程 $A \rightarrow B$, 我们如果不对体系做非体积功, 也不去吸收非体积功, 那么根据最小功原理吉布斯自由能必须减小. 也就是 A 的化学势必须比 B 的化学势高可以利用化学势的差对外输出机械功, 这便是内燃机的原理. 达到平衡时有 $\mu_A = \mu_B$, 此时两相之间的粒子交换已经是准静态的过程. 不会造成过程的不可逆性.

在引入化学势后对于共存的, 或多相混合的, 甚至非均匀的 $p - V$ 系统我们统一地写出热力学的基本关系式:

$$dU = TdS - pdV + \sum_{\alpha} \mu_{\alpha} dN_{\alpha}$$

上式是一个恒等式, 不再像之前的讨论那样由于缺少内部的参数而导致成为不等式. 这一点容易理解, 因为考虑到化学势实际上说明我们对不同的组分进行了区分, 考虑到了实际 $p - V$ 系统的内部构成而不是仅仅研究其宏观彻体性质. 在这一点上, 上式中各个量都是明确的微观含义, 对于平衡态和下一节要介绍的近平衡态都是可以使用的. 不会造成不相等的情形.

7.6 近平衡态热力学 *

从古至今,物理学对**非平衡态**(non-equilibrium state) 的研究兴趣丝毫不亚于对平衡态的研究.著名的**开尔文勋爵**(Lord Kelvin) 就曾试图处理不可逆过程.直到 20 世纪三四十年代才有**昂萨格**(Onsager) 等人建立了**近平衡态**(near-equilibrium) 的热力学与统计理论.而在六七十年代又由**普利高津**(Prigogine) 等人对**远离平衡态**(far from equilibrium) 的研究而大放异彩.远离平衡态的性质将在本书的讨论范围外,我们在这儿对近平衡态中的线性特性做一个简介.

7.6.1 线性输运现象

当系统处于非平衡态时,系统内部一般存在压强,温度,浓度,电势梯度.分别将引起质量¹⁰,热量,粒子与电荷的迁移,称为**输运现象**(transport phenomenon).形象地看这种因果关系,我们把各种非平衡的梯度称为**力**(force),而导致的输运称为**流**(flow).在力较小时流近似与力成正比,这个线性的区域正是近平衡态热力学所研究的区域.历史上人们相继发现了如下经验规律:

傅里叶 (Fourier) 热传导定律

$$\mathbf{J}_q = -\kappa \nabla T$$

其中 \mathbf{J}_q 表示热流密度.热交换是指通过非粒子交换,也非物质团集体运动而通过热接触,分子碰撞而交换的能量部分.而系数 κ 称为**热导率**(thermal conductivity).

菲克 (Fick) 扩散定律

$$\mathbf{J}_n = -D_0 \nabla n$$

其中 \mathbf{J}_n 表示粒子流密度.但为了热力学的统一处理我们又经常把上式写为:

$$\mathbf{J}_n = -D \nabla \mu$$

注意到 $\mu = \mu(n, T)$,一般一定温度下化学势都是与浓度正相关的,两种定义因此而能够互相换算 $D_0 = D \left(\frac{\partial \mu}{\partial n} \right)_T$,仅仅是**扩散率**(diffusivity) 的取法不同.然而如果不等温的情形如何处理?这不会带来问题,因为温度梯度也能造成粒子流,见下.

欧姆 (Ohm) 定律

$$\mathbf{J}_e = -\sigma \nabla \varphi$$

其中 \mathbf{J}_e 表示电流密度.而 σ 为导体的**电导率**(electrical conductivity).

索瑞 (Soret) 效应

温差传热与浓度差传粒子的现象存在耦合,温度差(而不是浓度差)也将导致粒子的扩散.这个效应称为**热迁移**(thermophoresis) 或**热扩散**(thermodiffusion).类似的可以由以下经验公式表示:

$$\mathbf{J}_n = -D_T \nabla T$$

¹⁰这是黏滞现象,本书将不讨论

其中 D_T 称为热扩散率(thermodiffusivity). 从而在傅里叶热传导实验中, 如图7.20左, 在温度梯度下实际上存在着热流与粒子流两种同时存在的现象.

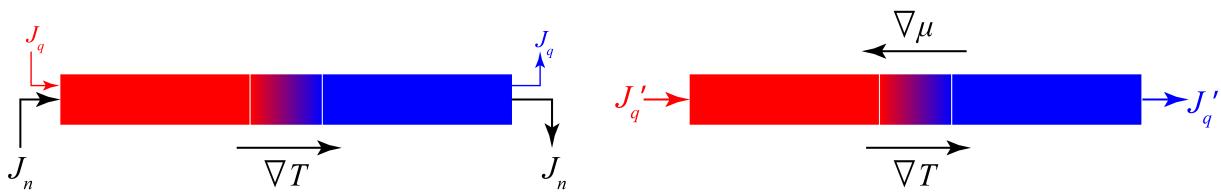


图 7.20: Soret 效应与 Dufuer 效应

杜福尔 (Dufuer) 效应

如图7.20右所示, 在傅里叶热传导实验中阻断两侧的粒子流, 那么由于热扩散现象的存在将造成两侧粒子的浓度差. 两个效应由此产生的粒子流也因此而达到平衡而抵消. 在此时对热流进行测量, 其值将不同于有粒子流时的值, 这可以被解释为由于粒子浓度梯度所造成的. 经验规律可以写为:

$$\mathbf{J}_q - \mathbf{J}'_q = \kappa_\mu \nabla \mu$$

伽伐尼 (Galvani) 接触电势差

金属内的电子体系不是一个简单的体系, 在量子理论建立后才有能带理论(energy band theory) 能给出电子性质的正确描述. 一般来讲, 不同金属内的电子可以认为具有不同的化学势 μ_i . 此时如果让两种金属 1 与 2 相互接触, 化学势高的金属, 设为 1, 其电子将自发地向化学势低的金属 2 上运动. 这直接改变的不是化学势, 而是更直接地导致了电磁学的结果—2 上电势降低而 1 上电势升高. 我们可以引入电化学势(electrochemical potential) 来表示两种驱动电子运动的效果的和, 达到平衡时应有:

$$\bar{\mu}_1 = \bar{\mu}_2 ; \quad \bar{\mu}_i = \mu_i - e\varphi_i$$

这一点也很好理解, μ 的含义为从无穷远移动一个电荷进入导体中外界所需要的功, 现在只需要把整体电荷分布所带来的电势能改变所需要的功也计及在内即可. 上式给出了两种金属的接触电势差(contact potential difference):

$$\varphi_1 - \varphi_2 = e(\mu_1 - \mu_2)$$

然而它既不能产生回路电流, 甚至也不能被直接测量: 如果把第三种金属制作的电极加在两种金属两端, 由于三个界面上都有接触电势差而使得第三种金属的两个电极上电势差为零.

塞贝克 (Seebeck) 效应

1821 年物理学家塞贝克发现由两种材料的导线构成的闭合回路如果把两个连接点置于不同的温度环境下则会使得小磁针偏转. 他把这个效应称为“热磁效应”, 这是因为当时人们对电磁现象之间的联系缺乏认识, 毕竟恰好在几乎同时奥斯特 (*Ørsted*) 才发现电流能够产生磁场. 之后奥斯特就意识到这个现象是典型的热电耦合现象, 展开了人们对热电耦合现象研究的序幕. 类似于温差传热与浓度差传粒子的耦合, 这个现象实际上是有以下经验规律:

$$\mathbf{J}_e = -\sigma \eta \nabla T$$

其中 η 被称为塞贝克系数(Seebeck coefficient), 而实际温差电现象有以下特点: 温差电动势不依赖于导线的长度, 不依赖于导线上温度的分布, 既使中间有连接第三种材料的导线, 只要连接点等温, 也不影响温差电动势的大小. 这很容易理解. 首先意识到接触电势差不会带来电动势, 其次在第 i 段导线上的非静电力为:

$$\mathbf{J}_e = \sigma \mathbf{K} \Rightarrow \mathbf{K} = -\eta_i \nabla T$$

从而两端的总电动势为:

$$\mathcal{E}_i = \int_{H_i}^{L_i} -\eta_i \nabla T \cdot d\mathbf{l} = \eta_i (T_{H_i} - T_{L_i})$$

上式以高温端向低温端为正,两个温度分别用 T_H 和 T_L 表示.通过上式发现的确这个积分与路径无关,而两端等温的导线段将不产生电动势.故如果把导线 1 和 2 连成回路,以从 1 上高温到低温,2 上低温到高温为回路正方向的回路电动势为:

$$\mathcal{E} = (\eta_1 - \eta_2)(T_H - T_L) = \eta_{12}\Delta T$$

而 η_{12} 则是实用的温度系数(temperature sensitivity).塞贝克系数是可正可负的,类似于热扩散现象,如果载流子为正电荷,如 p 型半导体,那么热扩散导致电荷从高温端向低温端迁移,系数为正.如果载流子为负电荷,如 n 型半导体,相反的系数就为负.金属的塞贝克系数就都是负的.由此现象制作的热电偶(thermocouple)是很好的宽温度范围测温仪器.下图 7.21 左演示了其半导体型热电偶的工作原理.

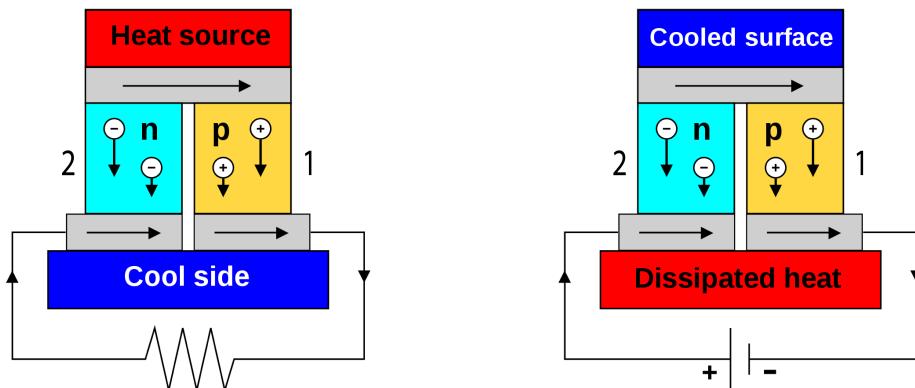


图 7.21: Seebeck 效应与 Peltier 效应

佩尔捷 (Peltier) 效应

类似于索瑞效应与杜福尔效应的关系,温差可以耦合电流,那么这个效应也会反过来对热传导造成影响.第一个效应是佩尔捷效应,当电流 I 以其正向流过以上连接点处时,连接点会产生一部分独立于热传导外的吸放热.这个热量被称为佩尔捷热,以电流从 1 到 2 为正,热量产生为正,则经验公式为:

$$\dot{Q} = (\pi_1 - \pi_2)I$$

其中 π_i 为材料 i 的佩尔捷系数.

汤姆孙 (Thomson) 效应

汤姆孙也就是开尔文爵士,他在 1851 年前后系统地研究了热电耦合的两个效应,并在理论上预言和实验上观测了连续版本的佩尔捷效应---即汤姆孙效应.这个效应只需要把上式中的热产生率换成热产生率密度,而佩尔捷系数之差理解为某种和温度有关的性质的梯度,电流改为电流密度,得到:

$$\dot{q} = -k \mathbf{J} \cdot \nabla T$$

这样一些线性现象可以代表在统计力学方法出现前人们对近平衡态研究成果.

7.6.2 传热的熵产生

我们先建立简单的热传导过程的热力学理论。在热传导的过程中每一处的熵增都是由于左右两侧的传热造成的，根据热力学第一定律，单位体积内的内能增加为：

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{J}_q$$

而凡是内能增加即体现为熵增：

$$\delta u = \delta q = T \delta s$$

那么熵的等式可以化为：

$$\frac{\partial s}{\partial t} = -\frac{1}{T} \nabla \cdot \mathbf{J}_q$$

至此已足够明显，如果定义**熵流**(entropy flow)，它由其他流共同体现，在热传导时热流即也携带熵流：

$$\mathbf{J}_s = \frac{\mathbf{J}_q}{T}$$

那么熵流本身不能完全反应每一点处熵的增加，还要引入**熵产生**(entropy production) 的概念，比如在此种情形下为：

$$\theta = \mathbf{J}_q \cdot \nabla \frac{1}{T}$$

那么有以下熵连续性方程：

$$\frac{\partial s}{\partial t} = -\nabla \cdot \mathbf{J}_s + \theta$$

7.6.3 普遍理论与昂萨格倒易关系

一般来讲，熵密度的改变往往与体积内的某种“荷” q_α 与“势” f_α 有关，一般可以表示为：

$$\delta s = \sum_{\alpha} f_{\alpha} \delta q_{\alpha}$$

而每一种“荷”都有其连续性方程：

$$\frac{\partial q_{\alpha}}{\partial t} = -\nabla \cdot \mathbf{J}_{\alpha}$$

从而一样地我们写出：

$$\frac{\partial s}{\partial t} = -\nabla \cdot \mathbf{J}_s + \theta$$

$$\mathbf{J}_s = \sum_{\alpha} f_{\alpha} \mathbf{J}_{\alpha} ; \quad \theta = \sum_{\alpha} \mathbf{J}_{\alpha} \cdot \nabla f_{\alpha}$$

从而流就是 \mathbf{J}_{α} 而力就是 $F_{\alpha} = \nabla f_{\alpha}$ ，那么近平衡区的线性输运特性可以表示为：

$$\mathbf{J}_{\alpha} = \sum_{\beta} L_{\alpha\beta} \nabla f_{\beta}$$

中间的经验系数构成的矩阵就是**昂萨格矩阵**(Onsager matrix)。通过对微观过程可逆性的研究，昂萨格断言：

$$L_{\alpha\beta} = L_{\beta\alpha}$$

即这个矩阵是一个对称的(半正定)矩阵，任意两种流与力之间的交叉耦合强度实际上应该相等。这就是**昂萨格倒易关系**(Onsager reciprocal relation)，由于其普适性，它经常被当成热力学第四定律¹¹。从而整体熵产生被表示为一个二次型：

$$\theta = \sum_{\alpha\beta} L_{\alpha\beta} \nabla f_{\alpha} \cdot \nabla f_{\beta}$$

¹¹本书将不涉及热力学第三定律

7.6.4 推证热电耦合的普遍规律

热电耦合问题中存在两种流,一是热流 \mathbf{J}_q ,定义为单位面积单位时间内通过的热量.二是电流 \mathbf{J}_p ,我们令载流子为正电荷,电量为 p ,而把化学势约化为单位电荷的: $\zeta = \mu/p$ 则 μn 项变为 $\zeta \rho \cdot \rho$ 为电荷密度.在导体导电问题中有一些不同于普遍理论之处:一是有外电场的存在,它会对流做功而直接导致输运,二是化学势的变化一般可以忽略,而仅在两种导体接触处才需要考虑.我们写出此时的内能连续方程:

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{J}_q - \nabla \cdot (\zeta \mathbf{J}_p) - \mathbf{J}_p \cdot \nabla \varphi$$

第一项是传热输入的能量,第二项是传入粒子所带来的能量,第三项还有外电场 $\mathbf{E} = -\nabla \varphi$ 对电流所做的功.而电荷连续方程写为:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}_p$$

熵的大小为:

$$\delta s = \frac{1}{T} \delta u - \frac{\zeta}{T} \delta \rho$$

代入化简得:

$$\begin{aligned} \frac{\partial s}{\partial t} &= -\frac{1}{T} \nabla \cdot \mathbf{J}_q - \frac{1}{T} \mathbf{J}_p \cdot \nabla (\zeta + \varphi) \\ &= -\nabla \cdot \left(\frac{1}{T} \mathbf{J}_q \right) + [\mathbf{J}_q + (\zeta + \varphi) \mathbf{J}_p] \cdot \nabla \frac{1}{T} - \mathbf{J}_p \cdot \nabla \frac{\zeta + \varphi}{T} \\ &= -\nabla \cdot (f_1 \mathbf{J}_1 + f_2 \mathbf{J}_2) + \mathbf{J}_1 \cdot \nabla f_1 + \mathbf{J}_2 \cdot \nabla f_2 \\ &= -\nabla \cdot \mathbf{J}_s + \theta \end{aligned}$$

其中第一个流是能量量纲的 $\mathbf{J}_1 = \mathbf{J}_q + (\zeta + \varphi) \mathbf{J}_p$,对应的“势”为 $f_1 = 1/T$;第二个流是电荷量纲的 $\mathbf{J}_1 = \mathbf{J}_p$,对应的“势”为温度约化的电化学势 $f_2 = -(\zeta + \varphi)/T$.这样我们写出昂萨格线性关系:

$$\begin{aligned} \mathbf{J}_q + (\zeta + \varphi) \mathbf{J}_p &= L_1 \nabla \frac{1}{T} - M \nabla \frac{\zeta + \varphi}{T} \\ \mathbf{J}_p &= M \nabla \frac{1}{T} - L_2 \nabla \frac{\zeta + \varphi}{T} \end{aligned}$$

对应欧姆定律,命上式温度梯度为零,在一种导体中 ζ 始终认为是一个常数,有:

$$L_2 = \sigma T$$

而对应温差电效应,我们消除外电场,让 φ 是一个常数,用塞贝克系数可以确定 M :

$$M = \sigma T (\eta T + \zeta + \varphi)$$

最后再由傅里叶定律,让电流 \mathbf{J}_p 为零,得到:

$$L_1 = \kappa T^2 + \sigma T (\eta T + \zeta + \varphi)^2$$

最后整理得:

$$\begin{aligned} \mathbf{J}_q &= -(\kappa + \sigma \eta^2 T) \nabla T - \sigma \eta T \nabla (\zeta + \varphi) \\ \mathbf{J}_p &= -\sigma \eta \nabla T - \sigma \nabla (\zeta + \varphi) \end{aligned}$$

首先考虑伽伐尼接触电势差,显然当且仅当上式中两个梯度为零时所有流都将消失.从而相互接触的两种介质有

$$\zeta_1 + \varphi_1 = \zeta_2 + \varphi_2 \implies \varphi_1 - \varphi_2 = -\frac{1}{p} (\mu_1 - \mu_2)$$

如果是金属 $p = -e$ 我们便得到了之前熟知的结果. 再考虑佩尔捷效应, 我们依然认为两介质满足以上式子, 这是因为此时由 $\nabla\varphi$ 驱动的电流应该认为是发生在两种介质中, 而分界面上仍是这个固定的接触电势差. 那么对于电流从介质 1 流向 2 的情况, 再命上式中温度均匀不变, 有:

$$\mathbf{J}_{q,i} = \eta_i T \mathbf{J}_{p,i}$$

也就是说电流恰好携带着一个热流. 由内能连续性:

$$\dot{Q} = \frac{dU}{dt} = \eta_1 T I - \eta_2 T I + (\zeta_1 + \varphi_1) I - (\zeta_2 + \varphi_2) I$$

后两项表示输入输出的电流粒子带来的能量与电场力的做功, 这两项恰好消掉了. 于是得到了佩尔捷效应的公式, 并发现佩尔捷系数实际上就是:

$$\pi = \eta T$$

这是开尔文爵士当年得出的两个关系之一, 称之为第二汤姆孙关系(the second Thomson relation). 我们再研究连续情形的内能连续性, 再次使用之前的表达式:

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{J}_q - \nabla \cdot (\zeta \mathbf{J}_p) - \mathbf{J}_p \cdot \nabla \varphi$$

首先根据昂萨格关系, 注意到任意情况下都有:

$$\mathbf{J}_q = \eta T \mathbf{J}_p - \kappa \nabla T$$

再注意到在我们要研究的问题中电流仍连续:

$$\nabla \cdot \mathbf{J}_p = 0$$

代入上式, 通过昂萨格关系化简便得:

$$\dot{q} = \nabla \cdot (\kappa \nabla T) + \frac{\mathbf{J}_p^2}{\sigma} - T \mathbf{J}_p \cdot \nabla \eta$$

上式第一项表示热传导积累的热, 第二项是焦耳热, 第三项就是汤姆孙热了. 可见它是由于塞贝克系数 η 不均匀而导致的, 一般就是其与温度有关¹², 温度梯度也造成了一个塞贝克系数的梯度 $\nabla\eta = \frac{d\eta}{dT} \nabla T$, 从而我们也得到了开尔文系数:

$$k = T \frac{d\eta}{dT} = \frac{d\pi}{dT} - \frac{\pi}{T}$$

上式即被称作第一汤姆孙关系(the first Thomson relation).

这就是近平衡态线性热力学理论对热电耦合现象的解释大意.

¹²既然 η 可以随温度改变, 那么 κ 随温度的改变其实也应考虑在内. 当然这一项是纯导热现象而与电流没有耦合.

第八章 液体与固体的性质

在讲解热力学基本规律的过程中, 我们已经对气体, 尤其是理想气体这样的热力学体系有了一定的了解. 但是对于与我们的生活息息相关的液体(liquid) 与 固体(solid) 两种凝聚态(condensed matter) 我们还知之甚少. 在此稍作一步展开介绍.

8.1 固体晶格论

8.1.1 经典晶格论

固体在此处实际上特指晶体(crystal). 晶体应具有长程序(long-ranged order), 也就是说组成的晶体的原子, 分子, 或者离子应体现出大范围的空间周期性. 如图8.1所示, 一整块物质如果完全由严格的周期性结构组成, 那么一般其表面也是规则的几何形状, 称作单晶体(single crystal). 但是即使单晶体内部是充满着缺陷(defect) 的, 无论是原子的缺失(vacancy) 还是填隙(interstitial), 还是更复杂的面缺陷与体缺陷, 都会使得晶体的不同区域变得孤立起来, 常见的金属, 冰, 陶土, 岩石都是由大量可视作晶体的晶粒拼接而成, 在物理性质上丧失了单晶体所具有的各向异性而成为多晶体(precrystal). 最后还有丧失了长程序的无定形体(amorphous solid), 又做非晶体, 玻璃体. 比如玻璃, 蜡, 塑料等, 它们都不是晶体, 但一般仍认为是固体.

固体的结合方式在极大程度上决定了其热学性质, 而固体的结合又在很大程度上依赖于其外层的电子在量子统计规律下的诸多行为. 根据最后的结果我们可以粗浅地分类:

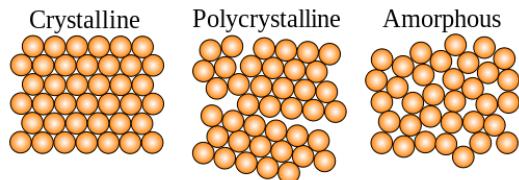


图 8.1: 单晶体, 多晶体与无定形体

| 晶体类型 | 结合方式 |
|------|---|
| 原子晶体 | 局域在原子间的共用价电子对与原子实之间的相互吸引 |
| 离子晶体 | 一种原子的电子几乎完全被另一种原子剥夺形成离子, 两种离子相互吸引 |
| 金属晶体 | 非局域的共享电子们与原子实们之间的相互吸引 |
| 分子晶体 | 分子与分子之间十分微弱的氢键, 范氏相互吸引, 还包含电偶极矩磁偶极矩相互作用 |

在某些较为简单的情形下, 我们会认为固体可以拆解为两个体系, 一是由原子实(atomic core) 构成的点阵(lattice), 一是参与原子间相互作用的所有价电子(valence electron) 构成的电子气(electron gas). 后者是那么地轻盈, 但是却带有与原子实等量反号的电荷, 这一点造成了如下的局面: 体系的动能几乎由具有绝大多数质量的点阵承担¹; 而势能也由点阵的位置与变形来计算, 对于电子的分布, 总是遵循泡利不相容原理和

¹其实这也是个量子效应, 详见之后章节.

最小势能原理,故可由点阵位置计算.这种平均场论(mean field theory)的思想使我们写出体系的总能量来:

$$H = \sum_g T(\mathbf{v}_{i,g}) + V(\mathbf{r}_{i,g}) \quad , \quad i = 1, 2 \dots n ; g \in G$$

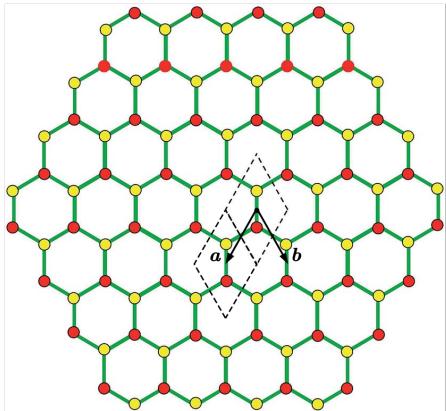


图 8.2: 石墨烯的结构

式中 G 表示所有原胞(unit cell)的集合,将一个原胞按照 G 的方式平移便得到了整个晶体,称作布拉伐格子(Bravais lattice).而每一个原胞都是全同的,内含 $1, 2 \dots n$ 号原子.以石墨烯(graphene)为例,石墨(graphite)实际上即由一层一层的石墨烯通过范德瓦尔斯力吸引而形成.每一层可以认为是一种平面型的原子晶体.其原胞由两个并不具有平移等价性的原子构成.原胞形状为菱形,故可以在 $x-y$ 平面上按密铺,具有平移对称性,如果我们把沿两个 $x-y$ 平面上独立方向的对称平移矢量 a, b 找到,那么所有平移对称性就能够被记做:

$$G = \{p\mathbf{a} + q\mathbf{b} | p, q \in \mathbb{Z}\}$$

从而能量的一种最简单的写法就是:

$$\begin{aligned} H = & \sum_{pq} \frac{1}{2} m (\dot{x}_{1,pq}^2 + \dot{y}_{1,pq}^2 + \dot{x}_{2,pq}^2 + \dot{y}_{2,pq}^2) + \frac{1}{2} k [(x_{1,pq} - X_{1,pq})^2 + (y_{1,pq} - Y_{1,pq})^2 \\ & + (x_{2,pq} - X_{2,pq})^2 + (y_{2,pq} - Y_{2,pq})^2] + \frac{1}{2} m' (\dot{z}_{1,pq}^2 + \dot{z}_{2,pq}^2) + \frac{1}{2} k' (z_{1,pq}^2 + z_{2,pq}^2) \end{aligned}$$

上式 1 表示上图中每一个原胞中的黄色原子的动能和势能,2 则表示红色原子的能量.这背后的原理是,首先动能自然是正比于速度的平方的,而势能在平衡位置 $(X_{i,pq}, Y_{i,pq}, 0)$ 附近泰勒展开也至少是从二阶项开始.由于只考虑原子偏离平衡位置不多的情况,所以只保留最低阶的项.所以都是二次型.由于红色黄色原子在中心反演下对称,故两者系数 m, k 也是相等的.而在平面上的能量具有三重旋转对称性,具有这样的对称性的二次型只能在所有方向上都具有相同的系数,故连 x, y 的系数都是一样的.但是 z 方向的运动则可以具有不一样的系数 m', k' .

这样极端的近似存在一个致命的缺陷:忽略了格点与格点之间的相互作用,下文我们将采取一种弹性原子链模型来弥补这个缺陷.但这个处理足以带来第一个关键的信息:我们发现上面一共具有 12 个平方项,这就是说,根据之前说过的,之后还将讨论的能均分定理,该单元的总平均能量为:

$$\varepsilon_{\text{cell}} = \frac{1}{2} kT \cdot 12 = 6kT$$

或者说对于这样的原子晶体(离子晶体,金属晶体也是类似的),其每一个原子上的平均能量就是:

$$\varepsilon = 3kT$$

这其实是一个普遍的性质,任何原胞里面如果具有 n 个原子,那么对应的动能项和势能项就会各有 $3n$ 个,相当于每个原子平均具有三个动能项和势能项,对应三个运动的维度.从而平均能量就是 $3kT$.那么整个晶体的总内能就是:

$$U = 3NkT = 3\nu RT$$

即晶体的摩尔热容是:

$$C_m = 3R$$

这是杜隆-裴替定律(Dulong-Petit law).有四点值得注意:

- 以上摩尔热容指的是每一摩尔原子的热容.例如 NaCl 晶体,每一摩尔的 NaCl 应当按 2 摩尔原子计算.

2. 以上内能只计算了振动导致的热运动形式的能量. 晶格自己的结合能自然也是内能的一部分, 被我们直接处理为不变的常数而从“有效”内能中略去了. 如果考虑热膨胀甚至晶格的解离, 如熔化或溶解的过程, 显然这个过程内能的改变就会用到这一部分. 另, 对于固体来说等体热容和等压热容的差别很小可以忽略不计, 因为固体温度不同的体积膨胀与气体相比实在太细微.

3. 电子对热容的贡献被忽略了, 这个做法甚至是合理的. 背后有一定的量子力学原理: 电子被量子效应束缚在定态上: 它们要么是束缚电子处于原子轨道中, 要么是价电子处于分子轨道上, 甚至金属中的自由电子也其实是在一定的能带结构中, 本质上还是被所有原子实束缚. 故计算表明室温下这些电子产生的热容是可以忽略不计的. 在室温下, 即使是金属, 它们的热容与杜隆-裴替定律符合地也是相当地好的. 但是在较高的费米温度(Fermi temperature)量级开始, 其热容明显大于 $3R$. 量级基本上在几万开尔文. 这个温度下基本上物质也称为等离子体态而不是固态了.

4. 最后, 晶格热容低温下杜隆-裴替定律也会失效. 也是与量子效应有关. 典型的温度称作德拜温度(Debye temperature). 明显低于德拜温度时晶格热容将迅速减小. 常见金属的德拜温度在室温附近, 而非金属单质的德拜温度往往较低. 如果温度特别低, 那么可以证明晶格的热容将正比于温度的3次方, 而电子造成的热容将正比于温度的1次方. 之后我们会做简要讨论.

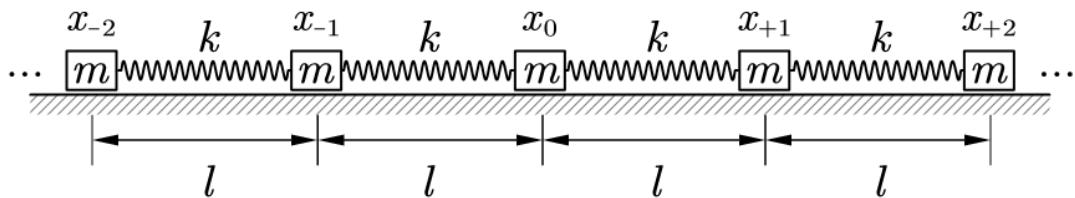


图 8.3: 弹性

认为固体单元之间有十分简单的平方形式的弹性势能, 这是近似但实用的. 正如弹性(elasticity)的概念正发源于其中, 在简化的一维原子链模型下, 正如力学的弹性体中所描述的那样, 这一个体系就会产生弹性. 其纵向的弹性模量为:

$$F = K \frac{\Delta L}{L} \Rightarrow K = kl$$

那么沿三维晶体的一条原子链方向的拉伸压缩形变的杨氏模量, 就可以不严谨地认为是以上的 K 除以每一条原子链“占据”的截面积. 但是这个模型有一个缺点: 温度升高以后原子仅仅是围绕既定的平衡位置做热振动, 从而就无法产生正常固体具有的热膨胀(thermal expansion)现象.

为了解释热膨胀现象我们就需要对既定的模型进行适度的扩充. 其实热膨胀可以简单地来自于其相互作用的非谐性(anharmonicity). 简谐的受力是线性回复的, 简谐的势能是至多二次的多项式的形式, 势能函数的图像是标准的抛物线. 但是只要在势能函数中最小增添一个三次项, 就可以导致势能的非谐性. 对于任意非线性力的势能函数组成的一维原子链, 热膨胀其实就来自这个势能的非零三阶导数. 例如, 我们依然认为一维原子链中, 只有相邻的两个原子之间才有相互作用, 例如形式为李纳-琼斯势能(Lennard-Jones potential)²:

$$V(u) = V_0 \left[-2 \left(\frac{u_0}{u} \right)^6 + \left(\frac{u_0}{u} \right)^{12} \right]$$

这样在平衡的时候(无热振动), 相邻原子之间的距离为 $l = u_0$. 即势能曲线的最低点对应的距离. 在该点处, 该势能的二阶展开给出了这个模型对应的简谐模型中的劲度系数 k :

$$V = V(u_0) + \frac{1}{2} V''(u_0)(u - u_0)^2, \quad k = V''(u_0) = 72 \frac{V_0}{u_0^2}$$

²然而这个势能适用于不带电、无偶极矩、不成化学键的两个原子或分子之间的作用力. 其他的力一般更强而具有更低的幂次.

现在 $u - u_0$ 表示的是两个质点的间距. 现在用 x 表示它, 如果势能仅仅是上式给出的:

$$V = -V_0 + \frac{1}{2}kx^2$$

这样的势能就不会产生热膨胀的效果, 经典物理的观点是, 如果体系以一定形式在这样一个势能平衡位置处振动, 往往往右位移的幅度和停留时间都是对称的, 那么平均起来相当于两原子平均距离没有改变. 其实更精确的观点应当是: 其原子间的距离应当满足一个玻尔兹曼分布. 现在让我们考虑泰勒展开的三阶项:

$$V'''(u_0) = -1512 \frac{V_0}{u_0^3}$$

即势能变为:

$$V = -V_0 + \frac{1}{2}kx^2 + \frac{V'''(u_0)}{6}x^3 = -V_0 + cx^2 - gx^3, \quad g = 252 \frac{V_0}{u_0^3}$$

根据玻尔兹曼分布 (之后章节会有介绍) 的理论, 分子之间的平均距离对应的 x 的计算公式应当是:

$$\langle x \rangle = \frac{\int xe^{-V/k_B T} dx}{\int e^{-V/k_B T} dx}$$

在计算中做如下近似:

$$e^{gx^3/k_B T} \approx 1 + \frac{gx^3}{k_B T}$$

得到以上积分为:

$$\langle x \rangle = \frac{\frac{3\pi^{1/2}(k_B T)^{3/2}g}{4c^{5/2}}}{\frac{\pi^{1/2}(k_B T)^{1/2}}{c^{1/2}}} = \frac{3g}{4c^2} k_B T$$

这就可以给出热膨胀系数 (thermal expansion coefficient) 的值:

$$\frac{\Delta L}{L} = \alpha T \Rightarrow \alpha = \frac{3k_B g}{4u_0 c^2}$$

这个热膨胀系数实际上是线膨胀系数, 体膨胀系数, 由小量近似方法容易得到, 应是线膨胀系数的三倍. 在温度变化范围不大的情况下, 这两个系数根据公式都是常数, 所以固体的长度和体积分别按照以下公式变化:

$$\alpha_V = 3\alpha_L = 3\alpha$$

$$L = L_0(1 + \alpha T), \quad V = V_0(1 + \alpha T)$$

其中 L_0, V_0 为假想的外推到绝对零度时的原长, 原体积.

类似于热容的性质, 把简谐的势能改为非谐的情况下, 尽管产生了一个热膨胀系数, 但是它其实完全是个常数. 在温度较低的情况下, 热容的杜隆-裴替定律和线膨胀系数的上理论公式将会同时失效.

8.1.2 量子晶格论 *

此前在力学中我们讨论过两个模型:

一是连续介质的弹性波模型. 在一个杨氏模量为 E 的密度为 ρ 的无限长的自由棒³的弹性运动模式中, 值得注意的是以下基础的行波解:

$$\xi(t) = A \cos(\omega t - \kappa x + \varphi), \quad \frac{\omega}{\kappa} = \sqrt{\frac{E}{\rho}}$$

³我们也说过, 如果是连续的三维介质, 一个波矢 κ 下就存在更复杂的三个振动模式, 纵波与其平行, 横波与其垂直. 其波速依赖于材料剪切模量 G 和体弹性模量 K :

$$v_P = \sqrt{\frac{K + 4G/3}{\rho}}, \quad v_S = \sqrt{\frac{G}{\rho}}$$

二是考虑宏观物质实际上是由离散的单元所构成的格波模型. 我们也得到了以下行波解:

$$x_n(t) = A \cos(\omega t - n\kappa l + \varphi) , \quad \omega^2 = \frac{4k}{m} \sin^2 \frac{\kappa l}{2}$$

在两个问题中, ω, κ 都是满足色散关系下可连续取值的. 而研究本征模式可替代研究单个质元或格点的运动, 因为整个体系的运动除了看做所有单个质元或格点的运动的组合, 也可以分解为本征模式的叠加:

$$x_n(0) = \int_{-\pi/l}^{\pi/l} [A(\kappa) \cos n\kappa l + B(\kappa) \sin n\kappa l] d\kappa$$

上式其实是一种离散傅里叶变换(discrete fourier transform, DFT)⁴, 这一点在接下来的分析中将会变得明了. 其中 κ 的取值范围在第一布里渊区(first Brillouin zone):

$$-\frac{\pi}{l} < \kappa \leq \frac{\pi}{l}$$

这是因为 $\kappa + j \cdot 2\pi/l$ 与 κ 对应的两种振动模式没有任何区别: 相邻两个格点振动的相位差只是多了 2π 的整数倍.

出于三个考虑: a. 以上离散傅里叶变换的 $A(\kappa), B(\kappa)$ 算法并不直观. b. 真实热学体系并不是无限长原子链, 而是具有粒子数 N . c. 虽然真实热学体系总是存在复杂的边界条件, 但是在远离边界的区域选取一个粒子数为 N 的子系统时, 边界与内部变得没有区别. 综合考虑我们重新建立一个类似于格波模型的有限原子链模型:

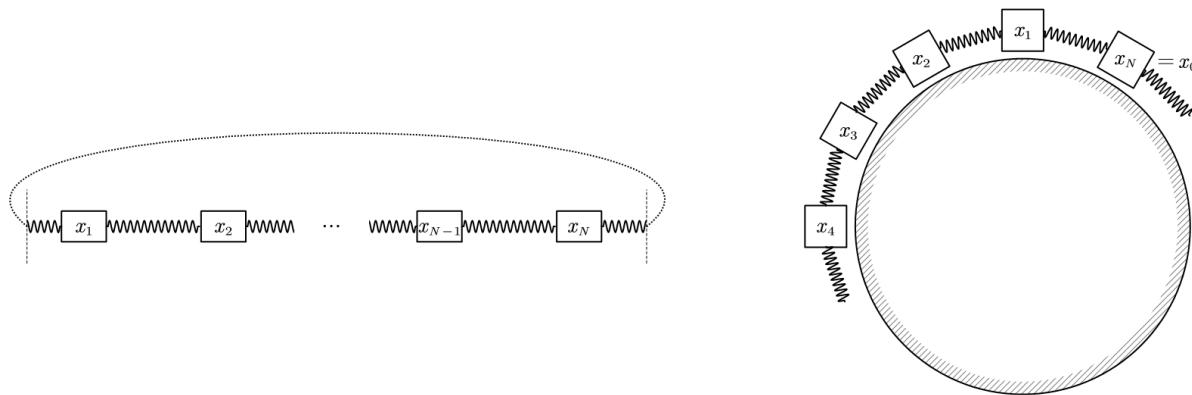


图 8.4: 周期性边界条件

如上图左, N 个原子排成链状. 并给予其周期性边界条件:

$$x_0 = x_N$$

物理上可以理解为这 N 个原子绕成了环状. 如果 N 很大, 这样的体系就能反映一般的热学链条内部的性质. 在这个条件下, κ 的取值就变得有限:

$$N\kappa l = j \cdot 2\pi \quad j = 1, 2, 3 \cdots N$$

故总共只有 N 种可能的振动模式, 第 j 种模式下:

$$x_n(t) = A_j \cos(\omega_j t - n\kappa_j l + \varphi) , \quad \omega_j^2 = \frac{4k}{m} \sin^2 \frac{\kappa_j l}{2}$$

为了使得推导更简洁, 引入复数表示法. 将以上模式叠加计算系数:

$$x_n(0) = \sum_{j=1}^N A_j e^{-2\pi i \frac{n j}{N}}$$

⁴更准确的来讲是 DTFT

十分关键的一点是, 我们不是单纯把上式看做零时刻的坐标确定每一种模式的强度系数 A_j 的决定式. 而是直接看做一种任意时刻都恒成立的坐标变换:

$$x_n = \sum_{j=1}^N q_j e^{-2\pi i \frac{n_j}{N}}$$

这个变换就是真正意义上的离散傅里叶变换. 它的逆变换与通常的傅里叶逆变换形式类似. 只需要计算:

$$\sum_{n=1}^N x_n e^{2\pi i \frac{n_j}{N}} = \sum_{n=1}^N \sum_{j=1}^N q_j e^{-2\pi i \frac{n_j}{N}} e^{2\pi i \frac{n_j}{N}} = \sum_{j=1}^N \left(\sum_{n=1}^N e^{2\pi i \frac{n(j'-j)}{N}} \right) q_j$$

括号内的求和式在 $1 \leq j' \leq N$ 时仅有 $j = j'$ 才不为零, 此时指数函数取 1 求和 N 次. 故:

$$\sum_{n=1}^N x_n e^{2\pi i \frac{n_j}{N}} = N q_j$$

从而得到此前坐标变换的逆变换:

$$q_j = \frac{1}{N} \sum_{n=1}^N x_n e^{2\pi i \frac{n_j}{N}}$$

这个变换可不止是 $q_j(0)$ 表示体系振动分解为本征模式的一个系数的作用那么简单, 它可以直接改写这个体系的动能和势能函数. 首先将动能和势能形式改写成:

$$T = \sum_{n=1}^N \frac{1}{2} m \dot{x}_n^* \dot{x}_n$$

$$V = \sum_{n=1}^N \frac{1}{2} k(x_n^* - x_{n-1}^*)(x_n - x_{n-1}) = \frac{1}{2} \sum_{n=1}^N 2kx_n^* x_n - \frac{1}{2} \sum_{n=1}^N k(x_n^* x_{n-1} + x_{n-1}^* x_n)$$

代入以上变换并化简得到:

$$T = \frac{1}{2} \sum_{j=1}^N N m \dot{q}_j^* \dot{q}_j$$

$$V = \frac{1}{2} \sum_{j=1}^N 2Nk \left(1 - \cos \frac{2\pi j}{N} \right) q_j^* q_j = \frac{1}{2} \sum_{j=1}^N 4Nk \sin^2 \frac{\pi j}{N} q_j^* q_j$$

这就可以直接看出. 整个体系变成了 $j = 1, 2 \dots N$ 对应的谐振子的直接合成:

$$H = \sum_{j=1}^N \frac{1}{2} N m (|\dot{q}_j|^2 + \omega_j^2 |q_j|^2) , \quad \omega_j^2 = \frac{4k}{m} \sin^2 \frac{\pi j}{N} , \quad \kappa_j = \frac{2\pi j}{Nl}$$

所以, 通过令 $N \rightarrow \infty$. 我们就可以得到此前的无限长原子链方案对应的模式系数与其确定方法:

$$x_n(0) = \int_{-\pi/l}^{\pi/l} [A(\kappa) \cos n\kappa l + B(\kappa) \sin n\kappa l] d\kappa$$

$$A(\kappa) = \frac{l}{2\pi} \sum_{n=-\infty}^{+\infty} x_n(0) \cos n\kappa l ; \quad B(\kappa) = \frac{l}{2\pi} \sum_{n=-\infty}^{+\infty} x_n(0) \sin n\kappa l$$

至此更值得我们注意的反而是 N 个粒子组成的满足周期性边界条件的情况. 我们通过坐标变换将初始的粒子坐标代换为 N 个广义坐标 q_j . 这样体系的能量直接变成了 N 个谐振子项:

$$H = \sum_{j=1}^N \frac{1}{2} N m (|\dot{q}_j|^2 + \omega_j^2 |q_j|^2)$$

那么容易看出每个谐振子的广义坐标就会随时间做简谐振动:

$$q_j = A_j e^{i\omega_j t} + A'_j e^{-i\omega_j t}$$

这里的“谐振子”，已经从经典物理中的谐振子概念迈进了更抽象的层面。经典物理的谐振子 $H = mq^2/2 + kq^2/2$ ，系数 m, k 通常有所指，而且指向某个具体的物理实体的属性。比如单摆若视作谐振子，摆角视作广义坐标 $q = \theta$ ，则系数 m 指单摆对悬挂点的转动惯量，而 k 值则是单摆质量，摆长与重力加速度三者的乘积。我们这里通过坐标变换得到的 N 个谐振子，每一个谐振子代表的是**集团行为**(aggregate behavior)，从因果性上看，任何一个格点的运动将对相邻的格点产生动力学影响，从而从来都无法看成是独立的振子。但是分解为广义坐标对应的谐振子后，振子与振子之间的影响就不再存在了。

基于这一点，从经典的热学理论就可以判定， N 个振子上的动能和势能就是体系的内能。但又由于能均分原理，这个内能就是：

$$U = N \cdot 2 \cdot \frac{1}{2}kT = NkT = \nu RT$$

这里是一维原子链。如果考虑三维情形，那么上式就回到了此前提过的杜隆-裴替定律：

$$U = 3\nu RT$$

但这个坐标变换处理的意义绝对不仅限于在考虑原子间相互作用的情形下得到了一个与不考虑相互作用而采取平均场论的思想得到的结果一致的结果，它的下一步是考虑能量尺度较低下的**元激发**(elementary excitation)。

我们应当已经知道，在氢原子的电子能量函数下：

$$H = \frac{\mathbf{p}^2}{2m} - \frac{e^2}{4\pi\epsilon_0 r}$$

可以得到其角动量和能量的量子化：

$$L_z = m\hbar \quad , \quad m = 0, \pm 1, \pm 2 \dots$$

$$E = -\frac{E_0}{n^2} \quad , \quad n = 1, 2, 3 \dots$$

其背后的物理规律为某在 \hbar 的尺度下才发挥显著功能的量子定律⁵。那么容易想见，虽然经典谐振子能量连续。但是如果考虑量子效应，一个量子谐振子：

$$H = \frac{1}{2}m(\dot{q}^2 + \omega^2 q^2) \quad , \quad q \text{ 存在不确定性}$$

只要其振动能量足够小，就也可以看到其能级的量子化现象。量子理论的推导给出了一个非常简洁的公式⁶：

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega$$

这就是说，谐振子的振动幅度不是连续的而是离散的⁷，不同的振动幅度之间只能以一份一份的形式叠加运动剧烈程度，每一份叠加的能量子我们就称作元激发：

$$E = \hbar\omega$$

在原子链中的元激发更值得留意，它是集团模式的元激发，每一份元激发代表了某种在总质量 M ，总长 L 的 N 原子链中传播的行波。而且由于：

$$H = \sum_{j=1}^N H_j \quad , \quad H_j = \frac{1}{2}M(|\dot{q}_j|^2 + \omega_j^2|q_j|^2) \quad , \quad \kappa_j = \frac{2\pi j}{L}$$

故这样的元激发一共有 N 种。每一种都具有具体的波矢 κ_j 与角频率 ω_j 。引人注意的是，如果计算元激发的总动量，不难发现：

$$p_j = \hbar\kappa_j$$

⁵暗示薛定谔方程。

⁶思考：如果用相空间面积量子化是否可以得到相同的结果？

⁷准确的说法是振动幅度是不确定的，但造成了离散的能级

单个元激发具有动量和能量，就好像水面上的一朵浪花，给人以类似移动的物体的直觉，当然真正在运动的是造成元激发的物质粒子。但是固体中物质粒子却不能像元激发那样长途迁徙。能否把单个元激发也形式地视作某种粒子？这种观点其实是被物理学普遍采用的。这样的粒子称作准粒子(quasiparticle)。它的含义是量子化的集团运动模式：

$$E = \hbar\omega , \quad \mathbf{p} = \hbar\boldsymbol{\kappa}$$

对于波来说，一个重要的结论为，如果考虑波包的运动，其中心速度应为群速度：

$$v_g = \frac{d\omega}{d\kappa}$$

这个式子即色散关系，它在量子理论下将具有更加直白的含义。我们考虑一个向前以 v_g 传播的波包，并把它看做元激发或若干元激发的相干态。那么作为准粒子，根据元激发具有的能量动量公式，比较不同 ω, κ 的准粒子就有：

$$v_g = \frac{dE}{dp}$$

甚至可以写为矢量式：

$$dE = \mathbf{v}_g \cdot d\mathbf{p}$$

如何理解这个式子？设想在准粒子上作用一个力 \mathbf{F} ，时间为 dt 。那么力为准粒子带来的动量和动能改变就分别可以参考经典的动量定律和动能定律：

$$d\mathbf{p} = \mathbf{F} dt$$

$$dE = \mathbf{F} \cdot \mathbf{v}_g dt$$

上式代入下式就可以得到：

$$dE = \mathbf{v}_g \cdot d\mathbf{p}$$

从这个推导来看，把元激发作为准粒子看是贴切的。它继承了经典物理中关于粒子的一些关键性质。

从原子链推广到固体。固体内部晶格振动分解为本征模式行波后的元激发，我们把这样的准粒子称为声子(photon)。

声波，连续介质之体波也。声子就是固体⁸中声波的量子。固体的声速，其实就是声子的群速度，它符合色散关系：

$$v_g = \frac{dE}{dp}$$

固体中传播声波时，完全可以说它看做大量声子在固体中做 v_g 的匀速直线运动。注意我们尚未考虑声子间的相互作用⁹，故声子运动过程是没有任何阻力的。这样热学固体系统就十分奇妙地简化为了一个由声子构成的理想气体系统。

然而话又反过来说，声子系统与理想气体系统还是有着不小的差异的：

理想气体，由 N 个粒子构成，每一个粒子 3 个平动自由度，还可能有内禀自由度，每一个自由度上能量连续¹⁰：

$$U = N(\overline{\varepsilon_x} + \overline{\varepsilon_y} + \overline{\varepsilon_z} + \dots)$$

声子系统，由 $3N$ 个独立振子构成，每一个振子上的元激发称作声子，而这样的元激发可以用任意个：

$$U = \sum_{j=1}^{3N} \varepsilon_j , \quad \varepsilon_j = \left(n_j + \frac{1}{2} \right) \hbar\omega_j$$

⁸液体中的声波的量子往往更复杂，如常常会产生旋子(rotion)

⁹它将由格点之间相互作用的非谐效应造成，当然晶格缺陷也会造成声子的散射。

¹⁰当然只要能量足够小，照样也可以提出量子版本的理想气体模型

这样的区别当然就能体现在热学性质上。具体推导过程还需要参考此后关于统计物理的章节。从热容和热导率结果上来说：

理想气体单分子平动动能服从玻尔兹曼分布，故平均动能为：

$$\overline{\varepsilon_x} = \overline{\varepsilon_y} = \overline{\varepsilon_z} = \frac{1}{2}kT$$

但是每一个振子上的平均能量却服从玻色-爱因斯坦分布，故平均能量只有：

$$\overline{\varepsilon_j} = \frac{\hbar\omega_j}{e^{\frac{\hbar\omega_j}{kT}} - 1}$$

可见只有当 $\hbar\omega_j \ll kT$ ，每一个振子上的平均能量才回归经典物理结果（动能 + 势能两个平方项）：

$$\frac{\hbar\omega_j}{kT} \ll 1 : \quad \overline{\varepsilon_j} \simeq kT, \quad U = 3\nu RT$$

故在高温下，的确固体热容与 $3R$ 较为符合。但是低温下，正如后文将论述的那样，较高角频率的振子很难被激发，故统计振子内能时将存在“截断”。这导致热容与温度的三次方成正比：

$$T \rightarrow 0^{11} : \quad U = CT^4, \quad C_V = 4CT^3$$

固体的导热与理想气体的导热机理相似吗？原则上来说相似：都是粒子在运动过程中不断受到散射作用而造成热导率。但是声子的散射比较复杂：考虑固体格点之间非谐势能会造成多声子散射过程，或是考虑固体内的缺陷也会造成声子散射。一般来说固体热导率不仅与温度呈现分段函数关系，还依赖于固体的制备方式，纯度或掺杂浓度，低温下甚至依赖于固体的尺寸。本书不做展开。

8.2 固体电子论 *

促使现代固体物理理论成型的早期工作涉及金属的物理性质的测量，一是纯粹的热学特性如热容和热导率，一是纯粹的电学（输运）特性如电导率，还有若干热电耦合特性与磁电耦合特性。即使从直接的感官上说，金属的大量性质异于通常的固体是显而易见的，特殊的光泽，良好的延展特性，优异的导电导热特性。稍加思考便不难想到这些都与金属中的电子运动的独特性有关。

但是对于金属电子论的早期研究，物理学家们遇到了很大的困难：大量实际测量结果并不能符合理论的预期。尤其是金属的热容：例如洛伦兹曾经就疑惑不解：电子能够参加电的传导，那么就应该视作在运动，怎么对于热容又没有贡献？实际观测金属热容扣除晶格热容以后，电子贡献的热容低于视作理想气体热容的 1%。这个问题只有到了泡利不相容原理与狄拉克-费米分布被发现之后才得到解答。

固体中的电子分为两大类：一类是可以自由运动的**传导电子**（conduction electron），在所有原子间自由游走¹²，低温下到了固体边缘被势垒反弹回来；第二类电子就是**束缚电子**（bound electron），它是被仅局限在空间中特定区域而无法转移的电子，细分又有两类：**紧束缚电子**（tight binding electron）与氢原子模型中的电子类似，在单个原子内存在，往往是原子外的内层电子，经典地看可以认为在原子轨道上运动；以及**价电子**（valence electron）或**成键电子**（bonding electron），它们在若干原子间来回运动。

可以想见，传导电子对固体的热学性质会有较大的影响，束缚电子主要只对固体的电学特性造成改变。金属含有大量传导电子，我们下面主要来研究它。

8.2.1 德鲁特模型

- 经典理论为**德鲁特模型**（Drude model），把电子视作类似理想气体，在原子实间做匀速直线运动，平均速率为 v ，频繁发生碰撞，碰撞自由程 λ ，自由时间 τ ，经典理论给出其输运的两个特性：电导率与热导率：

$$\sigma = \frac{ne^2\tau}{m} ; \quad \kappa = \frac{1}{3}nc\lambda v$$

¹¹更准确的说法是 $T \ll T_D$ ，温度远小于德拜温度。

¹²不一定所有方向，比如石墨导电只发生在层内。

- 经典理论对平均速率,自由程与自由时间的估计都严重偏离真实值好几个量级,电子质量也应该修正为有效质量才准确.但最后对电导率,热导率的估计仅差几倍,但仍然没有解释金属的热容为什么没有因为电子热运动而增加的问题.
- 考虑到电子的波动性,不确定原理告诉我们电子的波函数的弥散范围至少能扩展到原子的尺寸.把电子处理为波得出了单电子的布洛赫波理论(Bloch wave theory),而且得出了电子运动动量与能量之间的独特色散关系,也就是电子的能带结构(energy and structure),能量既不是像真空中电子那样连续也不是像原子内电子那样有能级,而是介于两者之间.价电子将一个能带填满处于满带,故不导电;而导电的电子处在未填满的能带.
- 同时,与原子中的电子一样,电子要符合泡利不相容原理(Pauli exclusion principle),能带中的电子往往是从最低能量的静止态往上填充,填充到其动能约为几个电子伏的费米能级上.即使在常温下,电子在各个能级上的自带能量也会远大于热力学特征能量 kT ,所以它不是符合麦克斯韦-玻尔兹曼分布,而是符合由于电子作为费米子(fermion)其微观特性带来的费米-狄拉克分布(Fermi-Dirac distribution).这样的电子体系是强简并气体(strongly degenerate gas).
- 部分绝缘体带隙(band gap)较窄,部分处于满带的电子可以因为热运动激发到空带.这就成为了本征半导体(intrinsic semiconductor).激发的电子数密度若为 n ,那么载流子就为数密度 n 的电子和数密度 $p = n$ 的空穴:

$$p \cdot n \propto e^{-\frac{V_g}{kT}} ; \quad \sigma = pe\mu_p + ne\mu_n$$

其中 $\mu = e\tau/m$ 为迁移率,为速度正比于外力的系数.

- 掺杂半导体分为 p 型和 n 型,主要载流子为空穴 p for positive 和电子 n for negative.用它们制作的pn 结(pn junction)进一步就是各半导体晶体管的基元了.在 pn 结内电子的输运有两大类:扩散电流(diffusive current)和漂移电流(drift current).

8.3 液体的彻体性质

8.3.1 液体性质综述与其微观成因



液体在很多方面与气体相似,而与固体区分开来.具体来说,液体没有固定的体积易于流动;没有弹性只有黏性;黏度与压缩系数比气体倒是大不少,这得益于其接近固体的高密度;热容,热导率等更细节的内容我们后几节详细阐述;微观细节则在以后章节中加以说明.最后,我们指出特殊条件下液体也具有某些与固体相似的性质,尤其是当变形率十分的急剧时.这种现象是非牛顿流体(non-newtonian fluid)的一种典型,比如图8.5中用玉米面粉与水和成的欧不裂(oobleck)就是一种典型的一旦作用力变化十分快,就会主要表现出弹性固体的性质而不是液体的流

图 8.5: 在有“弹性”的液体上行走 动性的材料,所以人才可以在液面上快步走动而不会陷进去.但使我们把液体固体彻底区分开来的主要原因是固体具有序(order),而且是长程(long range)的序¹³.

液体内一般无序.每一个单元周围的若干近邻粒子距离不像固体那样几乎取固定值,也不像气体那样剧烈变化.而是在一个特定的范围内分布.用近邻粒子随半径的分布的方式来研究液体和固体气体的区别是一种行之有效的方法.利用 X 射线衍射的方法可以得到物质内部单元之间距离分布的信息.这样可以

¹³凡事皆有例外,液晶(liquid crystal)就有长程序的液体,非晶体(non-crystalline)便是无长程序的固体.更准确的液体固体区分方式应该要看结合的强度.

得到系统的径向分布函数(radial distribution function), 即粒子体密度随与中心粒子距离的分布函数. 如图8.6所示为氩的三态的模拟计算后的径向分布函数.

径向分布函数为“表”, 相互作用形式为“里”. 从径向分布函数可知: 气体仅有当两个分子较为靠近时径向分布函数才略有升高并随半径减小迅速趋于零, 从中可以分析出两个分子靠近则产生吸引, 势能降低, 玻尔兹曼分布下造成概率增加. 而通过固体径向分布函数可以看出即使在半径较大的情况下分子也具有长程序. 而液体则短距离下径向分布函数类似固体, 长距离下类似气体. 即在短距离下吸引力占据主导作用, 分子团可以形成局部的“类晶区”, 但是在长距离下则不具有平移不变性.

而如果考察液体分子之间相互作用的形式. 液体可以分为以下类型:

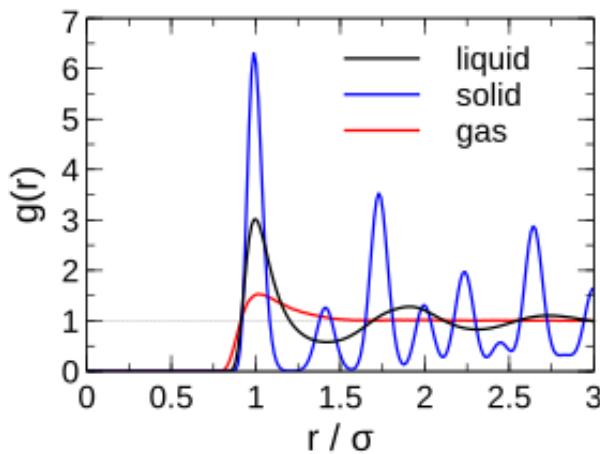


图 8.6: 径向分布函数

- **范德瓦尔斯液体**(van der Waals liquid)

由无固有电偶极矩分子构成的液体, 其分子之间相互作用力为范德瓦尔斯力, 这是一种由于量子效应导致的与分子之间感应产生电偶极矩有关的力. 经验公式是为其引入前文也介绍过的**李纳-琼斯势能**(Lennard-Jones potential):

$$V(r) = -\varepsilon \left[\left(\frac{\sigma}{r} \right)^6 - \left(\frac{\sigma}{r} \right)^{12} \right]$$

下一章相变将详细研究这种力同时造成的气体与液体的性质.

- **极性液体**(polar liquid):

构成液体的单元存在固有电偶极矩则为极性液体, 在极性液体中, 有存在一类单元之间有着较强吸引作用的液体, 水就是其中的典型代表. 称作**缔合性液体**(associated liquid). 缔合性液体性质非常不同于非缔合性的对应液体, 热学上通常具有较大的黏滞, 反常的压缩与热容特性. 甚至会造成**液液相变**(liquid-liquid phase transition, LLPT).

- **金属液体**(metallic liquid):

如果构成液体的基本可活动粒子带电, 如电子, 离子, 原子核. 那么这会大大增加液体内部的关联, 形成较好的导热和导电性能. 如熔解的金属, 稠密的等离子体等等.

- **量子液体**(quantum liquid):

在极低温或极高密等极端条件下的一些系统将会出现**超流性**(superfluidity). 典型例子如 2.4K 以下温度的液氦 (${}^4\text{He}[II]$). 它将体现出第二声 (温度波), 热容发散, 黏度消失等奇特的力热效应.

即使对于我们最熟悉的液体水, 很多问题至今依然是开放问题与前沿问题: 如何准确描写水分子之间的相互作用? 冰和液态水之间的相变和液态水的反常膨胀是怎样的过程? 水一共有多少种物相? 诚然液体系统是复杂的多体问题, 其难点在于从自由的理想气体模型引入相互作用导致的高密度气体模型和从具有序的晶体模型不断引入序的破缺导致的瓦解的固体模型都不能很好地反应液体独特行为的全貌. 在理论研究液体行为的场合较多被使用的方法为**密度泛函理论**(density functional theory, DFT). 对这个方法的介绍已经远远超出本书的范围. 故在以下对液体具体性质的讨论中我们将仅仅围绕现象进行介绍并做少量定性的原理注解.

8.3.2 热容与黏度

热容:一般来说,按原子的摩尔数计算,才有等压等体热容约为 $3R$. 热膨胀:部分液体与晶体反常膨胀,结晶过程也不是收缩反而是膨胀,一种解释是加热使得一些固体到液体成键倾斜以后更有利于在其他方向节省空间.

黏度:液体一般温度越高黏度越小.这是因为不同于气体黏度,液体黏度更像固体那样,靠的是相互作用形成的“激发”.而热运动加剧时液体分子间相互作用更难形成结果,比如两个原子很难形成一种亚稳态的“集团”.

8.4 液体的表面性质

- 液体与气体交界面为**表面**(surface),与固体交界面为**界面**(interface). 表面一般就若干个分子层那么厚,表面的分子数密度由于过渡到气体中所以明显会降低. 界面则可能降低可能升高.
- 分子与分子间相互吸引形成内聚力,它的计算一般包括在体的内能中. 但表面与界面的形成会改变能量,叫做表面能与界面能. 它正比于其面积 S . 习惯上:

$$W_s = +\sigma_s S = +\sigma_{lg} S$$

$$W_i = -\sigma_i S = -(\sigma_{ls} - \sigma_{gs}) S$$

前者 S 为液体表面积. 后者 S 为液体固体接触的界面面积,就是说要考虑由于液体与固体的接触使得固体与气体少接触了相同的面积可能导致的 $\sigma_{gs} S$ 能量,尽管这一项几乎总是可以忽略. σ_{ls} 是可正可负可为零的. σ_{lg} 一般大于零.

- 表面能对应**表面张力**(surface tension):

$$f_s = \sigma_s L \quad ; \quad f_i = \sigma_i L$$

前者使表面尽可能小,后者使界面尽可能大.

- 附加压强的**杨-拉普拉斯方程**(Young-Laplace equation):

$$\Delta p = \sigma \left(\frac{1}{\rho_x} + \frac{1}{\rho_y} \right)$$

其中 ρ_x 和 ρ_y 是可以任意选取的两个相互垂直的方向上曲面上法曲线的曲率.

- 液体固体接触的**接触角**(contact angle) 的**杨-杜普雷方程**(Young-Dupré) 方程:

$$\cos \theta = \frac{\sigma_i}{\sigma_s}$$

若 $\sigma_i > 0$ 称为**浸润**(wetting). 若 $\sigma_i > \sigma_s$ 上式无解称为**完全浸润**(perfect wetting), 液体会在固体表面尽可能铺展. 若 $\sigma_i \leq 0$ 称为**不浸润**(non-wetting), 若 $\sigma_i < -\sigma_s$ 称为**完全不浸润**(perfect non-wetting)

- 表面能实际上是自由能而不是内能,等温等压拉伸液膜要伴随着吸放热. 详情参考之前章节的计算.

8.5 极端条件下的其他物态

- 超冷 ($\sim \mu\text{K}$): **玻色-爱因斯坦凝聚**(Bose-Einstein condensate): 宏观数目的原子的动量 $p = \hbar k$ 凝聚到了基态. 其波函数可以被干涉实验探测到.

- 低温 ($1 \sim 10\text{K}$): 超流(superfluid) 现象: 如 He-4 液体黏度消失. He-3 由于是费米子其超流直到 1970 年代才被发现, 需要 $\sim \text{mK}$ 的低温.
- 低温物理中的高温 ($10 \sim 100\text{K}$): 超导(superconductor): 常规金属的电阻率会消失, 伴随着迈斯纳效应(Meissner effect): 完全抗磁性. 电子结合成库珀对(Cooper pair), 可以像 BEC 那样通过晶格作用而凝聚成整个波函数而传播.
- 高温低密 ($10^3 \sim 10^4\text{K}$): 等离子体(plasma), 所有电子都与原子核不再形成束缚态, 而是构成独立的气体体系, 可以不考虑相对论效应也不考虑电子的量子简并来处理.
- 超高温, 高密度 ($10^{11} \sim 10^{12}\text{K}, 10^{17}\text{kg/m}^3$): 引力场下坍缩形成的中子星. 电子数不再守恒而是在中子质子互变的反应中产生与消灭.
- 最高温 ($> 10^{12}\text{K}$): 各大型对撞机上已经实现的夸克-胶子等离子体(quark-gluon plasma): 质子数和中子数不再守恒, 内部的构成的夸克与胶子形成强相互作用的等离子体.
- 每一种粒子数不再守恒的极端相对论性的气体, 其内能密度与温度的关系都是: 对于每自由度的玻色子与费米子:

$$u_b = \frac{2\sigma}{c} T^4 \quad ; \quad u_f = \frac{7}{8} u_b$$

例如对于光子气, 两个偏振自由度再结合 $J = uc/4$ 便得到黑体辐射公式. 对于不包含第二代基本粒子的 QGP, 夸克有 2 味 3 色 2 自旋 2 电荷一共 24 自由度, 胶子有 8 色 2 偏振(螺度)共 16 自由度. 代入得:

$$u = \frac{74\sigma T^4}{c}$$

第九章 相与相变摘要

9.1 相平衡

- 相平衡条件: 等温等压, 化学势相等:

$$\mu_1(T, p) = \mu_2(T, p)$$

即三种平衡同时被满足: 热学, 力学与化学平衡.

- 若 $\mu_1 > \mu_2$, 粒子从势高的向势低的自发相变. 如某温度下未饱和的水蒸气内化学势就会低于水中的化学势.
- 除了这三个量相等, 摩尔熵 s , 摆尔体积 v , 数密度 n 两相平衡时都可以不相等.
- 相变潜热(latent heat) 指等温等压条件下的吸热, 所以是摩尔焓的增量:

$$Q = \nu l = h_2 - h_1 = u_2 - u_1 + p(v_2 - v_1)$$

- 0°C 下若把水的比热容记做 $1\text{cal/g} \cdot \text{K}$, 那么水的溶解热与汽化热为 80cal/g 与 600cal/g , 100°C 下汽化热降低到只有 540cal/g . 冰和水蒸气的比热容几乎刚好只有水的一半 $0.5\text{cal/g} \cdot \text{K}$

9.2 气液相变

首先范德瓦尔斯方程(van der Waals equation):

$$p = \frac{\nu RT}{V - \nu b} - \frac{\nu^2 a}{V^2}$$

- 用范德瓦尔斯方程同时可以描述气体和液体, 低于临界温度 $T_c = 8a/27Rb$ 时 $p - V$ 曲线不再单调. 提供了等温等压等化学势相平衡的可能性. 事实上在麦克斯韦等面积法则(Maxwell's equal area rule) 决定的两点处能够相平衡. 这定义了气液两相的分歧背后的数学原因.
- 汽化的熵变其实简单地决定了潜热:

$$l = T(s_g - s_l)$$

- 两相共存时压强与温度的导数关系, 也就是克劳修斯-克拉伯龙关系(Clausius-Clapeyron relation):

$$\frac{dp}{dT} = \frac{s_g - s_l}{v_g - v_l} = \frac{l}{T(v_g - v_l)}$$

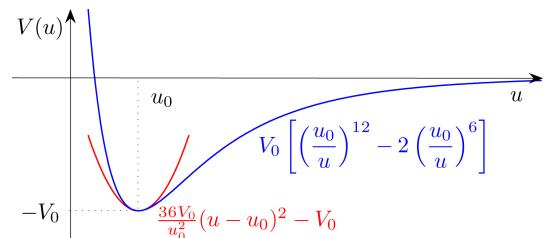


图 9.1: 范氏等温线

- **蒸发**(evaporation): 由于大气的存在,水蒸气在空气中仅是分压,不等于水的压强,但是大气压对水的压强使水压缩程度很小对水的化学势几乎可以忽略.故,可以认为某温度平衡时空气中的水汽分压就是该温度下水蒸气单独与水在该温度共存时的压强,称作**饱和蒸气压**(saturate vapour pressure). 0°C 时为 0.6kPa, 室温下变成 2 – 4kPa, 最后 100°C 变成 1atm. 蒸发条件是外界大气环境水蒸气分压不能达到该温度下的饱和蒸气压. 这样化学不平衡会让水变成蒸汽而增大水面上的总压强,但是力学不平衡又会让蒸汽转移向上方的大气形成蒸发流. 易挥发的气体会尤其明显. 高压锅也是利用这个特点,取消了力学平衡,用化学平衡来使压强单方面增加. 日常生活中的空气一般都是不饱和的,降雨的天气空气湿度可达 100% 而形成饱和的空气. 如果空气水蒸气分压较大而经历降温至**露点**(dew point) 以下,水蒸气就会析出而结露.
- **沸腾**(boiling): 当温度达到甚至高于某压强下两相共存对应的温度时. 剧烈的沸腾现象就发生了. 比如 1atm 下水在 100°C 沸腾, 此时水蒸气不断产生希望在更高的压强下实现化学平衡,但是外界大气压的力学平衡又不允许. 所以就发生了从内到外的剧烈汽化现象.

第十章 统计物理基础

10.1 数学基础

10.1.1 概率与独立性

现实生活中有很多偶然事件,偶然事件的成因是多种多样的,它们集中表现在相似的条件下进行试验,而能够得到完全不同的结果.数学上用抽象的集合来表示所有可能的结果:

$$x \in X, \quad X = \{x\}$$

其中每一个元素代表一种可能的结果.这些结果应该有以下特点:

忠实性:每一个不同的实验结果如实地反应为集合中的不同元素.也就是不能用一个元素代替一类实验结果.

互斥性:当一个结果发生时,另一个结果就必须排除,也就是不能有多个元素对应同样的实验结果.

完备性:所有的实验结果必须都有集合中的元素对应.

这样就能把集合 X 称为**样本空间**(sample space),而每一个元素 x 也称为一个**元事件**(elementary event).我们常说的**事件**(event),其实一般指样本空间的一些定义良好的子集 $A \subset X$.只要实验结果在这个子集中,就说这个事件发生了.所谓定义良好我们可以做如下理解:

在**古典概型**(classic models)情形下,样本空间是一个有限的集合,此时任意子集都可以视为某种事件.如投一颗骰子,样本空间为 $X = \{1, 2, 3, 4, 5, 6\}$,则投出偶数是一个事件 $A = \{2, 4, 6\}$.

但在**几何概型**(geometric models)情形下,样本空间一般具有与 \mathbb{R}^n 类似的结构,一般都是无限的集合,有一些“事件”的提法应当给予摒弃否则会引起矛盾.例如在闭区间 $[0, 1]$ 间任取一个点,这个点恰好是有理数这样的“事件”可能就不是那么定义良好¹.一般常见的事件如点落在区间 (a, b) 内等等.

设想同时做好几个实验,这几个实验互不干扰它们的结果是完全独立的,那么联合到一起就构成了一个大的实验,其结果应表示为一个数组:

$$\mathbf{x} = (x_i) ; \quad x_i \in X_i$$

而新的样本空间称为原来那些样本空间的**独立直积**(independent product).记做:

$$\bigotimes_i X_i = \{\mathbf{x} = (x_{ij}) | x_{ij} \in X_i\}$$

互斥,独立这样的一些概念如何用数学严格表述?事实上它们恰恰是用概率去定义的.概率是一种我们关于实验结果可能性的**先验假设**(a priori presumption),它是**随机变量**(random variable)的正实数值函数 P ,而随机变量既可以取为样本空间中的单个结果,也可以取为结果的集合:事件:

$$P(A) = P(\bigcup_{x_i \in A} x_i) = \sum_{x_i \in A} P(x_i)$$

¹事实上有理数集是定义良好---可测的,但有些集合不可测从而不能讨论它们的概率.参考https://en.wikipedia.org/wiki/Non-measurable_set

这其中已用到了 P 的属性: **互斥事件**(mutual-exclusive events, disjoint events) 的加法原理:

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

再加上:

$$P(A) \geq 0 ; P(X) = 1$$

就构成了一个合适的概率定义. 而对于随机变量可取连续样本空间, 比如区间 $[a, b]$ 的情形, 引入**概率密度函数**(probability density function, pdf):

$$P([x, x + dx]) = p(x)dx ; p(x) > 0, \int_a^b p(x)dx = 1$$

便是一个合适的概率定义.

现在来讨论事件的独立性. 两个非互斥的事件可能同时发生, 同时发生这一个新的事件即被定义为 $A \cap B \neq \emptyset$, 此时可以定义事件 B 在事件 A 下的**条件概率**(conditional probability):

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

显然, 若条件概率为零, 那么实际上两个事件互斥. 而实际上如果 B 在 A 下的条件概率等于 B 的概率, 那么这种情况称为两个事件相互独立:

$$P(B|A) = P(B)$$

注意到上式实际上也可以写为:

$$P(X)P(A \cap B) = P(A)P(B)$$

所以两个事件相互独立的确是一个相互的关系, 此时 A 的条件概率也有 $P(A|B) = P(A)$. 两个事件相互独立是一种很微妙的关系, 这意味着一个事件的发生既不会阻碍另一个事件, 也不会促成另一个事件. 而是完全没有影响. 物理上看, 很有可能两个事件没有因果关系.

我们之前构造的独立直积, 用概率表示即为:

$$P((x_{ij})) = \prod_i P(x_{ij})$$

而对于连续概率分布, 两个样本空间的独立直积给出的新概率应该是有以下概率密度:

$$p(x, y) = p_1(x)p_2(y)$$

10.1.2 随机变量及其数字特征

物理实验中的随机现象, 常常体现为实验测量中测量结果数据的上下波动. 此时一般认为这个测量数据 x 就是一个随机变量, 而具有先验的某种概率密度函数 $p(x)$. 有时候我们关心测量数据的某种函数 $y = f(x)$, 那么不同的 y 其出现概率分布应该修改:

$$p_y(y) = \frac{dp}{dy} = \frac{p(x)dx}{dy} = \frac{1}{f'}p(x)$$

例如, 如果考虑在 x 方向的粒子速度 v 具有分布律:

$$p(v) = \frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}$$

可以验证这个概率归一, 那么其能量 $e = \frac{v^2}{2}$ 的分布为

$$p_{e+} = \frac{p(v)}{v} = \frac{1}{2\sqrt{\pi e}}e^{-e}(v > 0) ; p_{e-} = \frac{p(v)}{-v} = \frac{1}{2\sqrt{\pi e}}e^{-e}(v < 0)$$

注意到同一个能量对应着两种可能的运动方向,故把两个 de 对应概率分开算,最后合到一起:

$$p_e(e) = \frac{1}{\sqrt{\pi e}} e^{-e}$$

一个在统计中关心的数学处理,是在 n 次全同独立测量中获得的测量结果的平均数(average):

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

而著名的**大数定律**(law of large numbers) 指出:如果概率分布的一个重要数字特征:**期望**(expectation) 存在²:

$$\mu = \langle x \rangle = \int x dp$$

那么如果随着测量次数趋于无穷,平均数将会无限趋近于期望值:

$$\lim_{n \rightarrow \infty} P(|\bar{x} - \langle x \rangle| > \varepsilon) = 0, \forall \varepsilon > 0$$

如何理解这一点? 我们再定义另一个随机变量的数字特征:**方差**(variance), 物理上称为**涨落**(fluctuation), 数学上还称**中心二极矩**(second central moment):

$$\sigma^2 = \langle (x - \mu)^2 \rangle = \int (x - \int x dp)^2 dp$$

物理上还常常把涨落除以期望的平方称为相对涨落. 那么可以证明:

$$\langle \bar{x} \rangle = \langle x \rangle ; \quad \langle (\bar{x} - \mu)^2 \rangle = \frac{\langle (x - \mu)^2 \rangle}{n}$$

故随着实验次数增加,平均数作为新的随机变量它的期望不会变,而方差在不断减小,故有中心极限定理的说法.

以上讨论适用于任何 $y = f(x)$ 型的随机变量,其中 x 可以是一次实验的实数结果,也可以是多次实验的数组结果,也可以是不同独立甚至非独立测量结果的复合,此时 y 即被理解为事件,代表某些可能发生的元事件的集合. 期望与方差被定义为:

$$\begin{aligned} \mu &= \int y dp \\ \sigma^2 &= \int (y - \mu)^2 dp \end{aligned}$$

其中 dp 表示 y 落在该点附近范围内的概率.

10.1.3 信息熵

为了刻画样本空间与其概率的多样性,我们引入**信息熵**(information entropy) 的概念. 对于一个样本空间与随机分布,它具有的熵被定义为:

$$S = \sum_{x_i \in X} -P(x_i) \ln P(x_i)$$

而每一种可能性 x_i 带来的信息量被定义为 $I_i = -\ln P(x_i) > 0$. 而总熵是事件信息量的期望. 为什么如此定义? 比如我们设想投一枚骰子,那么投出 1 这个数的信息量应该是投出奇数的信息量加上在已知投出奇数后再投出 1 的信息量,投出奇数称作 A 事件那么这个性质写作:

$$I(x) = I(A) + I(x|A)$$

²值得指出,有些数学上诡异的概率分布是不存在期望的,比如典型的柯西分布:

$$p(x) = \frac{\gamma/\pi}{x^2 + \gamma^2}$$

当然这样的理想概率分布物理上几乎不存在,因为物理量总是有限的,在一定尺度下适用.

而应该有一种概率与信息量之间的普适对应,以上三个信息都可以理解为:

$$I(x) = f(P(x)), I(A) = f(P(A)), I(x|A) = f(P(x|A))$$

但三个概率本来就有:

$$P(x) = P(A)P(x|A)$$

这暗示着 $f(xy) = f(x) + f(y)$, 在假定 f 性质足够好 (可以求二阶导数) 情况下:

$$\begin{aligned} & \xrightarrow{\partial_y} xf'(xy) = f'(y) \\ & \xrightarrow{\partial_x} f''(xy) + xyf'(xy) = 0 \\ & \xrightarrow{xy=u} \frac{d}{du}(u \frac{df}{du}) = 0 \\ & \Rightarrow f = C \ln u + C' \end{aligned}$$

重新代入 f 性质, 得 $C' = 0$. 剩下的 C 可以任取, 一般理解为 -1 , 这样给出了正的信息量. 这个推理过程还给出了事件的信息量 $I(A)$ 与条件信息量 $I(A|B)$ 两个概念的定义. 根据定义, 可见 A 取整个样本空间的信息量就是零.

现在我们理解到信息量表示稀有程度, 具体事件发生概率越小则信息量越大. 那熵又表示什么含义呢. 它表示多样性. 我们设想如果某一种元事件 x_i 的可能性被发现还可以细分为 $x_{i1}, x_{i2}, \dots, x_{im}$ 种子可能性. 那么根据此式所引入的熵就会因此而增加, 增量为:

$$\begin{aligned} \Delta S &= \sum_{j=1}^m -P(x_{ij}) \ln P(x_{ij}) - (-P(x_i) \ln P(x_i)) \\ &= \sum_{j=1}^m -P(x_{ij}) \ln P(x_{ij}) - \sum_{j=1}^m -P(x_{ij}) \ln P(x_i) \\ &= \sum_{j=1}^m -P(x_{ij}) \ln \frac{P(x_{ij})}{P(x_i)} \\ &= \sum_{j=1}^m P(x_{ij}) I(x_{ij}|x_i) \\ &= P(x_i) \cdot \sum_{j=1}^m -\frac{P(x_{ij})}{P(x_i)} \ln \frac{P(x_{ij})}{P(x_i)} \\ &= P(x_i) \cdot \sum_{j=1}^m -P(x_{ij}|x_i) \ln P(x_{ij}|x_i) \\ &= P(x_i) S(x_i) \end{aligned}$$

一方面, 等于细化这一元事件所增加的信息量的概率平均. 另一方面, 也等于事件熵与概率的乘积. 此处也定义了事件细分熵:

$$S(A) = \sum_{x_i \in A} -P(x_i|A) \ln P(x_i|A)$$

可见单个元事件的细分熵是零. 而如果把样本空间可以分为一系列不相交的事件:

$$X = \bigcup_i A_i ; A_i \cap A_j = \emptyset (i \neq j)$$

那么有:

$$S(A) = \sum_i P(A_i)[I(A_i) + S(A_i)] = S(\{A_i\}) + \sum_i P(A_i)S(A_i)$$

可见把不相交的事件并在一起时, 总熵就是把 A_i 视作元事件的熵加上各个子事件的熵的加权和.

最后让我们来看看作为以上熵的定义所符合的一个十分重要的性质.为了方便起见我们把样本空间记做 Ω .并在样本空间的每个元事件引入两个随机变量 $x_i \in X, y_j \in Y$.还要求两个随机变量独立:

$$P(x_i, y_j) = P(x_i)P(y_j), \forall i \forall j$$

那么对于整个样本空间按照随机变量 X 来划分的熵与 Y 的熵:

$$S_X = \sum_i -P(x_i) \ln P(x_i) ; S_Y = \sum_j -P(y_j) \ln P(y_j)$$

而总熵:

$$S = \sum_{ij} -P(x_i, y_j) \ln P(x_i, y_j)$$

计算其关系,容易发现:

$$\begin{aligned} S &= \sum_{ij} -P(x_i, y_j) \ln P(x_i)P(y_j) \\ &= \sum_{ij} -P(x_i, y_j) \ln P(x_i) + \sum_{ij} -P(x_i, y_j) \ln P(y_j) \\ &= \sum_i \sum_j -P(x_i, y_j) \ln P(x_i) + \sum_j \sum_i -P(x_i, y_j) \ln P(y_j) \\ &= \sum_i -P(x_i) \ln P(x_i) + \sum_i -P(y_j) \ln P(y_j) \\ &= S_X + S_Y \end{aligned}$$

即,独立随机变量的熵是可以直接相加的.而我们再来分析一下若是非独立随机变量的叠加导致的熵:

$$\begin{aligned} S_X + S_Y - S &= \sum_{ij} -P(x_i, y_j) \ln P(x_i) + \sum_{ij} -P(x_i, y_j) \ln P(y_j) - \sum_{ij} -P(x_i, y_j) \ln P(x_i, y_j) \\ &= \sum_{ij} -P(x_i, y_j) \ln \frac{P(x_i)P(y_j)}{P(x_i, y_j)} \\ &\geq -\ln \sum_{ij} P(x_i, y_j) \cdot \frac{P(x_i)P(y_j)}{P(x_i, y_j)} \\ &= -\ln \sum_{ij} P(x_i)P(y_j) \\ &= -\ln 1 \\ &= 0 \end{aligned}$$

其中用到了 \ln 是上凸函数,所以自变量的加权平均处的函数值总是大于函数值的加权平均:

$$\ln(\sum_i p_i x_i) \geq \sum_i p_i \ln x_i ; \sum_i p_i = 1$$

以上结论告诉我们非独立的两个随机变量所带来的熵有一部分是重合的,从而导致总的熵小于两个随机变量熵的和.小的这一部分叫做两个变量之间的互信息(mutual information):

$$M = S_X + S_Y - S$$

互信息代表着已知一个随机变量的结果以后对另一个随机变量条件概率分布的影响.例如考虑自然光在某个方向的分振幅,以 x 方向的振幅作为第一个随机变量,如果另一个方向也是 x ,那么互信息恰好等于 x 方向振幅概率分布的熵;如果另一个方向为 y 方向,那么两个随机变量独立,从而互信息为 0;如果选取一个与 x 方向夹锐角的方向,则 x 方向的振幅在一定程度上也会影响该方向的振幅的概率分布,从而两个随机变量之间具有一定的互信息量,并随着夹角增大到 90 度而消失.

- 概率模型: **经典概型**(classic models) 中原子事件构成的集合是有限的,一般为等概率分布.也可以想象不是等概率的样本空间.事件是部分原子事件构成的子集.一般写作:

$$x \in \Omega, \quad \Omega = \{\omega\}$$

$$P(A) = P(\bigcup_{\omega_i \in A} \omega_i) = \sum_{\omega_i \in A} P(\omega_i)$$

- 概率模型: **几何概型**(geometric models) 中的样本空间是与 \mathbb{R}^n 具有类似的结构的连续空间,一般认为参数化.比如原子事件抽象为区间 $[a, b]$ 内的一个数,或是二维球面 $S^2 = \{(\theta, \varphi) | \theta \in (0, \pi), \varphi \in [0, 2\pi]\} \cup \{\theta = 0, \theta = \pi\}$ 上的一个点.而概率则是一个归一化的**概率分布函数**(probability distribution function):

$$\int_a^b dp = \int_a^b p(x)dx = 1$$

$$\oint_{S^2} dp = \oint_{S^2} p(\theta, \varphi)dS = 1$$

- 独立事件**(independent events): 对于经典概型,事件 A 与事件 B 独立:

$$P(A \cap B) = P(A)P(B)$$

如果把事件的交写作直接的乘积 $A \cap B = AB$,事件的补 $\Omega \setminus A$ 写作 \bar{A} ,上式等价于以下四式:

$$P(AB) = P(A)P(B) \quad ; \quad P(A\bar{B}) = P(A)P(\bar{B})$$

$$P(\bar{A}B) = P(\bar{A})P(B) \quad ; \quad P(\bar{A}\bar{B}) = P(\bar{A})P(\bar{B})$$

在样本空间定义一个函数变量 $X : \omega_i \mapsto x_i = X(\omega_i)$,那么可以根据不同的函数结果对样本空间进行分类,每一个结果就是一个事件,概率为 $P(x_i)$.此时变量 X 与变量 Y 独立即定义为:

$$\forall x_i \forall y_j : P(x_i y_j) = P(x_i)P(y_j)$$

对于定义在 $(x, y) \in \mathbb{R}^2$ 上的几何概型,概率分布关于两个坐标独立的条件为可以找到两个坐标各自的概率分布函数:

$$\forall x \forall y : p(x, y) = p_x(x)p_y(y)$$

- 两种概型下变量 X 的**期望**(expectation) $E = \langle X \rangle$ 与**方差**(variance) $\sigma^2 = \langle (X - E)^2 \rangle$ 的定义:

$$E = \sum_{\omega_i \in \Omega} P(\omega_i)X(\omega_i) = \int_{\Omega} X(x)dp$$

$$\sigma^2 = \sum_{\omega_i \in \Omega} P(\omega_i)(X(\omega_i) - E)^2 = \int_{\Omega} (X(x) - E)^2 dp$$

- 两种概型下**统计熵**(statistic entropy) 或**信息熵**(information entropy) 的定义:

$$S = - \sum_{\omega_i \in \Omega} P(\omega_i) \ln P(\omega_i) = - \int_{\Omega} \ln P(x)dp$$

- 在 \mathbb{R} 上的分布若要求 σ^2 取固定值,那么**正态分布**(normal distribution) 可以使得熵最大:

$$p(x) = N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

如果还要求期望为零,那么上式 $\mu = 0$.注意这等价于说,如果一个一维粒子的平均动量为 0 而平均动能为 $\langle mv^2/2 \rangle = kT/2$,熵最大的分布为:

$$f(v) = \sqrt{\frac{m}{2\pi kT}} e^{-\frac{mv^2}{2kT}} \quad ; \quad S = \ln \sqrt{\frac{m}{2\pi kT}} - \frac{1}{2}$$

- 如果已知 \mathbb{R}^2 上概率分布 $p(x, y)$ 的边缘分布(marginal distribution) p_x, p_y :

$$p_x(x) = \int_{-\infty}^{\infty} p(x, y) dy$$

$$p_y(y) = \int_{-\infty}^{\infty} p(x, y) dx$$

那么在 \mathbb{R} 分别以 p_x, p_y 作为分布函数的分布的熵与二维分布的熵满足:

$$S \leq S_x + S_y$$

成立条件是 X 与 Y 随机变量独立:

$$p(x, y) = p_x(x)p_y(y)$$

这就是说,两个随机分布合成时,独立合成能够使得熵最大.

10.2 统计假设

原则上来说,经典的物理体系应当是确定性(certainty)的,用经典力学语言可以描述(质点系的位置与动量)的,不包含随机性(randomness)的,不应引入概率描述的系统.然而一方面由于体系在大粒子数自由度数下所体现出来的统计复杂性,即使初态是一个高度特殊的状态:比如一个长方体气缸内理想气体初始时刻所有粒子排列为规则的点阵,具有完全相同的速度矢量.在历经一段时间之后也会体现出完全的随机性:体系在很长一段时间内看来各个状态都没有明显的序(order),显得杂乱无章.而稍微有序的状态几乎不可能再次出现.这期间的过程与演化,原则上说理应是确定性的过程,如果我们在某一时刻令所有粒子的速度方向反向,这个体系将会重新回到最初的状态.

但是我们将研究随机过程(stochastic process),在一个随机过程中,即使上一刻只有唯一的一种状态,下一刻的状态也不是决定性的,而是体现出不确定性或是随机性.从而对一个时刻系统完整的描述就要包含在各种可能状态下的概率,这实际上是一个系综(ensemble).用系综而不是单个的系统描述,从而引入概率的原因有以下几个:

- 外在随机性:多自由系统是混沌系统,初态微小的改变将会随时间以指数形式迅速放大.热力学体系总是处在外界环境中,外界总是对体系有各种各样的微小扰动.即使在之前的情况下在某个时间让所有粒子速度反向,也会因为外界对体系的影响而无法观察到体系回到初始状态.
- 内在随机性:以衰变为例,如果认为衰变发生在一个瞬间,粒子只可能处于衰变前或衰变后两个不同状态,那么衰变就是一个随机的过程.每一小段时间粒子均以均等的概率衰变,那么大量全同未衰变的粒子构成的体系的未衰变数期望随时间将指数衰减,但是每一个时刻未衰变粒子数仍然具有某种概率分布,粒子数对于其期望仍具有某种方差.事实上原子与原子的碰撞涉及到量子力学行为,同样的初始条件是会导致不同的结果的.
- 等概率原理(principle of equal a-priori probability):平衡态分布在相空间等能量曲面上所有点处的概率分布函数都应该相等.例如大量摆角固定的单摆构成的热力学体系,显然摆球出现在最高点的概率要大于最低点的概率(最高点速度慢停的久).但是如果把其摆动的 $p - x$ 空间画出来合适的比例下相轨迹为圆,而且一定做匀速圆周运动,所以肯定单位圆弧上出现摆球的概率相等.这个结论是理论力学里通过刘维尔定理(Liouville's theorem)的一种合情推理,但一般作为平衡态统计力学的最初假设.
- 通过等概率原理可以证明,如果理想气体也符合这个原理的话,平衡态分子速度作为随机变量分布为麦克斯韦分布律.这个推导参考热力学统计物理书籍的系综理论.

- 通过等概率原理 + 求最可几分布(most probable distribution) 也可以得到实际的分布为麦克斯韦分布律.这个推导也参考热力学统计物理书籍的最可几分布.
- 玻尔兹曼 H 定理**(Boltzmann H theorem): 通过分子混沌假说(molecular chaos hypothesis),认为从非平衡态分子通过碰撞向平衡态转变的过程中,每一时刻都存在的概率分布,即,真实世界的情况只是万千可能情况中的一种.而每一次分子的碰撞都是把单一的初态的可能性变成了更多的末态的可能性.数学上很容易证明,这样的概率演化过程熵只增不减.从而末态的熵一定达到了极值.再根据上一章的论述,粒子在三个方向的速度分布一定是正态分布,且彼此独立,平均动能一致时熵最大.也给出了麦克斯韦分布律.

10.3 麦克斯韦分布律

三维麦克斯韦速度分布律(Maxwell distribution law) 是指:

$$f(\mathbf{v}) = \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} e^{-\frac{mv^2}{2kT}} = \prod_1^3 \sqrt{\frac{m}{2\pi kT}} e^{-\frac{mv_i^2}{2kT}}$$

- 这是速度空间中的概率分布函数. 即若分子数 N 很大则 $dN = N f(\mathbf{v}) dv_x dv_y dv_z$.
- 由于三个方向分布律独立, 每个方向上的分布律也叫麦克斯韦速度分布律:

$$\sqrt{\frac{m}{2\pi kT}} e^{-\frac{mv_i^2}{2kT}}$$

$$f(\mathbf{v}) = f(v_x) f(v_y) f(v_z)$$

- 麦克斯韦速率分布律是指速度空间中等速率球之间的概率微元决定的函数:

$$\frac{dN}{N} = dp = F(v) dv$$

$$F(v) = 4\pi v^2 \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} e^{-\frac{mv^2}{2kT}}$$

- 平均速率与方均根速率:

$$\bar{v} = \sqrt{\frac{8kT}{\pi m}} ; \quad \sqrt{\bar{v}^2} = \sqrt{\frac{3kT}{m}}$$

- 泻流数与压强既可以用速率分布计算也可以用单方向的分布函数计算:

$$\Gamma = \frac{1}{4} n \bar{v} ; \quad d\Gamma = \frac{1}{4} dn v$$

$$p = \frac{2}{3} n \bar{\varepsilon} ; \quad dp = \frac{2}{3} dn \varepsilon$$

- 光子气的以上公式是类似的:

$$\Gamma = \frac{1}{4} nc ; \quad d\Gamma = \frac{1}{4} dnc$$

$$p = \frac{1}{3} n \bar{\varepsilon} ; \quad dp = \frac{1}{3} dn \varepsilon$$

10.4 能均分定理

- **近独立子系**(nearly independent system): 体系的内能可以写出各个粒子内能的和, 即忽略粒子之间的相互作用或者做**平均场**(mean field theory) 处理:

$$U = \sum u_i = \sum h(q_i, p_i)$$

- **玻尔兹曼分布律**(Boltzmann distribution law): 若粒子在某一自由度上的能量为 $h(q, p)$, 那么概率即为:

$$dp = C e^{-\frac{h(q, p)}{kT}} dp dq$$

其中 p 为该自由度上的广义动量, C 为归一化常数, 可积分计算. 或者直接我们写出最关心的平均能量:

$$u = \frac{\iint h(q, p) e^{-\frac{h(q, p)}{kT}} dp dq}{\iint e^{-\frac{h(q, p)}{kT}} dp dq}$$

- 每一个平方项形式的能量在麦克斯韦分布律下都具有如下相同的平均能量:

$$\overline{\frac{1}{2} M \dot{q}^2} = \overline{\frac{1}{2} K q^2} = \frac{1}{2} kT$$

- 非平方项故不符合能均分的例子: 重力场, 离散线性能级, 二能级系统, 三维电偶极子...

10.5 功, 热, 熵

- 通过以上近独立子系的分析可以发现, 体系的内能为温度和哈密顿量中参数 (比如系统的体积等整体广义坐标) 的函数:

$$\begin{aligned} U &= Nu = U(T, Q_\alpha) \\ u &= \iint h(q, p, Q_\alpha) \rho(q, p, Q_\alpha, T) dp dq \\ \rho(q, p, Q_\alpha, T) &= e^{-\frac{h(q, p, Q_\alpha)}{kT}} / \iint e^{-\frac{h(q, p, Q_\alpha)}{kT}} dp dq \end{aligned}$$

- 绝热功: 在一个过程中, 哈密顿量中参数改变, 温度改变, 导致分布不变:

$$\begin{aligned} d\left(\frac{h(q, p, Q_\alpha)}{kT}\right) &= 0 \quad \Rightarrow \quad d\rho(q, p, Q_\alpha, T) = 0 \\ du|_{\text{adiabatic}} &= \iint dh(q, p, Q_\alpha) \rho(q, p, Q_\alpha, T) dp dq \\ dU|_{\text{adiabatic}} &= \sum Y_\alpha dQ_\alpha \end{aligned}$$

- 热: 内能改变的其他部分:

$$dq = du - du|_{dQ=0} = \iint h(q, p, Q_\alpha) d\rho(q, p, Q_\alpha, T) dp dq$$

- 熵: 容易证明, 如果定义:

$$s = \frac{u}{T} + k \ln \iint e^{-\frac{h(q, p, Q_\alpha)}{kT}} dp dq$$

则有:

$$dq = T ds$$

第十一章 光波与光线

11.1 界面上的反射与折射

11.1.1 光波与光线

众所周知,光是电磁频谱的一段. 我们所说的光一般是指可见光(visible light), 学术上代表波长为 400 – 700nm 的电磁辐射. 它大致代表了人对光的色视觉的界限. 而一般的光学研究还包括了波长长的至 1mm 左右的红外光(infrared light, IR) 与短至 10nm 的紫外光(ultraviolet light, UV).

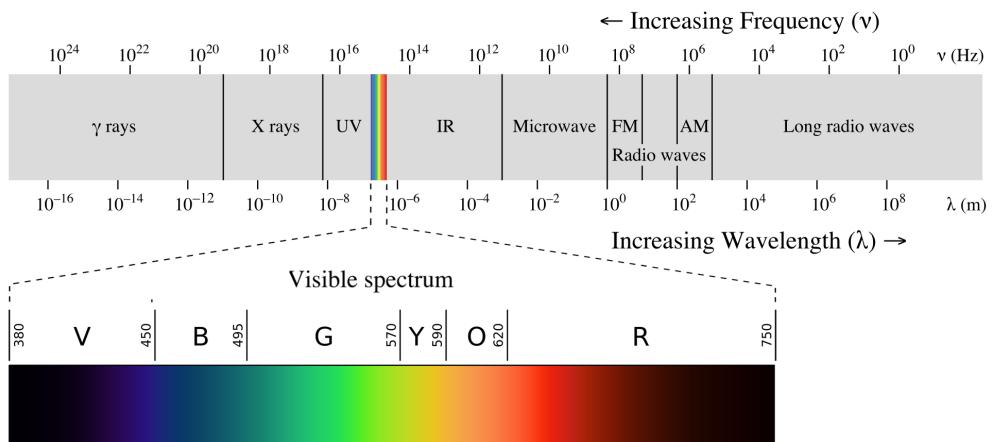


图 11.1: 光学研究范围

光的波动理论建立在众多干涉衍射实验上, 由 19 世纪后半页麦克斯韦 (Maxwell), 韦伯 (Weber) 与赫兹 (Hertz) 等人的理论与实验工作进行总结与升华. 最后通过迈克尔孙 (Michelson) 的干涉实验和随之由爱因斯坦 (Einstein), 洛伦兹 (Lorentz) 等人建立的狭义相对论来完善与归入到更加普适的经典电磁场论下. 然而整个 18 世纪实际上都处于光的粒子学说与波动学说的无休止争论中. 这是因为此时发现的光的几乎所有现象都可以同时用两种学说进行解释. 今天我们知道光实际上也有波粒二象性 (particle-wave duality), 最准确的描述光的语言应该是量子电动力学 (quantum electrodynamics, QED). 光是一份份的量子场, 被称为光子 (photon). 几何光学问题, 用经典的电磁场模型给出的描述最为精确, 而在特殊场合或者精度要求不高的情况下, 粒子光学的处理也是十分合适的.

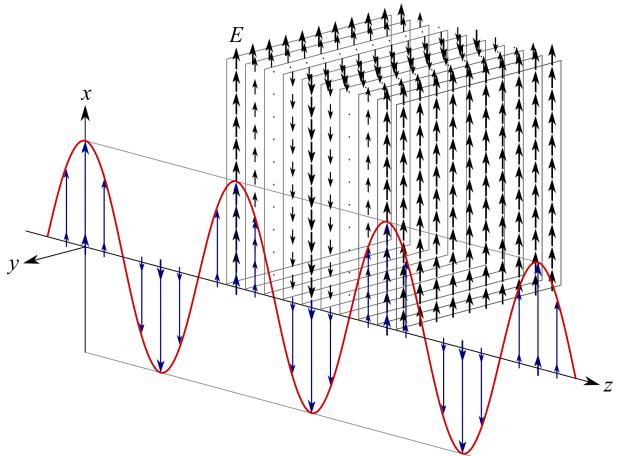


图 11.2: 平面偏振电磁波

真空中的平面波(plane wave) 提供了一种最为简单的情形, 如图, 此时 $\mathbf{E} = E\mathbf{e}_x \cos(\omega t - kz)$. 按照光学惯例, 我们把它写作一个复数, 模就是在每一点处场强的振幅, 而幅角我们取为 \cos 里的相位宗量的相反数, 而它的实部表示物理上的场强:

$$A(z, t) = E e^{i(kz - \omega t)}$$

有几点值得注意: 首先,之所以取相反数,是因为在频率甚高的光频波段我们不太用关心每一个时空点处场随时间 t 的变化,而是不同时空点 z 对振幅与相位的影响,故把 kz 项提前. 而对于 ωt 项由于每一点的振动都含这个随时间振动的因子,而且角频率都是 ω (单色波),所以我们往往又省去这个因子写成 $A = E e^{ikz}$, 相当于仅考虑 $t = 0$ 时的初相位与振幅. 这种操作意味着指数上的宗量随时间减小,这与我们平时对相位随时间增加的理解是相反的. 其次, 我们取消了其矢量性, 我们知道沿特定方向 z 传播的平面单色电磁波可以分解为 x 方向的偏振与 y 方向的偏振, 两者的同相反相组合可以造成不同方向的线偏振模式(linear polarized mode), 而相差 $\frac{\pi}{2}$ 的组合则可以造成不同的圆偏振模式(circular polarized mode). 但这些只有在特殊的光学器件下进行偏振光干涉实验才能发现区别, 故在一般的几何光学条件下我们只关心光的传播方向, 相位与振幅信息,而不关心其偏振时用标量 A 来代表任意模式下约化的光的振幅大小, A 具有电场强度的量纲¹, 相同的 A 代表相同的光强与能流. A 被称为标量振幅(scalar amplitude).

平面波是其他一大类波在局部可以采取的近似, 我们把等相位的面在局部近似为平面, 则波传播方向为垂直于平面的法线方向, 用波矢(wave vector) \mathbf{k} 代表这个方向, 它的大小就是上述平面波的 k , 代表单位长度所改变的相位大小, 正如角频率(angular frequency) 代表单位时间所改变的相位大小. 两者具有如下关系:

$$\frac{\omega}{k} = \frac{\nu}{\sigma} = \frac{\lambda}{T} = \frac{c}{n}$$

其中 ν 表示频率(frequency), 单位时间所经历的振动周期数, 由于一个周期相位改变 2π , 故 $\omega = 2\pi\nu = \frac{2\pi}{T}$, 同理, σ 表示波数(wave number), 有 $k = 2\pi\sigma = \frac{2\pi}{\lambda}$, λ 表示波长(wave length) 而 T 表示周期(period). 这里考虑到介质中的传播, 波速小于真空中的波速, n 为介质的折射率(index of refraction).

而我们所要研究的更加普遍的一类光波一般写为:

$$A(\mathbf{r}) = E(\mathbf{r}) e^{i\varphi(\mathbf{r})}$$

这类光波的共性是仍然以统一地角频率 ω 振动, 也就是所谓的单色波(monochromatic wave)². 所以我们省去包含时间的因子 $e^{-i\omega t}$, 直接写出只依赖于位矢 \mathbf{r} 的波动形式. 一般来说 E 是 \mathbf{r} 的缓变函数, 而 φ 则是 \mathbf{r} 的快速变化函数, 且其变化要与波矢大小 $k = \omega/c$ 相吻合:

$$\nabla\varphi = \mathbf{k}$$

例如, 波源在 $\mathbf{r} = 0$ 的球面波(spherical wave) 有 $E = \frac{u}{r}$, $\varphi = kr$, 而波源在 $\rho = 0$ 的柱面波(cylindrical wave) 有 $E = \frac{u}{\sqrt{\rho}}$, $\varphi = k\rho$.

¹这是因为与物质发生相互作用而感光时电场的作用占据主导地位.

²此时波的波动方程变为亥姆霍兹方程:

$$\left[\nabla^2 - \frac{1}{(c/n)^2} \frac{\partial^2}{\partial t^2} \right] A e^{-i\omega t} = 0 \Rightarrow (\nabla^2 + k^2) A = 0$$

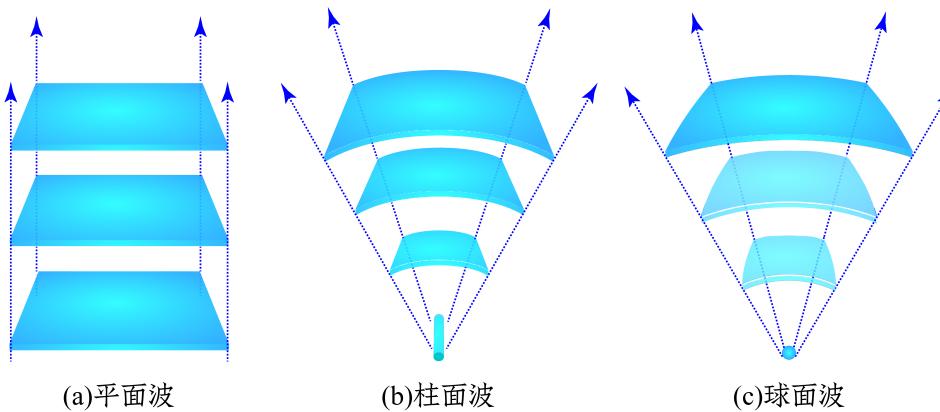


图 11.3: 三种常用波形式

显然这些理想的波不能看成是生活中的一般光源的波特性,一言以蔽之,日常光波具有很差的相干性(coherency).来自不同位置的点光源的不同偏振,不同频率,不同持续时间的波列非相关叠加,使得需要在十分仔细的实验条件下才可以观察到干涉衍射等现象.但这些对一般的几何光学性质没有影响,几何光学关心光的传播与光的强度,于是抽象出所谓的光线(ray)的概念,光线 L 始终沿着波传播的方向,与等相位面垂直:

$$L : \quad \mathrm{d}l // \mathbf{k} = \nabla \varphi$$

对于以上理想的波光线的方向是显而易见的(图中虚线方向),而强度(intensity),可以定义为 $I = A^2$ 平面波不变,柱面波与半径成反比,球面波则与半径的平方成反比.而日常生活中的线状点状光源其性质在几何光学意义下是完全相同的.我们也可以把光的传播视为光线的集合,如果遇到界面则会发生折射与反射,折射率的不均匀导致光线的偏折,而波前可以被部分的障碍物遮挡,没被遮挡的部分仍然以光线的方式向前传播.

也就是说,在几何光学研究中,光线是表,其实作为电磁场的光波才是里.光在两种介质表面的反射(reflection)与折射(refraction)就是波动理论应用的绝佳例证.

11.1.2 菲涅尔公式

如图,在一个典型的界面上反射与折射的过程中,反射光,折射光的强度实际上是与入射光的偏振状态有关的.这一点在接下来的计算中也可以表明.我们按照惯例,取两个独立的线偏振为:

- 横电波(TE wave),即垂直偏振波(³s-wave, s-polarized wave),它的电场方向垂直于主平面⁴.而磁场方向则在主平面内.
- 横磁波(TM wave),即平行偏振波(⁵p-wave, p-polarized wave),它的电场方向平行与主平面.而磁场方向垂直主平面.

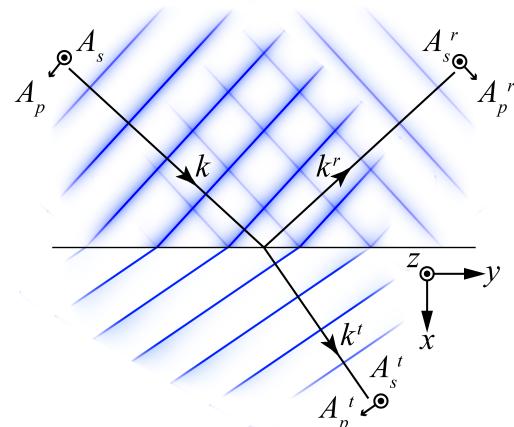


图 11.4: 界面与两种模式

³来自德语词 senkrecht: 垂直.

⁴由入射线与界面法线确定的平面,显然,出射波与折射波光线方向也在这个平面内.

⁵来自德语(英语)词 parallel: 平行.

它们的正方向如图所示, 满足 $\mathbf{A}_s \times \mathbf{A}_p // \mathbf{k}$. 而反射折射系数则定义为在 $(0, 0, 0)$ 点处的界面两侧各光振幅比:

$$A_{s,p} = E_{s,p} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}, A_{s,p}^r = E_{s,p}^r e^{i(\mathbf{k}^r \cdot \mathbf{r} - \omega^r t)}, A_{s,p}^t = E_{s,p}^t e^{i(\mathbf{k}^t \cdot \mathbf{r} - \omega^t t)}$$

$$r_{s,p} = \frac{E_{s,p}^r}{E_{s,p}}, t_{s,p} = \frac{E_{s,p}^t}{E_{s,p}}$$

对于 s -光, 电磁场在界面两侧切向电场连续, 切向磁场强度连续⁶. 而对于 p -光, 切向也有电场连续, 法向可以用电位移连续. 注意到边界条件对任意 t 成立, 这给出了:

$$\omega = \omega^r = \omega^t$$

也就是折射反射不能改变光的频率, 也就是颜色. 从光量子论的角度理解这代表弹性(elastic)作用, 光子的能量将不受改变, 介质不会以某种共振的方式吸收入射光的能量. 上式还意味着 $k = \frac{\omega}{c/n} = n \frac{\omega}{c} = nk_0 \propto n$, 介质内的波矢比同样颜色的真空光要大折射率 n 倍.

而这个边界条件对任意界面 y 坐标也是恒成立的, 这给出了:

$$k_y = k \sin \theta = k^r \sin \theta^r = k^t \sin \theta^t$$

即折射时平行于界面方向的波矢是不变的. 光量子论下可以理解为这个方向动量不变, 没有受到沿平行界面方向的相互作用. 但注意到 k 与 n 的正比关系, 将折射方折射率记做 n' , 折射角记做 θ' 我们得到了著名的斯涅耳定律(Snell's Law):

$$\theta^r = \theta; n \sin \theta = n' \sin \theta'$$

再注意到在光学问题中我们一般只讨论透明介质, 它们磁导率十分接近真空 $\mu \approx \mu_0$, 在此时实际上 $\frac{c}{n} = \frac{1}{\sqrt{\epsilon \mu}}$ 给出了 $n = \sqrt{\frac{\epsilon}{\epsilon_0}} = \sqrt{\epsilon_r}$. 而此时实际上电位移振幅就是 $D = \epsilon E = n^2 \epsilon_0 E \propto n^2 E$ 磁感应强度振幅就是 $B = \sqrt{\epsilon \mu} E = \frac{n}{c} E \propto n E$, 磁场强度 H 与它只差一个真空磁导率 μ_0 . 依此我们写出 s -光与 p -光的边界条件:

$$E_s + E_s^r = E_s^t; n E_s \cos \theta - n E_s^r \cos \theta = n' E_s^t \cos \theta'$$

$$E_p \cos \theta - E_p^r \cos \theta = E_p^t \cos \theta'; n^2 E_p \sin \theta + n^2 E_p^r \sin \theta = n'^2 E_p^t \sin \theta'$$

解之, 我们得到的则是著名的菲涅尔公式(Fresnel's equations):

$$r_s = \frac{n \cos \theta - n' \cos \theta'}{n \cos \theta + n' \cos \theta'}; t_s = \frac{2n \cos \theta}{n \cos \theta + n' \cos \theta'}; t_s - r_s = 1$$

$$r_p = \frac{n' \cos \theta - n \cos \theta'}{n' \cos \theta + n \cos \theta'}; t_p = \frac{2n \cos \theta}{n' \cos \theta + n \cos \theta'}; \frac{n'}{n} t_p - r_p = 1$$

都满足:

$$n(1 - |r_{s,p}|^2) \cos \theta = n' |t_{s,p}|^2 \cos \theta'$$

其中反射率(reflectance) 与透射率(transmittance) 分别被定义为:

$$R_{s,p} = |r_{s,p}|^2; T_{s,p} = \frac{n' \cos \theta'}{n \cos \theta} |t_{s,p}|^2$$

透射率与透射振幅系数 $t_{s,p}$ 关系前多出来的系数, 一是由于在介质中能流密度正比于折射率 $I \propto |\mathbf{E} \times \mathbf{H}| \propto n E^2$, 二是折射后方向改变, 两侧对应光束部分的横截面积也会发生改变. 下图展示了空气 ($n = 1$) 与玻璃 ($n = 1.5$) 界面上的外入射(external incindition) 与内入射(internal incindition) 时的情形.

⁶法向电位移为零.

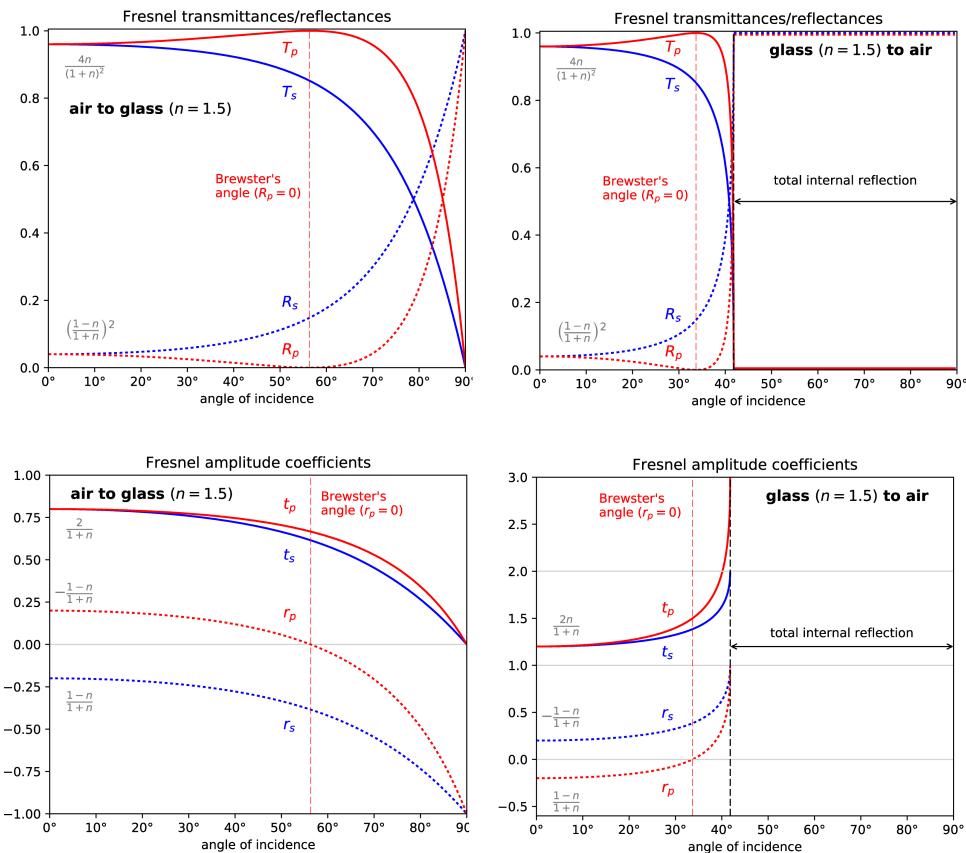


图 11.5: 界面反射与折射

有几点值得注意:

半波损失发生条件

当正入射时, 反射光可以与入射光发生叠加. 我们关心在 $x = 0$ 处, 也就是界面上的叠加结果. 当 $n' > n$ 时, 由上图可见 $r_p > 0$ 而 $r_s < 0$. 但注意到在此处我们预设的入射反射电场正方向对 p -光本来也相反而 s -光相同. 故事实上都是相干相消. 发生了半波损失(half-wave loss). 而若 $n' < n$ 时情况恰好相反, 两种光都是相干相长, 没有半波损失.

而对于入射角为 90° 的掠入射情形, 观察可知 $r_{s,p} = -1$, 即反射光完全与入射光交叠区域几乎完全相干相消. 这个结论对两种偏振模式与两种入射方式均给出了半波损失的结果.

注意透射光几乎在所有情形下都与入射光同相. 唯一的例外发生在全内反射情况下. 之后将仔细讨论之.

布儒斯特角

注意到:

$$r_p = \frac{n' \cos \theta - n \cos \theta'}{n' \cos \theta + n \cos \theta'} = \frac{\sin \theta \cos \theta - \sin \theta' \cos \theta'}{\sin \theta \cos \theta + \sin \theta' \cos \theta'} = \frac{\tan(\theta - \theta')}{\tan(\theta + \theta')}$$

可以发现在 $\theta + \theta' = \frac{\pi}{2}$ 时 $r_p = 0$, 横电的 p -光将发生消光的情形. 这种特殊的情况在内入射与外入射时都可以发生. 对应的入射角称为布儒斯特角(Brewster's angle). 其原因可以这样理解: 将入射波变为折射波与反射波实际上是介质电偶极子在电场

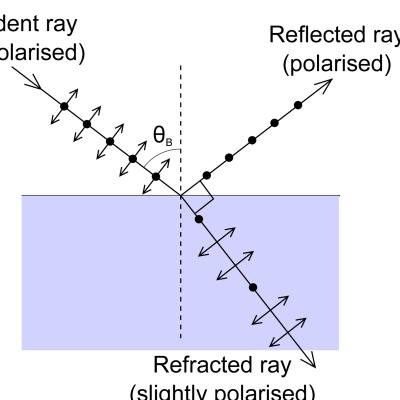


图 11.6: 布儒斯特角

驱动下振动的集体结果. 而对于 p -光, 反射方向恰恰是介质内电场激发的偶极振动的振动方向. 按照电磁波的辐射理论, 这个方向上恰好是辐射强度为零的. 所以恰好反射波消失.

依照这一种特性制作出来的布儒斯特窗(Brewster window) 常见于各种气体激光器的内置谐振腔或发射口. 它将一块透明介质倾斜至布儒斯特角的位置, 它有效地将 s -光进行反射, 而 p -光没有反射的直接发射, 使得出射激光一般都是沿 p -方向高度偏振的.

全内反射与隐失波

在内入射状态下如果入射角足够大, 根据斯涅耳定律将无法确定折射角 $\sin \theta' > 1$. 这种情形为全内反射(total internal reflection). 因为接下来的计算可以表明反射系数 $R_{s,p} = 1$. 而实际上此时并不是完全没有折射光场, 而是折射光场被局限在有光照射的表面区域无法传播开来.

首先, 折射光场由于边界条件在边界上每个时空点都成立, 它依然符合:

$$\omega' = \omega, \quad k'_y = k_y = k \sin \theta$$

但此时由于 x, y 方向波矢仍然满足 $k_x'^2 + k_y'^2 = k'^2$, 而由于入射角足够大使得 $k_y > k'$, 所以实际上 k'_x 只能成为一个虚数:

$$k'_x = \pm i\kappa; \quad \kappa = \sqrt{k^2 \sin^2 \theta - k'^2} = \frac{\omega}{c} \sqrt{n^2 \sin^2 \theta - n'^2}$$

取正号还是取负号呢? 这实际上取决于我们想研究的问题. 在此处我们关心 $k_x = +i\kappa$ 的解. 这是因为这使得相位因子 e^{ik_xx} 成为衰减因子 $e^{-\kappa x}$ 而不是正的指数爆炸因子⁷. 这样, 再将菲涅尔公式中的各个 $\cos \theta$ 理解为 $\frac{k_x}{k}$, 我们得到:

$$r_s = e^{-2i\delta_s}, \quad t_s = \frac{2n \cos \theta}{\sqrt{n^2 - n'^2}} e^{-i\delta_s}, \quad \delta_s = \arctan \frac{\sqrt{n^2 \sin^2 \theta - n'^2}}{n \cos \theta}$$

$$r_p = e^{-2i\delta_p}, \quad t_p = \frac{2nn'}{\sqrt{(n^2 - n'^2)(n^2 \tan^2 \theta - n'^2)}} e^{-i\delta_p}, \quad \delta_p = \arctan \frac{n \sqrt{n^2 \sin^2 \theta - n'^2}}{n'^2 \cos \theta}$$

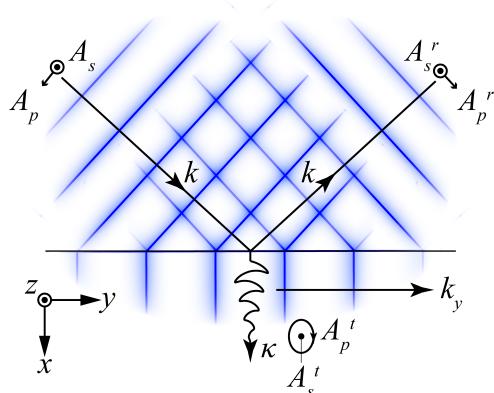


图 11.7: 全内反射与隐失波

可见, 此时最显著的特点除了反射振幅系数的模为一, 就是反射波将会与入射波拉开一个非 0 非 π 的相位差 $2\delta_s$, 它随着从临界全内反射到掠入射从 0 不断增加至 π . 而折射波相位则恰好介于两者之间, 其振幅在临界全内反射时有极大值, 此后随着入射角增加而不断减小至零.

更细致的分析可以发现折射光场的 p -光模式实际上是 $x-y$ 平面上的顺时针椭圆偏振光. s -光模式则显然还沿垂直平面的 z 方向. 两者都要在 x 方向上做衰减, 指数衰减可以表示为:

$$A(x) = A(0)e^{-\frac{x}{H}}$$

其中 $H = \frac{1}{\kappa} = \frac{\lambda_0}{2\pi\sqrt{n^2 \sin^2 \theta - n'^2}}$ 称为衰减深度(attenuation depth) 或趋肤深度(skin depth), λ_0 是真空中的该色光波长. 由表达式可见其一般 H 就光的波长的数量级. 正是因为这个原因这种折射波被称为隐失波(evanescent field). 它一定会产生但是由于存在厚度很薄所以难以被探测到. 如果在隐失波未衰减到零处又重新进入下一层折射角正常的介质, 那么将会有有限光强的光出射出来. 这种现象叫做隐失波耦合(evanescent wave coupling). 在光量子论的角度看这实际上是光子的量子隧穿(quantum tunnelling) 行为.

⁷这意味着下方的场才是需要关心的而界面上的场仅仅是弱地可以忽略的边缘效应.

值得指出,如果入射光并不是无限宽的平面波,而是可以确定其入射点位置的窄光束,而且也不是时刻连续的光,而是一个光脉冲,那么它的行为体现出更加有意思的结果.首先是著名的古斯-汉森效应(Goos-Hänchen effect),光脉冲似乎能进入下层介质,并在其中传播一段时间,随后在另一处离开下层介质出射.如图所示,这造成的前后光束中心位置的横向移位称为古斯-汉森位移(Goos-Hänschen shift).这个位移的计算需要把入射光束看成有一定角度展宽的单色平面波的叠加,而在入射点处恰好各个平面波相干相长,而光束出射点也是各个出射平面波相干相长的位置,不难说明:

$$\Delta_{s,p} = 2 \frac{d\delta_{s,p}}{dk_y} = \frac{\lambda \cos^2 \delta_{s,p}}{\pi \cos \theta} \cdot \frac{d \tan \delta_{s,p}}{d\theta}$$

计算可得:

$$\Delta_s = \frac{2H \tan \theta}{n} ; \quad \Delta_p = \frac{\Delta_s}{(1 + \frac{n^2}{n'^2}) \sin^2 \theta - 1}$$

另一种效应是英伯特-费多罗夫效应(Imbert-Fedorov effect).如果入射光是两种圆偏振光,那么可以证明隐失波场将具有垂直于纸面方向的能流,如果入射光是窄光束,那么反射光束将在垂直于纸面的方向发生位移,这个结果与光的角动量守恒相协调.在这里不加论述,详情参见光学教材.

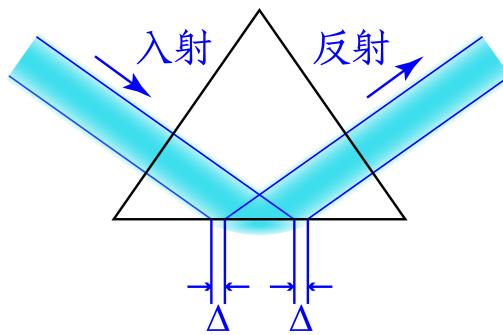


图 11.8: Goos-Hänchen Shift

11.2 光线方程

11.2.1 光线方程与折射定律

在给定折射率 n 随着空间位置连续变化的情况下,光将如何进行传播呢?把光的传播简化为光线,光线垂直于等相位面,那么光线方程(ray equation)给出了这个问题的答案.这个方程写作:

$$\frac{d}{ds}(n\tau) = \nabla n$$

其中 ds 表示光线经过的路程.我们来考虑证明这个式子.首先我们先观察上式在法向与切向的两个分量式:

$$(\frac{dn}{dl})_\tau \tau = (\nabla n)_\tau \tau$$

$$n \frac{d\tau}{ds} = n \frac{d\theta}{ds} \mathbf{n} = (\nabla n)_n \mathbf{n}$$

那么第一个式子是一个重言式(恒等式),而第二个式子则给出了由光线传播方向(给出了切向量 τ 与法向量 \mathbf{n})和局域折射率分布(给出了 ∇n)决定的光线弯曲程度(有曲率 $k = \frac{d\theta}{ds}$ 描述).是名副其实的能够决定光线传播规律的光线方程.事实上,如图,如果把在该点局域处折射率分布理解为沿 ∇n 的方向进行分层,把光线传播方向与局域统一的 ∇n 方向的夹角记为 φ ,则当光线走了 ds 时折射率变化为:

$$dn = \nabla n \cdot d\tau = |\nabla n| ds \cos \varphi$$

由折射定律:

$$d(n \sin \varphi) = 0 \Rightarrow d\varphi = -\frac{\sin \varphi dn}{n \cos \varphi} = -\frac{|\nabla n|}{n} ds \sin \varphi$$

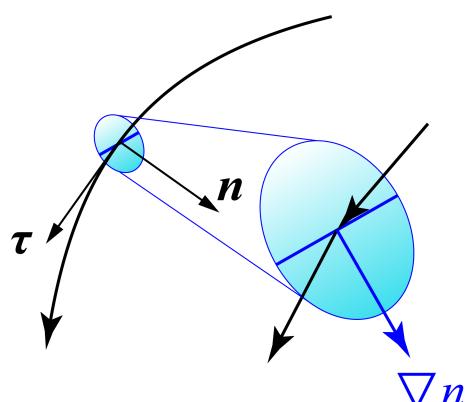


图 11.9: 光线方程推导

这便是:

$$n \frac{d\theta}{ds} = n \frac{|d\varphi|}{ds} = (\nabla n)_n$$

从而给出了光线方程的表达式.

具体的,如果研究在二维平面上,折射率 n 仅仅与 y 有关而与 x 无关的问题,如稳定大气中光线的传播问题,此时原光线方程化为:

$$\frac{d}{ds}(n \sin \theta \mathbf{e}_x + n \cos \theta \mathbf{e}_y) = \frac{dn}{dy} \mathbf{e}_y$$

注意这时候我们一般取 θ 为光线传播方向与界面法线 (y 方向) 间的夹角.上式一方面意味着

$$n \sin \theta = \text{Const.}$$

这可以看成是两层界面上折射定律的推广,它给出了每一个 y 处(因而知道 n) 的光线传播方向,从而本质上余下工作只需要解出具体轨迹即可.只需解方程:

$$\sin \theta = \frac{1}{\sqrt{1+y'^2}} = \frac{(n_0 \sin \theta_0)}{n(y)}$$

其中 $(n_0 \sin \theta_0)$ 是根据初始条件确定的不变量.

11.2.2 光力类比

牛顿提倡光的粒子说,他的粒子光学理论中光粒子⁸运动状态由位置与速度描述,而运动由轨迹方程与运动方程描述.的确,由光线方程可知,这是一个定义良好的动力学方程,它类似于势场中的牛顿第二定律:

$$\frac{d}{dt}(mv\tau) = -\nabla V$$

由粒子的初位置与初速度便可决定论地的给出之后的运动.这一种类似性为我们进行光力类比提供了良好的依据.然而要注意仍然有一个不同之处:光的速度大小不可控.在一个典型的光线求解的过程中,我们知道初始时刻的光的传播方向即可,而传播速度无需给出,事实上它也不能被改变,因为光速是常数,换一个角度理解,改变速度就是改变动量与能量,那么实际上光子的动量能量改变实际上反应在了光的颜色(频率)上.这是与相对论相吻合的一个结论.

那么如何进行光力类比呢?还是从以上两个公式的相似性出发,我们想象有一个质量为 $m=1$ 的粒子在光的真实传播轨迹上运动,规定其速度为 $v=n$ ⁹,那么其受力应该为:

$$\mathbf{F} = \frac{d}{dt}(mv\tau) = \frac{d}{ds/v}(n\tau) = v\nabla n = n\nabla n = -\nabla\left(-\frac{n^2}{2}\right)$$

也就是说如果认为粒子是在 $V=-\frac{n^2}{2}$ 的势场中运动的,那么其模仿光线的轨迹与用该处折射率来定标自己的运动速度的做法又恰巧是符合动力学规律的.恰巧,这个运动的守恒能量:

$$E = T + V = \frac{v^2}{2} - \frac{n^2}{2} = 0$$

光粒子的运动只能是能量为零的,如果忽略光的色散,那么不同频率的光在同一个折射率场中将具有完全相同的传播轨迹,实际上没有光粒子运动的“快慢”的说法.

我们总结光力类比法求光线传播问题的大意:

1. 明确折射率场分布 $n(\mathbf{r})$,明确光线初始入射位置与方向 $\tau_0(\mathbf{r}_0)$.
2. 光力类比:设置势场 $V(\mathbf{r}) = -\frac{n^2}{2}$,设置粒子,质量为 1,初始位置 \mathbf{r}_0 ,速度 $\mathbf{v}_0 = n\tau_0$.

⁸与光子概念有本质区别

⁹注意这不是实际光传播的相速度! $v \propto n$ 而 $v_g = c/n \propto 1/n$

3. 求解粒子的运动, 得出运动轨迹.

4. 这个轨迹也就是光线的轨迹, 再由已知的轨迹求解其他问题 (如传播时间等).

这种方法往往是方便的, 它会使得很多我们在力学问题中有明确结论的结果可以直观地呈示出来. 例如在之前的折射率随 y 改变的情形, 它对应着水平方向平移对称的势场, 从而有 x 方向的动量守恒, 这就是:

$$n \sin \theta = \text{Const.}$$

同样的道理, 如果光线在一个球对称的折射率场 $n(r)$ 中运动时, 其轨迹恰好对应有心力场中粒子运动的轨迹. 例如取:

$$V = -\frac{k}{r} + C \Rightarrow n = n_0 \sqrt{\frac{2a}{r} - 1}$$

那么这个粒子的运动是在环域 $0 < r < 2a$ 上的椭圆, 而且计算可以表明改变初始条件, 其半长轴固定为 a , 保证其轨迹范围不会超过这个环域.

而此时应该有角动量守恒, 用光线方向与径向夹角 θ 表示则为:

$$nr \sin \theta = L = \text{Const.}$$

11.3 费马原理

类似于力学变分原理, 光线方程的存在也意味着光学的变分原理, 这个原理即为著名的费马原理:

$$\delta \int_A^B n ds = 0, \text{ if } \delta \mathbf{r}_A = 0 \text{ and } \delta \mathbf{r}_B = 0$$

如何理解这个原理呢, 首先这个折射率对路程的积分称为光程(optical path), 在光力类比下它实际上就是满足能量守恒条件下的相积分:

$$S = \int_{A \xrightarrow{\Gamma} B} n ds = \int_{A \xrightarrow{\Gamma} B} \mathbf{p} \cdot d\mathbf{r}$$

而从 A 点到 B 点的光传播可以由某种参数方程描述:

$$\Gamma : \quad \mathbf{r} = \mathbf{R}(t), \quad \mathbf{R}(t_A) = \mathbf{r}_A, \quad \mathbf{R}(t_B) = \mathbf{r}_B$$

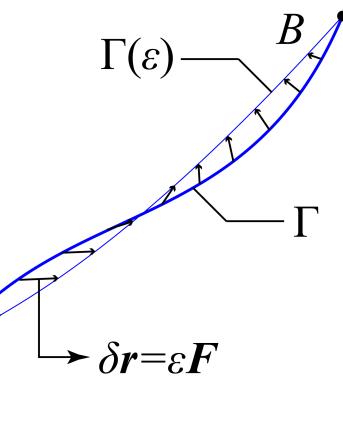


图 11.10: 光程变分

而变分的意思是我们另取一种可能的轨迹, 比如:

$$\Gamma(\epsilon) : \quad (\mathbf{r} + \delta \mathbf{r})(t) = \mathbf{R}(t) + \epsilon \mathbf{F}(t)$$

那么费马原理要求在轨迹端点不变的条件下, 新的光程关于所有可能的微扰方向 $\mathbf{F}(t)$, 关于其微扰强度的一阶导数都要等于零:

$$S(\epsilon) = \int_{A \xrightarrow{\Gamma(\epsilon)} B} n ds \quad ; \quad \left. \frac{dS(\epsilon)}{d\epsilon} \right|_{\epsilon=0} = 0$$

这个原理是等价于光线方程的, 证明这一点需要用到较为专门的矢量分析技巧, 在此略去. 用文字一般可以表述为:

两点之间光线的实际轨迹, 是光程取平稳值的路径.

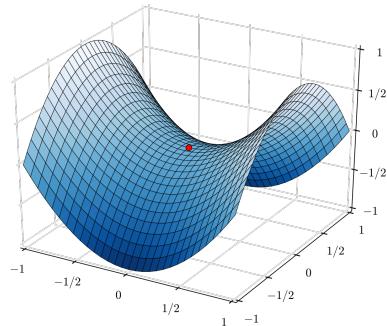


图 11.11: 二维鞍点处势取平稳值

其中的“平稳值”的说法,既包含了极大,也包含了极小,还包含了一些复合的情况,比如如图,我们考虑一个马鞍面上有一个质点在鞍点处于平衡状态,那么在一个方向上势能最高,另一个方向上势能最低,然而该点能平衡,因为势能取得了“平稳值”.二自由度下已经出现了两个自由度方向微扰后极大极小性质出现差异的情形,而光程问题则是在所有可能的轨迹空间中进行无穷多自由度的微扰的问题,普遍来说符合之前我们定义的一阶变分为零即可以符合费马原理的条件,故冠以“平稳值”的说法.

最后,值得指出以上结论都是在波动理论尚未发展到一定程度时的替代性的粒子理论.在有了波动理论以后以上两个结论都将得到极为简单的解释.我们把变折射率问题中的光的等相位面找到,光线即将垂直于等相位面.而相位场¹⁰ $\varphi(\mathbf{r})$ 与光传播方向,折射率间关系即为:

$$\nabla\varphi = \mathbf{k} = k_0 n \boldsymbol{\tau}$$

该式两边同时对传播方向取方向导数:

$$\frac{d}{ds}(\nabla\varphi) = k_0 \frac{d}{ds}(n\boldsymbol{\tau})$$

而此式左边,在沿传播方向求的导数应该理解为与下一个梯度算符独立的,故可以交换顺序:

$$\frac{d}{ds}(\nabla\varphi) = \nabla\left(\frac{d\varphi}{ds}\right) = \nabla k = k_0 \nabla n$$

两式联立便得到光线方程:

$$\frac{d}{ds}(n\boldsymbol{\tau}) = \nabla n$$

而费马定理也非常容易理解:光程积分:

$$\int n ds = \int n \boldsymbol{\tau} \cdot d\mathbf{r} = \frac{1}{k_0} \int \nabla\varphi \cdot d\mathbf{r} = \frac{\Delta\varphi}{k_0}$$

所以在光程微扰下只要起点与终点固定,光程就不会有改变.

¹⁰ 在粒子论里称为程函(eikonal)

第十二章 光学成像

12.1 傍轴光成像

12.1.1 物与像

成像不是一个平凡的过程. 延续上一章我们对变折射率问题的讨论, 现在我们引入所谓物(object)的概念, 它指的是一个点光源 S , 它能够向各种不同的方向发射光线, 每一个既定的方向则可以根据光线方程求解之后的光线传播. 这样它所发出的光场构成了物的光场. 而传播过程中到达的任何一点 P , 满足 S 与 P 间的真实光线有且只有一条. 它恰好由费马原理确定. 但传播过程中可能发生一种奇异的现象, 就是在空间中另一点 S' 处成了像(image). 从光线角度看, 便是:

一定范围内, 物发出的每一条光线都经过像.

而根据费马原理, 真实光线旁边的微扰光线光程对此光线的一阶变分为零, 但此时有一大组真实光线, 彼此之间可以连续的过渡, 我们必然得出:

物与像之间的所有真实光线, 其光程都是相等的.

以上就是物像之间的等光程原理(the principle of equal path length). 它实际上也是作为能否成像, 成像在何处的判据. 这一个结论也可以依靠干涉理论, 或者更普遍的, 光场概率幅的路径积分理论来理解: 如果两点间光程沿某种路径不是取得极值, 那么这种情形将会由于存在很多其他类似的路径, 使得最后在终点处贡献的概率幅相干相消. 仅仅是在光程取得极值的路径将对最后终点处的概率幅有主要贡献. 但在物像点间如果存在大量路径其光程完全不变, 最后像点的概率幅将会非常地大, 这也就是成像的过程了.

以上描述的实际上只是几何光学中的实物与实像的概念. 在一般的光学系统中, 光路都是傍轴的. 即满足傍轴条件(paraxial condition): 有一条中心光线作为主光轴(optical axis) 参考. 光具组(optical system) 把光路分为物方(object space) 与像方(image space). 两个空间一般都是均匀折射率空间, 故一般考虑光锥(pencil)型光束的传播(就是球面波). 注意两个空间有可能重合(反射情况). 这样的光束偏离主光轴的角度满足:

$$\theta \ll 1$$

或者表述为

$$\frac{z}{l} \ll 1$$

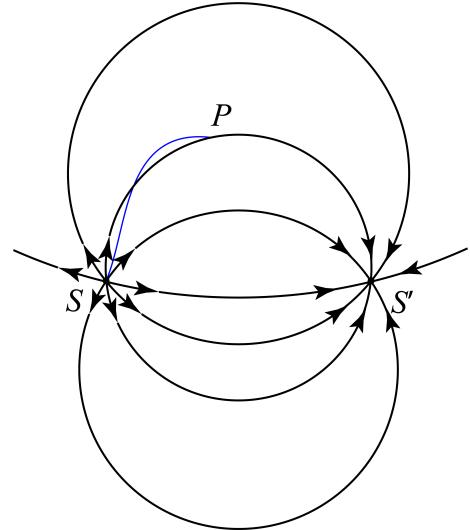


图 12.1: 物与像

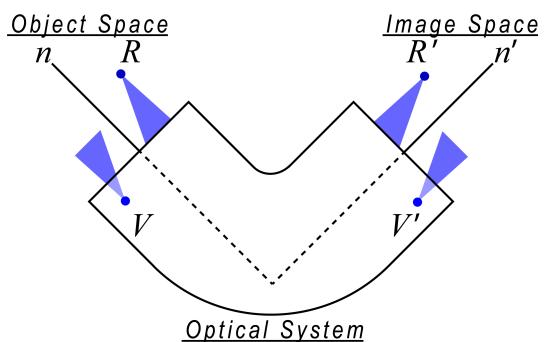


图 12.2: 实与虚

其中 z 是光路的特征偏离主光轴的距离. 而 l 是光路沿主光轴传播的特征长度. 如果进入光具组时光束呈现发散(diverging)态, 则说明进入光具组前已有物, 尽管这一个物有可能也不在真实空间中(如前一次经过光学组后成虚像作为这次的实物), 也把这样的物称为实物(real object). 它一定位于物方. 而相反的情况是进入光具组时为会聚(converging)光的情况, 那么物就不可能在物方找到了, 它位于光具入射面的另一侧(不一定是像方), 这样的物称为虚物(virtual object). 同理, 对于像, 出射光会聚的像位于像方, 是实像(real image), 而出射光发散, 像位于像方的另一侧, 则是虚像(virtual image).

12.1.2 球对称成像系统与符号法则

顾名思义, 球对称成像系统指的是主光轴在一条直线上而整个光具组具有对主光轴的旋转对称性. 当然实际光锥光路不一定要求关于主光轴对称. 两种典型的基础光具是折射球面与反射球面. 如下图所示:

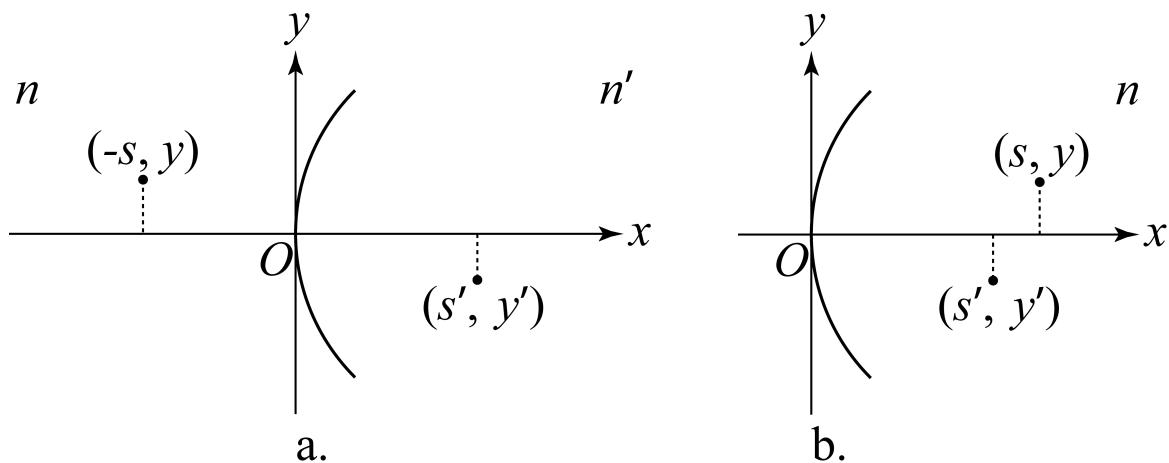


图 12.3: 球面折射与球面反射

在球面折射情况12.3下, 我们先考虑向像方凹的, 曲率半径为 R 的近似球面的傍轴折射成像(作左图). 各量的符号设定如图, s 为物距(object distance), s' 为像距(image distance), y, y' 则为物, 像的高(height). 由物像的等光程性, 应有:

$$f(X, Y) = n\sqrt{(X + s)^2 + (Y - y)^2} + n'\sqrt{(X - s')^2 + (Y - y')^2} = \text{Const.}$$

由于傍轴条件, 上式中参量 Y 是理解为小量变量即可, 而 X 近似为:

$$X = \frac{Y^2}{2R}$$

代入, 小量近似, 表示为 Y 的二次多项式:

$$\begin{aligned} f(X, Y) &= n\sqrt{s^2 + y^2} + n'\sqrt{s'^2 + y'^2} \\ &\quad - \left(\frac{ny}{\sqrt{s^2 + y^2}} + \frac{n'y'}{\sqrt{s'^2 + y'^2}} \right) Y \\ &\quad - \left[\frac{n[(1 + s/R)(s^2 + y^2) - y^2]}{2(s^2 + y^2)^{3/2}} + \frac{n'[(1 - s'/R)(s'^2 + y'^2) - y'^2]}{2(s'^2 + y'^2)^{3/2}} \right] Y^2 \end{aligned}$$

从而一次项与二次项系数都必须为零. 由傍轴条件, s, s' 是 u, v 的小量. 从而整理得:

$$\begin{aligned} \beta &= \frac{y'}{y} = -\frac{ns'}{n's} \\ \frac{n}{s} + \frac{n'}{s'} &= \frac{n' - n}{R} \end{aligned}$$

上式给出了成像的横向放大率(transverse magnification)与物像距关系(object-image distance relation),抽象地看,在傍轴加二阶近似意义下物方空间每一点都能成像.而给定任意物距处的像距与横向放大率,就决定了一个光学系统傍轴成像的所有特性.以上成像的两个公式的特点是:物像距都从球面的顶点开始计算,这个点称为光心(optical centre).这两个公式称为高斯型公式(Gaussian formula).一般把第二个物像距公式右端的结果定义为光焦度(focusing power):

$$\Phi = \frac{n' - n}{R}$$

其单位是屈光度(diopter),中文简称“度”. $1D = 1m^{-1}$.

当物距和像距取特殊值时会出现另一个趋于正负无穷,而对应的物像高也趋于正负无穷的情形,此时意味着入射,出射光成为了平行光(collimated light).这些特殊的物像距称为焦距(focal length),而对应的主光轴上的点称为焦点(focus).物方焦距与像方焦距就分别为:

$$f = \frac{n}{\Phi} = \frac{nR}{n' - n} ; f' = \frac{n'}{\Phi} = \frac{n'R}{n' - n}$$

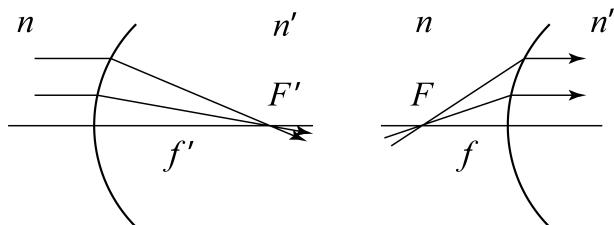


图 12.4: 焦点

自然地有:

$$f : f' = n : n'$$

通过引入两个焦距我们可以把成像公式写作:

$$\frac{f}{s} + \frac{f'}{s'} = 1 ; \beta = -\frac{f s'}{f' s}$$

实际上, n, n' 都不是通过这个折射成像系统可以测量的量,而两个长度量纲的 f, f' 代表了这个系统成像的所有特征.

为了更直观地反应物像距指尖关系的特性,一般把顺光传播方向,物到物方焦点的距离记为另一种物距 $x = s - f$,而像方焦点到像的距离记为另一种像距 $x' = s' - f'$.那么代入高斯型公式化简可得:

$$x x' = f f' \\ \beta = -\frac{f}{x} = -\frac{x'}{f'}$$

以上两个公式更为精简,称为牛顿型公式(Newtonian formula).图像也更加清晰:物像距恰好是一个反比关系.而横向放大率取决于物离物方焦点有多近,也等效于像离像方焦点有多远.运用起来往往更加方便.仔细观察还会发现,仅需 ff' 作为整体的大小实际上已经确定了物像距的关系.但如果要求的是横向放大率, f, f' 各自取多少就是必要的了.也就是说不仅需要知道两个焦点的位置,还得已知光心离各自有多远.

对于12.3右图的球面反射镜,只需要把两个焦距定为

$$f = f' = \frac{R}{2}$$

那么高斯型公式与牛顿型公式都以完全一样的形式适用,只需注意此时的主光轴正方向在反射时发生了反转,物像距的定义会产生相应的改变,参照图理解即可.

当我们考虑不同的球面凹凸性,不同的折射率分布,不同的物像实虚性时显然不希望每次都重新进行推导,而是希望我们之前推出的结论具有普适性.那么实际上之前我们涉及的每个量都可以有正负(折射率除外).约定不同量的正负性的方法称为符号法则(sign convention).本书介绍一下符号法则:

1. 光具球面曲率半径:出射方看,凹正凸负.出射方也就是单次成像的像方,如果曲率中心位于这一侧,那么曲率半径计为正.如果在另一侧,计为负.

2. 物像距, 焦距: 实正虚负. 如果是物点发出的发散光沿主光轴正向传播到物距的零点处则为实物, 物距计正 (实物). 如果是会聚光传播到物距零点处还未到达物点, 则物距计负 (虚物). 同理, 像距零点处出射会聚光可以沿主光轴传播形成像点, 像距计正 (实像). 若像距零点处出射发散光, 像点可以认为应在之前形成, 像距计负 (虚像). 物方焦距是使像距趋于无穷的特殊物距. 像方焦距是使物距趋于无穷的特殊像距.
3. 物像高: 上正下负. 不管任何情形, 都规定主光轴的一侧高度大于零, 一侧小于零. 也就是把与这个方向垂直的轴上坐标定义为高度. 如果把主光轴理解为水平方向, 一般上半空间物, 像高为正, 下半空间物像高为负.
4. 光线的倾角: 增正减负. 有时候我们不仅关心物像点间的对应关系, 也关心一根光线与另一根光线的对应关系. 而实际上相互对应的光线不仅是物方入射光线与像方出射光线的关系, 由于每一点某能成像, 两条线上的点也是一一对应的物像点. 此时光线的角度被定义为沿传播方向单位长度所增加的物或像高. 如果光线从下往上传播, 角度为正, 从上往下则角度为负.

我们之前推出来的, 针对于球面折射, 反射的公式自然是在以上符号法则下普适的. 同样的道理也适用于之后所有情况下提出的公式. 而且不失一般性地, 高斯型公式与牛顿型公式其实是对任意球对称成像系统傍轴适用的. 推广到非傍轴情形便是下一节讨论的理想成像系统.

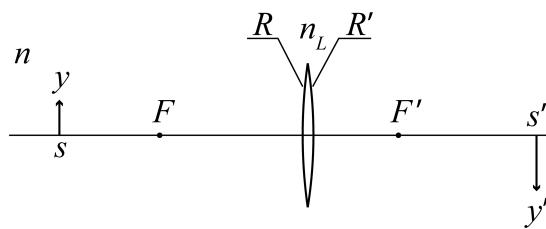


图 12.5: 薄透镜

我们接着讨论**薄透镜**(thin lens) 模型. 这是一个两次成像的过程, 第一次成像的像方零像距点, 第二次成像的物方零物距点恰好重合在透镜的光心位置. 从而有:

$$\frac{n}{s} + \frac{n_L}{s'_1} = \frac{n_L - n}{R} ; \quad \beta_1 = -\frac{ns'_1}{n_L s}$$

$$\frac{n_L}{s_2} + \frac{n'}{s'} = \frac{n' - n_L}{-R'} ; \quad \beta_2 = -\frac{n_L s'}{n' s_2}$$

注意在第二个公式中 R' 是在右表面凸是取正的曲率半径 (以双凸透镜为正), 它与我们既定的符号法则相反, 故在照抄公式时在分母加以符号以示区别. 再有上讨论的两个物像距参考关系:

$$s'_1 + s_2 = 0$$

消去中间的物像距, 得到:

$$\frac{n}{s} + \frac{n'}{s'} = \frac{n_L - n}{R} + \frac{n_L - n'}{R'} ; \quad \beta = -\frac{ns'}{n's}$$

这说明把两次成像的密接的结果是光焦度的直接相加:

$$\Phi = \Phi_1 + \Phi_2$$

而除此以外余下公式都没有变化:

$$f = \frac{n}{\Phi} ; \quad f' = \frac{n'}{\Phi}$$

$$\frac{f}{s} + \frac{f'}{s'} = 1 ; \quad \beta = -\frac{fs'}{f's}$$

这几个规律也适用于密接的透镜组. 任何发散和会聚光的能力直接用光焦度评价, 密接时这些规律都直接叠加.

12.1.3 光具组成像

如图,如果考虑厚透镜(thick lens) 模型,或者更普遍地讨论两个薄透镜相隔一定距离放置的情况.那么一般把前透镜像方焦点到后透镜物方焦点的以传播方向为正的距离即为光学距离(optical distance).即 $\Delta = \overline{F'_1 F_2}$,两透镜间距离:

$$d = f'_1 + f_2 + \Delta$$

上图的 Δ 实际上取负,事实上当 $\Delta = -f'_1 - f_2$ 时变为密接的情形,而 Δ 增大也代表着 d 增大. 我们写出此时的成像物像距公式:

$$\begin{aligned} \frac{n}{s} + \frac{n''}{s'_1} &= \Phi_1 \\ \frac{n''}{s_2} + \frac{n'}{s'} &= \Phi_2 \\ s'_1 + s_2 = d &= \frac{n''}{\Phi_1} + \frac{n'}{\Phi_2} + \Delta \end{aligned}$$

得到:

$$(s - f_1 - \frac{f_1 f'_1}{\Delta})(s' - f'_2 - \frac{f_2 f'_2}{\Delta}) = \frac{f_1 f'_1 f_2 f'_2}{\Delta^2}$$

上式即成像公式的牛顿形式. 而由牛顿型的放大率公式:

$$\beta = \left(-\frac{f_1}{s - f_1} \right) \left(-\frac{s' - f'_2}{f'_2} \right)$$

代入上式可得:

$$\beta = -\frac{-\frac{f_1 f_2}{\Delta}}{s - f_1 - \frac{f_1 f'_1}{\Delta}} = -\frac{s' - f'_2 - \frac{f_2 f'_2}{\Delta}}{-\frac{f'_1 f'_2}{\Delta}}$$

从中可以发现,这个成像公式恰恰等价于一个物像距的零点不位于两个透镜光心. 而两侧的焦点也不与原来的焦点重合的成像系统,仍然有其焦距为:

$$f = \overline{FH} = -\frac{f_1 f_2}{\Delta}, \quad f' = \overline{H'F'} = -\frac{f'_1 f'_2}{\Delta}; \quad f : f' = n : n'$$

而各个特征点位移为:

$$\begin{aligned} \delta_F &= \overline{F_1 F} = -\frac{f_1 f'_1}{\Delta}; \quad \delta'_F = \overline{F' F_2} = -\frac{f_2 f'_2}{\Delta} \\ \delta_H &= \overline{O_1 H} = -\frac{f_1 d}{\Delta}; \quad \delta'_H = \overline{H' O_2} = -\frac{f'_2 d}{\Delta} \end{aligned}$$

其中 H 被称作等效光学系统的物方主点, H' 被称作等效光学系统的像方主点. 以上公式的理解中如果 $\Delta < 0$,那么各个量就都是正的. 图画的正是这个情况. 此时的成像规律,可以认为是在 H 处具有光焦度为 $\Phi = -\frac{n'' \Delta}{f'_1 f'_2}$ 的透镜,再把其直接成像的结果进行一个平移,平移距离即为两个主点间距 $\overline{HH'}$. 正是因为这个原因,考虑多次成像系统的复合时可以再把两个光具等效的光学系统再与第三个光具按照以上公式再进行等效,直到化为最基础的成像公式的情形.

12.2 理想成像系统

把傍轴成像系统的特性进行数学上的抽象与大胆的推广,我们得到了理想光具组成像系统. 它是一个全空间中点到全空间中点的可逆映射:

$$\mathcal{J}: S \rightarrow S', \quad S \in \mathbb{R}^3 \cup \{\infty\}$$

在这个映射中无穷远点也是有意义的.一般正负无穷远处的点其实没有区别(发出平行光),不同的无穷远点根据其发出的光与主光轴的夹角定义.而理想光具组需要有以下特征:

1. 主光轴必须被映射到主光轴.
2. 映射关于主光轴对称. 可以认为是一个平面上的映射绕主光轴旋转而生成的.
3. 保直线, 平面的形状. 直线点集与平面点集映射到直线与平面.

仅这三个条件, 足以把映射的性质确定到仅含极少参数, 也就是我们上一节介绍的成像系统. 这是因为可以根据以上定义证明以下性质 (作为思考题请读者自己思考):

1. 垂直于主光轴的平面一定被映射到垂直于主光轴的平面.
2. 在垂直于光轴的平面之间的映射是简单的放缩变换 (横向放大率为常数).
3. 不同的物距的横向放大率, 要么始终保持为常数, 要么连续, 单次地取遍任意正负值. 前者称为**望远系统**(afocal system).

所以我们把这个映射的性质总结为在一个过主光轴的平面内的问题, 这个平面称为**子午面**(meridinal plane), 只需要物像距公式和放大率公式两个成像公式. 而定义若干的辅助点, 辅助面可以帮我们有效的研究非望远系统的性质. 这些点与面被称作**基点**(cardinal point) 与**基面**(cardinal plane). 除了望远系统, 所有理想光具组成像都有两对关键基点, 它们都是主光轴上的特殊点, 而有基面即过基点的垂直于主光轴的平面, 代表特殊的物像距处物像高的所有可能性. 它们是:

主点 (principal points): 横向放大率为 1 的那对共轭点即为物方主点与像方主点.

焦点 (focal points): 横向放大率发散 (∞) 时, 物方为物方焦点, 像方为无穷远点. 横向放大率为 0 时, 物方为无穷远, 像方则为像方焦点.

有了以上定义以后, 便可以用作图法确定成像公式.

12.2.1 作图法

作图法的基础是以下两个简单结论:

- 凡入射到物方主平面上一点的光线必然在像方主平面上等高处出射.
- 平行光入射, 出射光聚焦于像方焦平面上某点. 物方焦平面上一点入射, 则出射平行光. 更特别地, 经过物方焦点的光线出射光平行于主光轴, 平行于主光轴的光入射后出射光经过像方焦点.

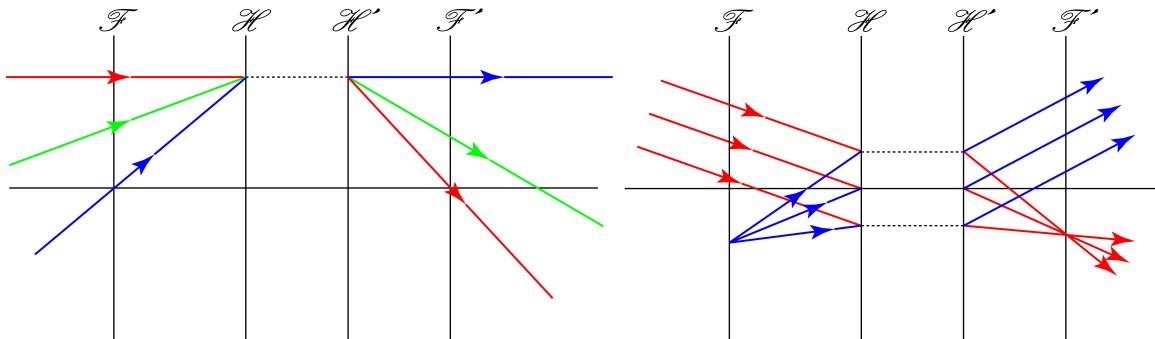


图 12.7: 特殊光线

通过以上特殊光线我们可以确定任意一条入射光线对应的出射光线, 也能进一步明确到每一个物对应的像的位置. 如下图左所示. 给出一条入射光线 AB . 如果它入射到物方主面上的 B 点. 那么必然从像方主面上的等高 C 点出射. 于是做出两条辅助光线: 一是从入射光线经过的物方焦面上的点作平行与主光轴

的光线(红色),那么我们很容易找到它的出射光线.那么原光线与此特殊光线必然出射平行光而找到出射 CD 的方向.二是也可以做出过物方焦点且与 AB 平行的辅助入射光线.也可以找到它的出射光线从而确定出 CD 来.而如果要找物 S 的像 S' 的位置,也是利用特殊的光线确定的,如下图右.通过几何上三角形1与2,3与4的相似关系:

$$\frac{y}{-y'} = \frac{x}{f} ; \quad \frac{-y'}{y} = \frac{x'}{f'}$$

整理即为牛顿型的成像公式:

$$xx' = ff' ; \quad \beta = -\frac{f}{x} = -\frac{x'}{f'}$$

再利用 $s = f + x, s' = f' + x'$ 可以确定从主面出发计算物像距的对应高斯型成像公式:

$$\frac{f}{s} + \frac{f'}{s'} = 1 ; \quad \beta = -\frac{fs'}{f's}$$

- 凡入射到物方主平面上一点的光线必然在像方主平面上等高处出射.
- 平行光入射,出射光聚焦于像方焦平面上某点.物方焦平面上一点入射,则出射平行光.更特别地,经过物方焦点的光线出射光平行于主光轴,平行于主光轴的光入射后出射光经过像方焦点.

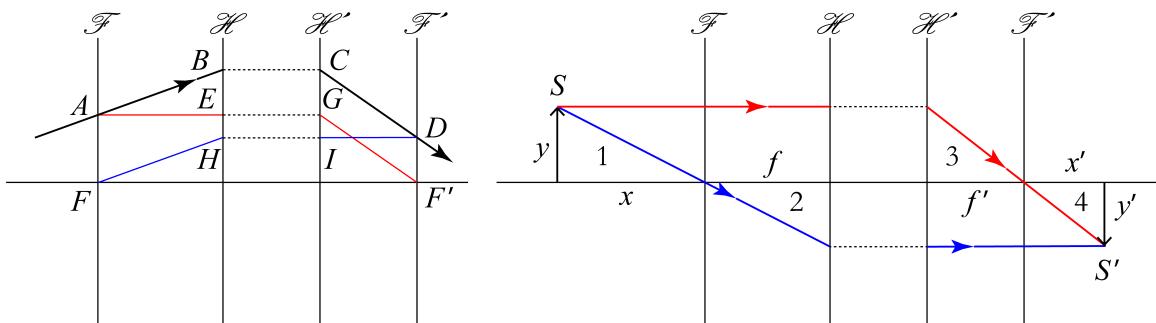


图 12.8: 共轭线与共轭点

12.2.2 基点基面性质

在实际傍轴成像过程中还有一个我们比较关系的量.就是在一定物距与对应的像距处的成像的角度放大率(angular magnification).它被定义为物点傍轴发出一个顶角为 u 小角的光锥时¹,成像处对应的光锥顶角 u' 的放大率.在傍轴的小角近似下,也可以定义为过零物像高处的共轭点的共轭光线的夹角比.后一个定义能产生正负号,我们采用后一种定义(右图中红色的角).

$$B = \frac{u'}{u}$$

而如果考虑的是理想光具组情形,我们可以将上式写为更严格的形式:

$$B = \frac{\tan u'}{\tan u}$$

从图中可以发现,这样定义角度放大率后它与物像距的关系从图中可以写出:

$$s \tan u = s' \tan(-u') \Rightarrow B = -\frac{s'}{s} = \frac{f'}{f} \beta$$

¹这里考虑平面问题,空间情形可以比较类似地讨论,此时应该考虑立体角.

考虑牛顿公式 $xx' = ff'$, 除去焦点与焦面, 还有有四种独特的物像距分别具有独特的性质. 它们对应的主光轴上的点与平面都被称作基点与基面. 而各个点性质如下:

主点 (principal points) : $x = -f, x' = -f' \Rightarrow \beta = 1$. 成等大正立像.

二倍焦距点 : $x = f, x' = f' \Rightarrow \beta = -1$. 成等大倒立像.

节点 (nodal points) : $x = -f', x' = -f \Rightarrow B = 1$. 角放大率为一.

物像距等值点 : $x = f', x' = f \Rightarrow B = -1$. 角放大率为负一.

其中节点概念在光学测量中有重要的作用. 它的特点是, 入射到物方节点 N 光线, 方向不改变地从像方节点 N' 出射. 注意到这个角放大率为一的事实也仅仅在物像距为节点时能够取到. 节点也是经常讨论的关键基点

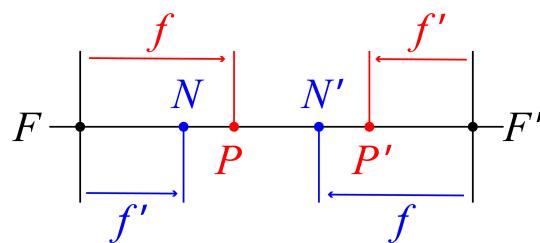


图 12.10: 三对关键基点

我们可以这么理解这三对基点的关系, $xx' = ff'$ 提示我们, 只需要知道 ff' 乘积整体的值便可以确定物像距之间的对应关系, 而两者之比 $f : f' = n : n'$ ² 则告诉我们放大率的信息. 实际上就是分别确定物方主点在焦点后方多远, 像方主点在像方焦点前多远. 而把这两个距离对换, 就变成了对应的节点位置. 换句话说, 如果把两焦点的中点看做对称中心, 则两个主点对这个点中心反演就到了两个节点.

这也就是说, 如果物方像方折射率相等 (这实际上是比较常遇到的情况), 那么也就没有 f 与 f' 的区别, 那么上面列举的四对基点退化为两对. 节点与主点必然重合, 横向放大率与角度放大率都是 1. 而二倍焦距处的那两个共轭点 (物方 $2F$ 点, 像方 $2F'$ 点, 成像系统常称 $4F$ 系统), 物像互为倒立关系, 横向放大率与角度放大率为 -1 .

12.2.3 实例与望远系统

球面折射, 即可以傍轴近似为两个主点重合到光心 (与两节点都不重合), 物像方焦距分别为:

$$f = \frac{nR}{n' - n} ; \quad f' = \frac{n'R}{n' - n}$$

如果 $R \rightarrow \infty, \Phi \rightarrow 0$. 即平面折射, 那么成像系统化为望远系统. 此时物像距为简单线性关系, 横向放大率处处一样:

$$\frac{s'}{s} = -\frac{n'}{n}, \beta = 1$$

横向放大率为 1 是系统沿平面界面平移对称的结果.

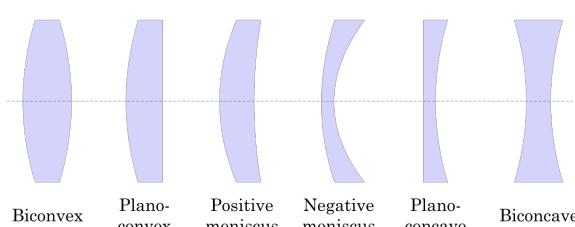


图 12.11: 透镜种类

而对于更复杂而常见的厚透镜系统, 一般我们做如图所示的分类. 即分为**双凸透镜**(biconvex lens), **双凹透镜**(biconcave lens), **平凸透镜**(planoconvex lens), **平凹透镜**(planoconcave lens), **正弯月透镜**(positive meniscus lens) 与**负弯月透镜**(negative meniscus lens). 一般置于折射率为 1 的空气中, 透镜折射率计为 n , 厚度记为 d . 则有以下**磨镜者公式**(lensmaker's equation):

$$\Phi = \frac{1}{f} = (n - 1) \left[\frac{1}{R_1} + \frac{1}{R_2} - \frac{(n - 1)d}{nR_1 R_2} \right]$$

²这个表达式的普遍性参考下一节与下一章.

推导参考之前的证明过程. 注意此处的焦距是要从两个主面开始计算的, 而两个主面的位置一般并不会与左右两折射面与主光轴交点重合. 而是相距:

$$\overline{HH'} = \frac{d(\Delta + f_1 + f'_2)}{\Delta} ; \quad \Delta = d - f'_1 - f_2$$

$$f_1 = \frac{R_1}{n-1} ; \quad f'_1 = \frac{nR_1}{n-1}$$

$$f_2 = \frac{nR_2}{n-1} ; \quad f'_2 = \frac{R_2}{n-1}$$

我们讨论四种特殊的情形. 一是 $R_1 + R_2 = d$ 的情形. 此时恰有:

$$\Delta + f_1 + f'_2 = d - (f'_1 - f_1) - (f_2 - f'_2) = d - R_1 - R_2 = 0$$

即两个主点重合. 其实, 两个主点必然重合到左右球面的公共球心. 这是因为如果入射光线过这个点, 那么出射光线方向恰好不改变, 必然也过这个点, 从而这个点是节点, 而对称成像系统(物像方折射率相等, 焦距相等)的主点与节点重合的. 从而主点也在这里. 只需要确定焦距便可以决定的所有特征. 它由下式给出:

$$f = \frac{nR_1R_2}{(n-1)d}$$

第二种情况是 $\Delta = 0$, 即 $\frac{nR_1}{n-1} + \frac{nR_2}{n-1} = d$. 此时光焦度变为零, 成像系统成为望远系统. 望远系统由于各种量的发散性, 其物像距关系与放大率都必须重新讨论. 如图12.12, 我们把物距 x 仍然计做物到左折射球面左焦点的距离, 而像距 x' 仍计做像到右折射球面右焦点的距离, 有:

$$xx'_1 = f_1f'_1 ; \quad \beta_1 = -\frac{f_1}{x}$$

$$x_2x' = f_2f'_2 ; \quad \beta_2 = -\frac{x'}{f'_2}$$

再考虑到两次成像有 $x'_1 = -x_2$, 得到:

$$\frac{x'}{x} = -\frac{f_2f'_2}{f_1f'_1} = -\frac{R_2^2}{R_1^2}$$

$$\beta = -\frac{f_2}{f'_1} = -\frac{R_2}{R_1}$$

可见物像距间有线性关系, 而横向放大率是一个常数. 我们也定义**纵向放大率**(longitudinal magnification):

$$\Xi = \frac{dx'}{dx} = -\beta^2$$

纵向放大率的大小为横向放大率的平方, 这个结论其实对对称薄透镜也是适用的(请读者证明), 此时物像距关系非线性, 故应采用严格的导数来定义局域放大率.

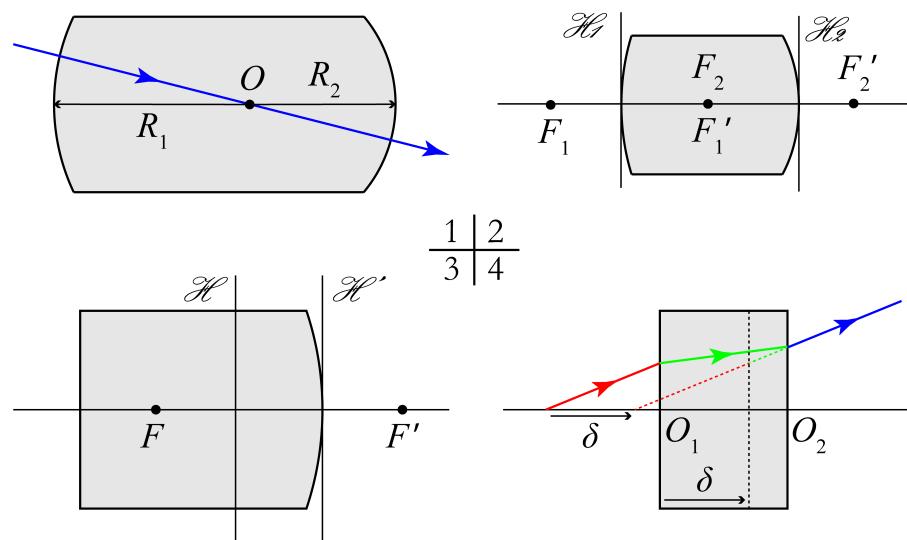


图 12.12: 四种特殊厚透镜

第三种情况是有一个球面退化为平面的情况,此时各公式也因为发散而失效.如 $R_1 \rightarrow \infty$ 成为平面.如图 12.12,我们可以根据此时成像系统的特征确定等效的各基点.像方焦点显然位置仍然在右折射面像方焦点的位置,这是因为平行光经过平面折射仍然是平行光,就应该会聚在这个焦面上.而像方主面也仍然在右折射面顶点处.这可以从平行于主轴的光线的传播确定出来.从而焦距即为右球面的像方焦距 $f = \frac{R_2}{n-1}$.物方主面位于像方主面经左平面折射成像的位置,而物方焦点还在其左侧距离 f 处,这些都很容易理解,留给读者思考其原因.

最后一种情况是两个球面都退化为平面的情况.这是不仅公式发散需要重新计算,它的结果还是一个望远系统.事实上我们可以考虑傍轴的几何光学,用简单的光线追迹法(ray-tracing method),我们发现如果把玻璃砖的厚度进行压缩到 d/n ,那么所有光线都变成了直线传播.从而物就重合到了像.所以我们发现此时的物像映射实际上就是一个固定距离 $\delta = \frac{n-1}{n}d$ 的平移,其横向纵向放大率都是 1.

12.3 更多讨论 *

12.3.1 理想成像本质

理想光具组是一种十分强大的工具,但一些普遍的规律仍然悬而未决:为什么根据光线的对应性能把成像确定到如此小的范围.物像方的焦距为什么总是与折射率正相关,同正同负.而各个放大率间的关系又是如何.所以我们提出理想成像的数学本质抽象:射影变换.

首先是数学上有著名的射影几何基本定理(fundamental theorem of projective geometry):

实线性空间上的共线性变换(collineation)是更高维度内的射影变换(projection).

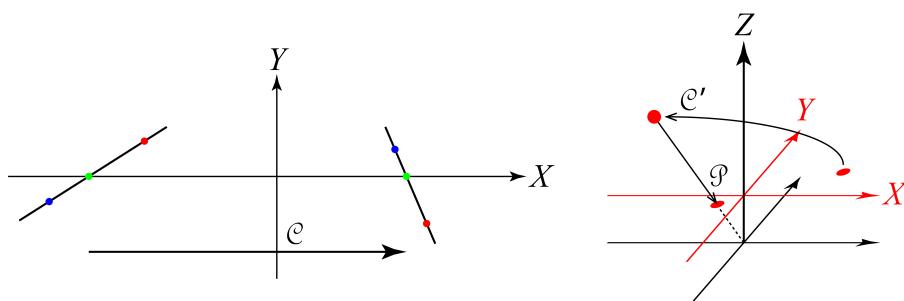


图 12.13: 射影映射

那么成像作为二维平面上的点到点之间的共线性映射:

$$\mathcal{C} : (X, Y)^T \mapsto (X', Y')^T$$

可以加一个第三维度 $Z, (X, Y, Z)$ 可以以原点为中心投影到 $Z = 1$ 平面, 那么原映射理解为三维内的非退化仿射变换加上投影:

$$\mathcal{P} : \{(X, Y, Z)\} \mapsto (X/Z, Y/Z, 1)$$

$$\mathcal{C}' : \{(X, Y, 1)^T\} \mapsto \{(Z'X', Z'Y', Z')^T\}$$

$$\begin{bmatrix} Z'X' \\ Z'Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \quad \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \neq 0$$

此即:

$$X' = \frac{aX + bY + c}{gX + hY + i}; \quad Y' = \frac{dX + eY + f}{gX + hY + i}$$

但为了使得 $Y = 0$ 的 X 轴上的点仍然被保留在 $Y' = 0$ 轴, 且 Y 改变正负号时 X' 不能变, 这给出:

$$b = h = 0; \quad d = f = 0$$

$$X' = \frac{aX + c}{gX + i}; \quad Y' = \frac{eY}{gX + i}$$

为了使变换非退化, 要求 $e \neq 0, gi \neq 0, ac \neq 0$.

情况一是 $g = 0, i \neq 0$, 那么:

$$X' = \frac{a}{i}X + \frac{c}{i}; \quad Y' = \frac{e}{i}Y$$

这代表所有的望远系统, 物像距间为线性关系, 横向放大率在整个空间都是常数. 两个共焦放置的薄透镜, 平面镜与平面的折射系统在傍轴近似下都是这样的系统. 三个放大率分别为:

$$\begin{aligned} \beta &= \frac{Y'}{Y} = \frac{e}{i} \\ \Xi &= \frac{dX'}{dX} = \frac{a}{i} \\ B &= \frac{dY'}{dX'} / \frac{dY}{dX} = \frac{\beta}{\Xi} \quad \text{i.e. } \beta = \Xi B \end{aligned}$$

从推导过程可以发现 $\beta = \Xi B$ 其实是个恒等式.

情况二是 $g \neq 0$, 对应所有的非望远系统. 先考虑纵向放大率:

$$\Xi = \frac{dX'}{dX} = \frac{ai - cg}{(gX + i)^2}$$

我们指出, 沿主光轴 X 方向传播的光线如果在像方仍然沿同样的 X 方向传播, 那么物沿 X 正方向移动时必然有像也沿 X 正方向移动³, 此时 $ai - cg > 0$, 这被称为**折射式系统**(dioptric system). 相反地, 如果沿主光轴方向传播的光线如果在像方沿 $-X$ 方向传播, 此时物若沿 X 正方向移动, 像将沿 $-X$ 方向移动, 被称为**反射式系统**(catoptric system). 此时我们将 X' 变为 $-X'$ 再讨论这个问题, 就会发现对应系数也满足 $ai - cg > 0$ 了. 而更普遍的一类情况, 物像方的主光轴甚至不会在同一个空间中重合, 被称为**折轴式系统**(off-axis system). 此时 X, X' 本质上就位于不同空间. 我们接下来总是默认 $ai - cg > 0$ 即 $\Xi > 0$.

³请读者用费马原理证明它.

进行坐标平移:

$$x' = X' - \frac{a}{g} ; \quad x = -\left(-\frac{i}{g} - X\right)$$

把大写 Y 改为小写 y 便得到:

$$xx' = ff' ; \quad \beta = \frac{y'}{y} = -\frac{f}{x}$$

$$f = \frac{-e}{g} ; \quad f' = \frac{ai - cg}{(-e)g}$$

这便是牛顿式的成像公式. 我们从中可以发现:

$$\frac{f'}{f} = \frac{ai - cg}{e^2} > 0$$

可见两个焦距一定同号. 它与折射率的关系是我们接下来要解决的关键问题.

第十三章 光学仪器

我们暂时不讨论光的干涉与衍射, 故暂时不太需要光的波动理论. 但实际上到了实用的层面上, 各个光学仪器, 从原理上说尽管几乎只用到几何光学的内容, 似乎也是避不开对光的强度的讨论的. 故本章先对在光的波动学说成型之前就已经蓬勃发展起来的光度学(photometry) 进行说明. 然后从若干方面对实用度最高的一些光学仪器做未免以偏概全的介绍.

13.1 光度学

光源产生的光, 在光学系统中可能被光学仪器最终接收, 如光电二极管(photodiode), 电荷耦合元件(charge-coupled device, CCD) 等. 也可以直接被人眼所直接观察. 很明显, 不同仪器作为探测仪器用时, 其光谱响应将存在很大的区别. 然而, 既然是光学, 从古到今仪器的设计都是围绕人的视觉展开. 即最核心的波段是可见光波段. 只是, 随着科技的发展, 眼睛的作用逐渐被物理仪器所替代, 对光的探测也逐渐融入了更大范围的对电磁辐射的探测中, 后者即辐射度量学(radiometry). 比如当今天体物理中用到的各种探测手段, 从波长最长的关于宇宙微波背景的射电探测, 到波长最短的中子星 γ 射线暴的探测, 俨然不再是古人仰观星空可以达到的广度和深度. 辐射度量学以能量作为最基本的物理量进行测量, 单位就是焦耳, J. 但是如果是考虑日常生活中以人眼视觉为核心设计的各类照明, 媒体灯光, 红外线, 紫外线以外的波段几乎就被我们排除在外. 为了将辐射度量学过渡到光度学我们需要先对人眼的独特的色视觉做一个了解:

13.1.1 色度学 *

大部分人类的视网膜内有两种大类的视觉感受细胞¹: 视锥细胞(cone) 与视杆细胞(rod). 而视锥细胞又根据其所含视蛋白质分为感受 500 – 700nm 波长的视红素细胞 (erythrolabe, L, ρ); 感受 450 – 630nm 波长的视绿素细胞 (chlorolabe, M, γ) 和感受 400 – 500nm 波长的视蓝素细胞 (cyanolabe, S, β)². 而视杆细胞由于结构差异导致比视锥细胞敏感约 100 倍, 单光子即可激发其视紫素(rhodopsin, R) 使其产生信号. 视紫素的合成需要依赖维生素 A. 而维生素 A 的摄取除了动物性食物源还可以来自植物性植物中含有的 β -胡萝卜素.

四种感光细胞对不同波长的光响应强度^{13.1}是不一致的. 亮视觉(photopic vision) 才具有真正意义上的色视觉: 此时三种色素细胞 L, M, S 同时工作分管长, 中, 短波段的信号激励. 使得在可见光范围内的不同类似的光对三种色素细胞给出不同的激励值. 影响人对色彩的判断. 而在暗视觉(scotopic vision) 下前三种细胞失效, 视杆细胞积极发挥功能, 其视紫素 R

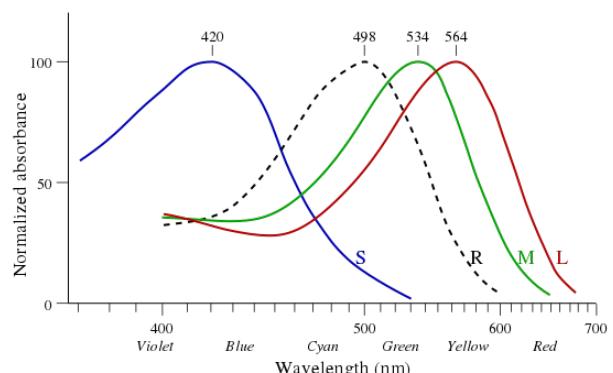


图 13.1: 四种感光细胞约化吸收率

¹还有内在光敏视网膜神经节细胞(intrinsically photosensitive retinal ganglion cells, ipRGCs), 它含有视黑素(melanopsin), 无法造成成像视觉, 但是在调节人昼夜节律方面发挥了主要功能.

²在女性中常见异常 X 染色体导致个体产生四种色素的视锥细胞从而其色视觉明显优于一般的男性或女性

在 $450 - 550\text{nm}$ 波长范围有较强的吸收, 但是无法鉴别光的颜色. 所以暗视觉给人感觉接近灰度图, 同时红色物体在黑暗环境下比其它颜色物体更加昏暗. 这被称作浦金野现象(Purkinje effect). 而有月光或星光的夜晚, 黎明与黄昏, 其实光线并没有暗到足够使得亮视觉失效. 此时的视觉称作中间视觉(mesopic vision).



图 13.2: 浦金野现象

不同人的视觉条件当然存在区别. 粗的层面讲, 一束 555nm 的位于人类视觉最敏感区的绿光和一束可见光边缘的 700nm 的红光, 都给出 1W/m^2 的平面波, 给人造成的亮度感觉一定是不一样的: 前者非常明亮而后者几乎不可见. 但是如果将一红一蓝两聚光灯聚焦于舞台再找两个人来比较哪个光斑更亮, 也许两个人就会给出不同的结果. 这种不同可能出于心理上或者生理上的效应. 但是如果要为照明制定标准, 就应该克服人与人之间的区别, 一般采取从大量受试者判断结果选取多数人的结果的方式. 这样就有了国际照明协会(Commission Internationale de l'éclairage) 在 1931 年通过的 CIE1931 标准. 它目前仍在广泛地被使用.

CIE1931 标准将人眼中三种色素细胞的响应抽象为三激励值(tristimulus values). 并将描述光的色彩, 亮度的空间抽象为色彩空间(color space). 而色彩空间中的点就表示一种混合光可能的色彩与亮度, 对应三激励值 L, M, S , 构成一种加性空间: 三激励值的大小代表抽象的人眼对三种抽象激励的反应. 如果以这三个参数建立坐标系表示色彩, 就称作 LMS 空间. 但是实用场合更常用的是另一套更直观的三激励值 XYZ 形成的空间. 它是这样定义的:

- Y 定义为 L, M, S 的某种组合使得其恰好与混合光视觉亮度基本一致.
- Z 被直接定义为 S 值的某个倍数.
- X 定义为 L, M, S 的另一种组合使得对于所有的单色光 X 总是非负且对于可见光的矩形光谱, $X = Y = Z$.

这样经过调整就得到 CIE1931 标准色匹配函数(color matching functions). 它表示不同单位能流密度的光对人眼造成的 XYZ 三激励值:

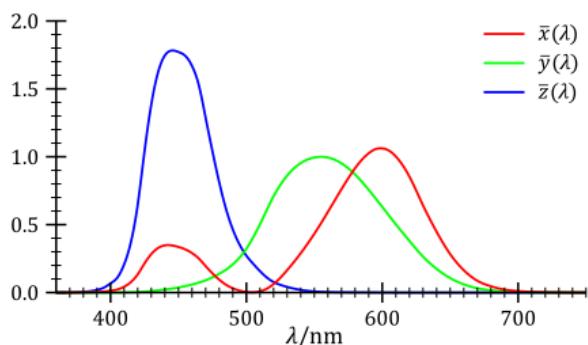


图 13.3: 色匹配函数

直观上来看, 色匹配函数可以根据通过入射光的谱计算三激励值. 例如入射光如果总能量为 E , 那么按照波长分解为:

$$E = \int e(\lambda) d\lambda$$

从而由三激励函数 $\bar{x}, \bar{y}, \bar{z}$ 三激励值就是:

$$X = \int \bar{x}(\lambda) e(\lambda) d\lambda$$

$$Y = \int \bar{y}(\lambda) e(\lambda) d\lambda$$

$$Z = \int \bar{z}(\lambda) e(\lambda) d\lambda$$

从图中定性上可以看出 XYZ 三值的含义为：光越亮 Y 越大，相对 X, Z 值 Y 激励大则说明光中绿光成分多；光中的蓝紫光成分越多 Z 约大；最后 X 值越大主要说明光中的红黄光比例多，但是实际上在 $400 - 500\text{nm}$ 的较短波长的光也能造成一定的 X 激励。

人类对色彩的理解可以分为两个大的部分：亮度(brightness) 与色度(chromaticity)。前者为与光源亮度相匹配的关于光的整体刺激值的指标，与三刺激值的 Y 值直接对应。后者则是独立于亮度的纯粹描述三刺激值混合比例的性质。而色度由于三刺激值比例接近 $1 : 1 : 1$ 时给出特殊的色彩感受：白色。故确定好白点之后又可以细分为色相(hue) 和饱和度(saturation)。饱和度为零对应的颜色即黑、白、灰。随着饱和度增加，较低饱和度的图像一般柔和、粉润；较高饱和度的图像则对比鲜明，大红大绿，视觉冲击力较强。色相则较为复杂，可以参考下面要介绍的色度图。

为了更准确地单独描述色度，我们引入导出的色度参数：

$$x = \frac{X}{X + Y + Z}, \quad y = \frac{Y}{X + Y + Z}, \quad z = \frac{Z}{X + Y + Z}$$

由于 $x + y + z = 1$ ，故通常 $z = 1 - x - y$ 由 x, y 来确定。我们就可以用 x, y 来描述色度，用 Y 来描述亮度。尤其是，把不同 (x, y) 值对应的颜色画在一个图中，就构成了著名的色度图(chromaticity diagram)³。

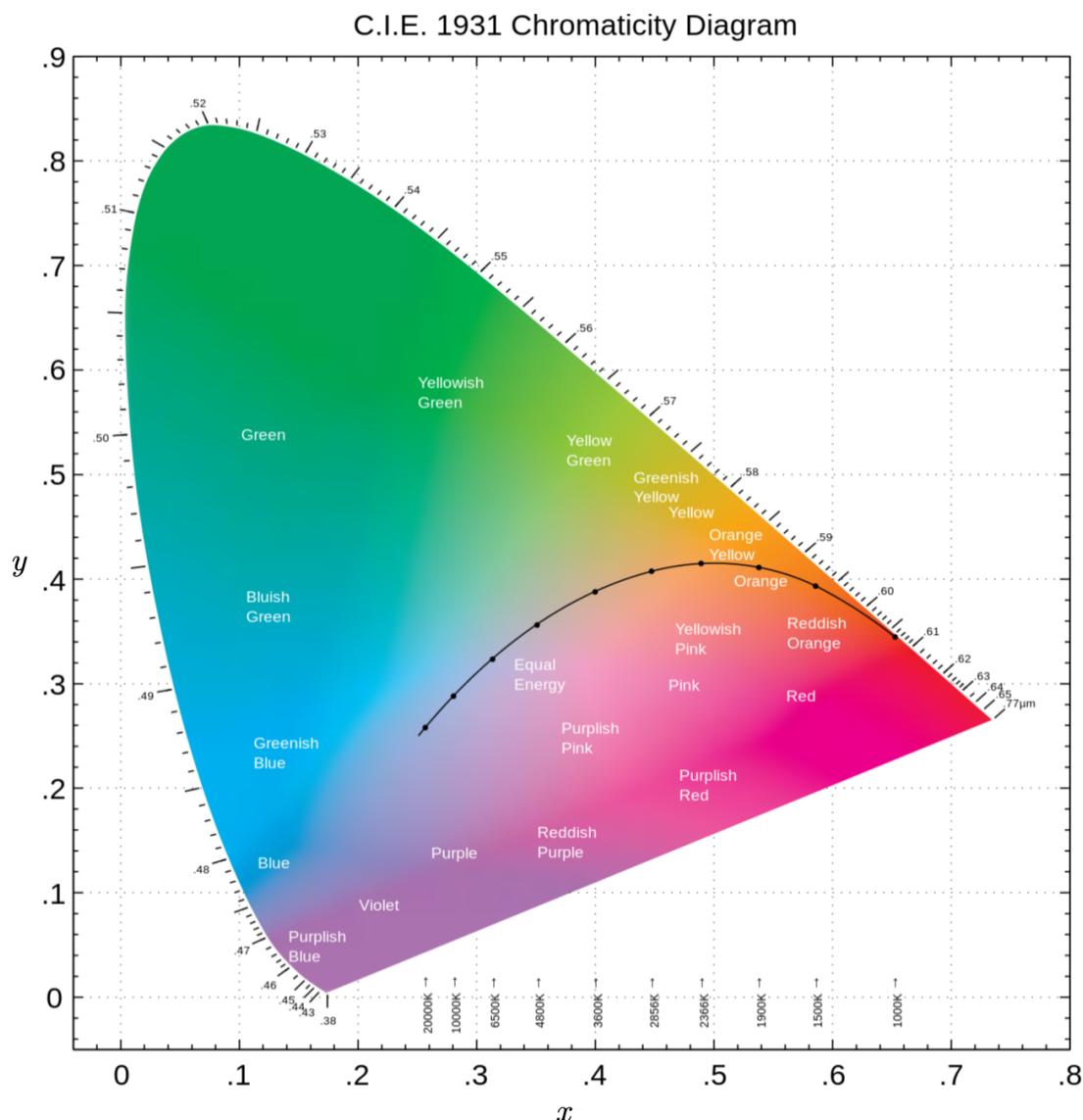


图 13.4: CIE1931 色度图

³由于打印机颜料限制你看到的色度图边缘颜色其实已经失真了。

从色度图上可以读出:

- 并不是色度图上每一个点都代表实际的某种光产生的三激励值. 而是分布在一个舌形的视觉色域(gamut of vision) 中. 例如 $(X, Y, Z) = (0.5, 0.1, 0.4) = (x, y, z)$ 就不是一个存在于自然界的色彩. 除非考虑生理视觉的非线性效应否则无法产生这样的三激励值, 它们对应不可能颜色(impossible color).
- 视觉色域被包含在由光谱色线(spectral locus) 和紫红色线(line of purples) 组成的凸轮廓内. 轮廓上就是饱和度为 1 的色彩, 它包括两类色相: 由所有单色可见光带来的色彩都在这个轮廓的曲边缘上, 即光谱色线, 色相就是我们常说的红, 橙, 黄, 绿, 青, 蓝, 蓝紫; 由可见光边缘的红色与紫色混合形成的色彩都在这个轮廓的直边缘上, 即紫红色线, 色相包括蓝紫, 紫, 红紫, 品红, 红.
- 光谱色线对光波长的分布绝不是均匀的. 小于 470nm 和大于 600nm 的紫, 红光造成的三激励值区别不大, 在色度图中压缩在较短的两端色线内. 黄绿蓝三色在色度图中具有较长的色线.
- 一般定义白点(white point) 为矩形光谱 (单位波长间隔具有相等能量 $e(\lambda) = \text{const.}$) 所对应的点, 按照三激励值定义它为 $(x, y) = (1/3, 1/3)$. 白点的饱和度定义为零.
- 白点与边缘点的连线上所有颜色共享一个色相, 饱和度从从中心的 0 过渡到边缘的 1.
- 混合两种颜色的光时得到的颜色将落在色度图两颜色代表的点的连线上某点代表的颜色上.
- 图中还标出了普朗克色线(Planckian locus). 它们表示不同温度下黑体辐射光谱在可见光波段进行混合导致的色彩. 它历经的颜色为: 红, 橙, 橘黄, 淡黄, 白, 淡蓝.

为了方便显示与印刷行业, 在色度图中选取原色(primary colors) 作为表示颜色的三个基, 进行混合就能产生色度图中一定范围内的色彩. 一般用的多的是RGB 三色模型(trichromatic RGB model). 在 1931CIE 标准中选取的三原色均来自可见光谱的红绿蓝三种单色光, 波长分别为 700nm, 546.1nm, 435.8nm, 定义其 (r, g, b) 值分别为 $(1, 0, 0), (0, 1, 0), (0, 0, 1)$.

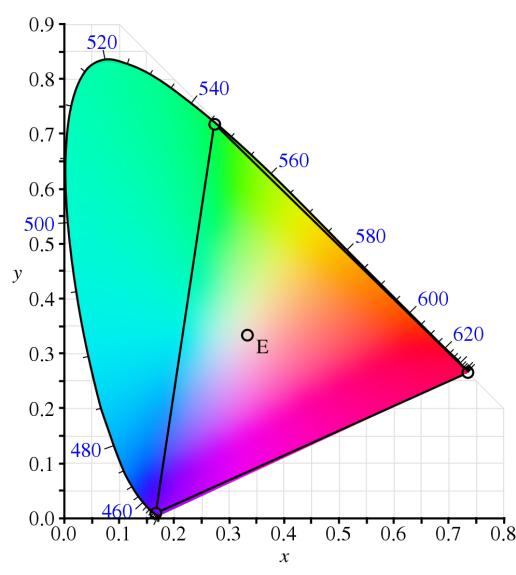


图 13.5: CIE1931-RGB 色域

这实际上对应了一个线性变换:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49000 & 0.31000 & 0.20000 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00000 & 0.01000 & 0.99000 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

并且 (r, g, b) 值定义为:

$$r = \frac{R}{R+G+B}$$

$$g = \frac{G}{R+G+B}$$

$$b = \frac{B}{R+G+B}$$

RGB 模型将三原色进行一定比例的混合后无法覆盖色度图中的每个颜色, 而是只能产生以三原色为顶点的三角形中的色彩. 而不同行业, 乃至同一行业的不同规范使用的三原色标准点也是各不相同的. 从而也就具有不同形状的色域(color gamut). 例如由惠普与微软与 1996 年提出了用于显示器, 打印机与因特网的 sRGB 标准就仅涵盖 40% 的视觉色域面积.

CIE 在 1931 年之后还相继公布了 1964, 1976 标准, 时代在进步, 关于色彩的理论和技术都在慢慢发展与变革. 后世新的标准将会逐渐取代旧标准.

13.1.2 国际单位与光度函数

对于光度学，国际单位制有一基本单位给一个光度学物理量：发光强度(luminous intensity)。这个基本单位为坎德拉(candela)。众所周知，物理量的大小总是与一定的测量过程对应，而其单位总是与测量所采用的标准相对应。那么以下国际单位中有关测量和其标准的规定就定义了发光强度和坎德拉两个概念：

坎德拉(cd) 的定义为：取千克、米和秒的国际单位定义，则频率为 540×10^{12} Hz 的单色光源辐射的发光效率具有固定的数值：683 cd · sr/W。

这其实也就是说，如果一个频率为 540×10^{12} Hz 的单色点光源向四面八方均匀发光，辐射强度为 1W/sr(即具有总功率 $4\pi W$)。则其具有的发光强度就规定为 683cd。这样任意一个同为 540THz，即约 555.016nm 的单色光源的发光强度就可以进行测量：只需要看其辐射强度为之前的辐射强度为 1W/sr 的光源的 x 倍，那么相应的发光强度就是 $x \cdot 683\text{cd}$ 。

但是这就带来一个问题，如何定义一个波长非 555.016nm 的单色光源，或是非单色光源的发光强度？这就又重新依赖于国际单位制以外的标准，依然以 CIE1924 标准为例，它定义了一个光度函数(luminosity function)。它可以想象为利用同为 1W/sr 的不同波长的单色光照明对人造成的相对视觉亮度。实验发现人在 555nm 左右的绿光最敏感，看上去光最亮。故作为标准规定此时光度函数值为 1。其他光度函数值均相对这个标准值比较。其值见下一页表格13.2⁴。

这个函数就给出了波长为 λ 的单色点光源辐射强度 I 与其发光强度 \mathfrak{I} 之间的换算关系，本书用斜体罗马字母表示光度量量而哥特体字母表示原始的辐射度量量：

$$I = 683.002^5 \text{cd} \cdot \text{sr}/\text{W} \cdot V(\lambda) \cdot \mathfrak{I}$$

最后，如果光源发光为连续谱，那么就定义其辐射强度谱函数 $i(\lambda)$ ：

$$\mathfrak{I} = \int i(\lambda) d\lambda$$

于是发光强度就为：

$$I = 683.002 \text{cd} \cdot \text{sr}/\text{W} \cdot \int V(\lambda) i(\lambda) d\lambda$$

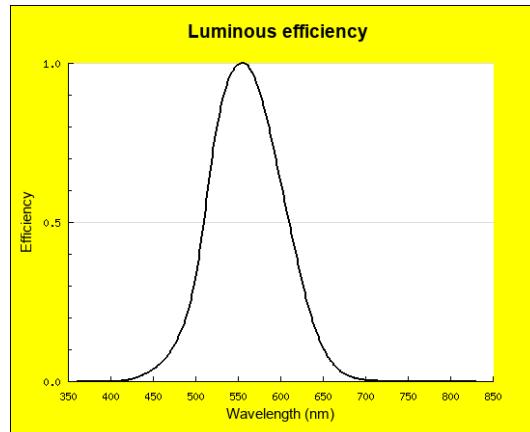


图 13.6: 光度函数图像

最好的方式是把辐射强度谱函数 $i(\lambda)$ 换算为发光强度谱函数 $i(\lambda)$ ：

$$I = \int i(\lambda) d\lambda$$

$$i(\lambda) = 683.002 \text{cd} \cdot \text{sr}/\text{W} \cdot V(\lambda) \cdot i(\lambda)$$

逐渐的我们体会到，任何描述辐射度量的物理量，按照波长进行分解后，乘以固定的系数 683.002 与对应波长的光度函数以后，再对不同波长重新合成，就构成了描述光度量的函数⁶。从而有以下对应关系，下一小节我们讲介绍关键的光度学量：

⁴此为较粗的数据表格，更详细的数据可以参考<http://cvrl.ioo.ucl.ac.uk/cie.htm>

⁵你没看错，由于国际单位制(SI)与国际照明协会(CIE)标准不统一(540THz定标与555nm定标)而是略有差距，导致非555nm的540THz处光度函数小于1而需要靠略大于683的系数来补偿。

⁶这是一种广义的“点乘”或“内积”，你体会到了吗？

表 13.1: 光度学与辐射度量学的对应

| 光度学量 | 国际单位 | 辐射度量学量 | 国际单位 |
|---------------------------|---------------------------|-------------------------|-----------------------|
| 光能量 (luminous energy) | lm · s, (T) | 能量 | J |
| 光功率 (luminous power) | lm | (辐射) 功率 (radiant power) | W |
| 光通量 (luminous flux) | | 辐射通量 (radian flux) | |
| 发光强度 (luminous intensity) | cd | 辐射强度 (radian intensity) | W/sr |
| 亮度 (luminance) | cd/m ² , (nit) | 辐射率 (radiance) | W/sr · m ² |
| 照度 (illuminance) | lm/m ² , lx | 辐照度 (irradiance) | W/m ² |
| 光出射度 (luminous exitance) | | 辐射度 (radiosity) | |

表 13.2: 光度函数表

| λ/nm | $V(\lambda)$ |
|---------------------|--------------|---------------------|--------------|---------------------|--------------|---------------------|--------------|---------------------|--------------|
| 380 | 3.90E-05 | : | : | : | : | : | : | : | : |
| 385 | 6.40E-05 | 455 | 4.80E-02 | 525 | 7.93E-01 | 595 | 6.95E-01 | 665 | 4.46E-02 |
| 390 | 1.20E-04 | 460 | 6.00E-02 | 530 | 8.62E-01 | 600 | 6.31E-01 | 670 | 3.20E-02 |
| 395 | 2.17E-04 | 465 | 7.39E-02 | 535 | 9.15E-01 | 605 | 5.67E-01 | 675 | 2.32E-02 |
| 400 | 3.96E-04 | 470 | 9.10E-02 | 540 | 9.54E-01 | 610 | 5.03E-01 | 680 | 1.70E-02 |
| 405 | 6.40E-04 | 475 | 1.13E-01 | 545 | 9.80E-01 | 615 | 4.41E-01 | 685 | 1.19E-02 |
| 410 | 1.21E-03 | 480 | 1.39E-01 | 550 | 9.95E-01 | 620 | 3.81E-01 | 690 | 8.21E-03 |
| 415 | 2.18E-03 | 485 | 1.69E-01 | 555 | 1 | 625 | 3.21E-01 | 695 | 5.72E-03 |
| 420 | 4.00E-03 | 490 | 2.08E-01 | 560 | 9.95E-01 | 630 | 2.65E-01 | 700 | 4.10E-03 |
| 425 | 7.30E-03 | 495 | 2.59E-01 | 565 | 9.79E-01 | 635 | 2.17E-01 | 705 | 2.93E-03 |
| 430 | 1.16E-02 | 500 | 3.23E-01 | 570 | 9.52E-01 | 640 | 1.75E-01 | 710 | 2.09E-03 |
| 435 | 1.68E-02 | 505 | 4.07E-01 | 575 | 9.15E-01 | 645 | 1.38E-01 | 715 | 1.48E-03 |
| 440 | 2.30E-02 | 510 | 5.03E-01 | 580 | 8.70E-01 | 650 | 1.07E-01 | 720 | 1.05E-03 |
| 445 | 2.98E-02 | 515 | 6.08E-01 | 585 | 8.16E-01 | 655 | 8.16E-02 | 725 | 7.40E-04 |
| 450 | 3.80E-02 | 520 | 7.10E-01 | 590 | 7.57E-01 | 660 | 6.10E-02 | 730 | 5.20E-04 |

13.1.3 亮度与照度

从实用主义的角度来看,首先我们关心一个点光源单位时间发出的能量,或者单位时间通过空间中某个指定的面的光具有的能量,这个量就是功率或辐射通量,两者具有共同的单位瓦特,前者其实就是后者当面闭合且包裹光源时的情形。与之对应的光度学物理量就是光通量(luminous flux, lumens),国际单位就是流明(lumen)。记做 lm。它是一个实用频率非常高的单位,从电灯发光的强度描述,到室内照明的规范标准都可以看到它的身影。对于 555nm 的单色绿光,1W 的辐射通量就相当于约 683.002lm 的光通量。这个常数称为最大光视效能(maximum lumious efficacy)。它的取值为:

$$K_m = 683.002 \text{ lm/W}$$

而一个其他波长的单色光源的光视效能就会降低为 $K(\lambda) = V(\lambda) \cdot K_m$ 。那么如果一个光源的辐射通量谱分布为:

$$\Phi = \int p(\lambda) d\lambda$$

它发出的流明数就可以这样计算得到:

$$\Phi = \int K(\lambda) p(\lambda) d\lambda$$

不同波长的 $K(\lambda)$ 以不同的大小把瓦特换算为流明。最后光源自己的光视效能(lumious efficacy) 就被定义为:

$$K = \frac{\Phi}{\Phi}$$

一支一体式荧光灯(俗称日光灯)的典型数据为 23W, 1500lm。而如果使用发光二极管(LED)灯,相同流明数的白光功率需求可以降低一半多,即达到 200lm/W 左右的能效系数。黑体辐射则最高只有 95lm/W 左右的能效系数,在黑体温度约 6500K 时达到极限。

点光源向不同方向发光的情况不一定是均匀的。为了描述不同方向的发光强度引入单位立体角内的光通量,即发光强度(luminous intensity):

$$I = \frac{d\Phi}{d\Omega}$$

电子行业的发展使得面光源发光成为照明的主要情形,任何电子产品的显示界面就是一个面光源。而我们也关心一个面受到照射的强度情况。从而就产生了照度和亮度两个新的概念。

照度(illuminance) 描述被照射面单位面积从四面八方被照射的总光通量:

$$E = \frac{d\Phi}{dS}$$

亮度(luminance) 则更为复杂,它本应是面光源单位面积朝某特定方向的发光强度。但是若该方向与发光面法向夹角为 θ ,则定义变为:

$$B = \frac{dI}{dS \cos \theta} = \frac{d\Phi}{d\Omega dS \cos \theta}$$

这是因为,如果一个发光面元朝四面八方的单位立体角内发光强度正比于该方向与发光面法向夹角的余弦 $\cos \theta$,这种余弦发光体称为朗伯体(Lambertian body),它就具有特殊的研究价值:比如黑体辐射与漫反射都遵从这种发光性质。从而把这种发光定义为各个方向亮度一致的,注意到上式中 $dS \cos \theta = dS_{\perp}$,故朗伯体的发光面积即发光方向看过去的投影面积。用来算亮度时的面积也用的是这个面积。

13.1.4 光度学守恒量

- 光扩展量(etendue) 守恒:

$$dG = n^2 dS \cos \theta d\Omega$$

- 拉格朗日-亥姆霍兹不变性(Lagrange-Helmholtz invariance):

$$nyu = n'y'u'$$

- 亮度不变性:

$$B = B'$$

13.2 光阑与光瞳

13.2.1 光阑

- **光阑**(stop) 定义: 阻碍光传播的客观对象 (如光圈, 透镜边缘, 镜筒等).
- **孔径光阑**(aperture stop): 起到限制最后像的范围或人能观察到的像的范围的作用. 缩小非孔径光阑的视场光阑能改变像的面积, 不改变像的照度.
- **视场光阑**(field stop): 起到限制系统能通过的光的夹角大小的作用. 缩小非视场光阑的孔径光阑能改变像的照度, 不改变像的面积.

13.2.2 光瞳

- **光瞳**(pupil) 定义: 光阑本身或光阑经过成像系统成的像.
- **入射光瞳**(entrance pupil): 物镜方的光瞳. 最好与物镜孔径匹配.
- **出射光瞳**(exit pupil): 目镜方的光瞳. 最好与人眼瞳孔直径匹配.

13.3 常见仪器简介

13.3.1 人眼

- 人眼明视距离 $s = 25\text{cm}$
- 最小分辨角 $1'$
- 瞳孔直径 $D = 5\text{mm}$
- 水平视场角 160 deg 竖直视场角 130 deg
- 如果一个助视光学仪器是为人眼设计, 那么要考虑三个角度量是否合适: 视角放大率, 最小分辨角与视场角.

13.3.2 显微镜

- 放大镜放大率公式:

$$M = \frac{s}{f}$$

- 光学显微镜放大率公式:

$$M = \frac{s\Delta}{f_O f_E}$$

13.3.3 望远镜

- 分类: 开普勒式, 伽利略式, 格里高利式, 卡塞格林式, 牛顿式...
- 开普勒望远镜放大率公式:

$$M = \frac{f_O}{f_E}$$