

高二物理竞赛 • 电磁学与近代物理

学而思

目录

第一章 序	5
1.1 说在一切前面	5
1.2 本书编排	5
1.3 预备知识	5
1.3.1 力学	5
1.3.2 电磁学	5
1.3.3 热学	5
1.3.4 光学	5
1.3.5 近代物理	5
1.3.6 数学	5
第二章 运动学	7
2.1 时空与物质	7
2.1.1 时空观, 坐标系	7
2.1.2 物质	9
2.1.3 参考系, 物理规律与其不变性	11
2.2 运动的描述	12
2.2.1 质点的运动	12
2.2.2 刚体的运动	19
2.3 参考系变换	21
2.3.1 点变换	21
2.3.2 刚体变换	23
2.4 运动的牵连	23
2.4.1 接触系	23
2.4.2 纯滚系	24
2.4.3 约束系	25
第三章 动力学	27
3.1 牛顿定律	27
3.1.1 概述	27
3.1.2 牛顿第一定律	28
3.1.3 牛顿第二定律	28
3.1.4 牛顿第三定律	29
3.1.5 质点系与它的牛顿定律	30
3.1.6 非惯性系的处理	32
3.2 动量定律	32

3.2.1 质点的动量	32
3.2.2 质点系的动量	33
3.3 角动量定律	34
3.3.1 质点的角动量	34
3.3.2 质点系的角动量	35
3.4 能量定律	36
3.4.1 质点的动能	36
3.4.2 质点系的动能	36
3.4.3 其它能量形式	37
3.5 位力定律 *	40
3.5.1 质点的位力	40
3.5.2 质点系的位力	41
3.6 碰撞问题	42
3.6.1 二质点正碰	42
3.6.2 自由刚体的碰撞	45
3.6.3 多体碰撞	45
第四章 静力学	47
4.1 约束	47
4.1.1 约束分类	47
4.1.2 广义坐标与自由度	50
4.1.3 约束力与广义力	52
4.2 力系化简	55
4.2.1 静力学公理体系	55
4.2.2 力系向一点简化	59
4.3 平衡问题: 矢量力学	61
4.3.1 平衡问题的要素	61
4.3.2 平衡条件与判据	61
4.3.3 亚静定, 静定与超静定	63
4.4 平衡问题: 虚功原理	64
4.4.1 理想约束	64
4.4.2 亚静定问题的虚功原理	65
4.4.3 静定问题的求解	65
4.5 分析力学初步 *	66
4.5.1 用广义坐标表示能量	67
4.5.2 拉格朗日方程	69
4.5.3 再论冲击问题	71
4.6 平衡态稳定性	72
4.6.1 一自由度体系	72
4.6.2 多自由度体系	72
第五章 简谐振动	73
5.1 方程与谐振	73
5.1.1 简谐振动的定义	73
5.1.2 简谐振动的性质	73

5.1.3 简谐振动的判定	74
5.1.4 小振动	75
5.2 阻尼振动与受迫振动	76
5.2.1 阻尼振动	76
5.2.2 受迫振动	78
5.3 多自由度小振动 *	82
5.3.1 基于线性代数与分析力学的简正坐标求解	84
5.3.2 基于对称性的简正模判定与简正频率求解	88
5.4 非线性摄动 *	94
5.5 格波	94
第六章 万有引力 99	
6.1 有心力下质点运动	99
6.1.1 运动的一般特征	99
6.1.2 有心力问题的求解	100
6.2 万有引力下天体运动	103
6.3 二体与潮汐	104
第七章 刚体 105	
7.1 刚体的物理描述	105
7.2 平面平行运动	107
7.3 空间刚体运动 *	108
第八章 弹性体 111	
8.1 弹性体的物理描述	111
8.2 弹性棒, 弹性绳, 弹性膜与弹性体	115
8.2.1 弹性棒	115
8.2.2 弹性绳	116
8.2.3 弹性膜	117
8.2.4 弹性体 *	118
8.3 弹性波	119
8.3.1 分离变量法	119
8.3.2 变量代换法	120
8.3.3 多维情况	120
8.3.4 再论格波	122
第九章 流体 123	
9.1 流体的物理描述	123
9.2 定常流动动力学	126
9.3 黏滞流体动力学	127
9.4 * 流体中的波	128
9.5 波的色散	128
第十章 热力学第一定律 129	
10.1 热力学第零定律	129
10.2 热力学第一定律	130

10.2.1 准静态与非准静态过程	130
10.2.2 内能	131
10.2.3 功	132
10.2.4 热量	132
10.2.5 耗散	133
10.3 理想气体	134
10.3.1 理想气体的定义	134
10.3.2 理想气体物态方程	135
10.3.3 混合理想气体	137
10.3.4 理想气体的过程	138
10.4 开放系统的理想气体	142
10.4.1 静态平衡问题——重力场中的大气	142
10.4.2 能量守恒——伯努利方程	143
10.4.3 动量守恒——欧拉方程	144
第十一章 热力学第二定律	147
11.1 循环过程	147
11.1.1 热机与热泵	147
11.1.2 热机循环	150
11.2 理想气体的熵	152
11.3 热力学第二定律	154
11.4 熵的计算	157
11.4.1 理想气体的熵	157
11.4.2 固定热容固体熵	158
11.4.3 大热容恒温热库	158
11.4.4 传热熵	158
11.4.5 混合熵	158
11.5 热力学函数与其特性	160
11.5.1 四个热力学函数与四个状态参量	160
11.5.2 若干定理的证明	161
11.5.3 自由能的含义	166
11.5.4 化学势	167
11.6 近平衡态热力学 *	168
11.6.1 线性输运现象	168
11.6.2 传热的熵产生	171
11.6.3 普遍理论与昂萨格倒易关系	171
11.6.4 推证热电耦合的普遍规律	172
第十二章 液体与固体的性质	175
12.1 固体晶格论	175
12.1.1 经典晶格论	175
12.1.2 量子晶格论 *	178
12.2 固体电子论 *	183
12.2.1 德鲁特模型	183
12.3 液体的彻体性质	184

12.3.1 液体性质综述与其微观成因	184
12.3.2 热容与黏度	186
12.4 液体的表面性质	186
12.5 极端条件下的其他物态	186
第十三章 相与相变摘要	189
13.1 相平衡	189
13.2 气液相变	189
第十四章 统计物理基础	191
14.1 数学基础	191
14.1.1 概率与独立性	191
14.1.2 随机变量及其数字特征	192
14.1.3 信息熵	193
14.2 统计假设	197
14.3 麦克斯韦分布律	198
14.4 能均分定理	199
14.5 功, 热, 熵	199
第十五章 光波与光线	201
15.1 界面上的反射与折射	201
15.1.1 光波与光线	201
15.1.2 菲涅尔公式	203
15.2 光线方程	207
15.2.1 光线方程与折射定律	207
15.2.2 光力类比	208
15.3 费马原理	209
第十六章 光学成像	211
16.1 傍轴光成像	211
16.1.1 物与像	211
16.1.2 球对称成像系统与符号法则	212
16.1.3 光具组成像	214
16.2 理想成像系统	215
16.2.1 作图法	216
16.2.2 基点基面性质	217
16.2.3 实例与望远系统	218
16.3 更多讨论 *	220
16.3.1 理想成像本质	220
第十七章 光学仪器	223
17.1 光度学	223
17.1.1 色度学 *	223
17.1.2 国际单位与光度函数	227
17.1.3 亮度与照度	229
17.1.4 光度学守恒量	229

17.2 光阑与光瞳	230
17.2.1 光阑	230
17.2.2 光瞳	230
17.3 常见仪器简介	230
17.3.1 人眼	230
17.3.2 显微镜	230
17.3.3 望远镜	231
第十八章 静电学	233
18.1 电荷与电场	233
18.1.1 电磁相互作用与电荷	233
18.1.2 库仑定律	234
18.1.3 电场	234
18.2 两个定律与电势	237
18.2.1 电场的高斯定律	238
18.2.2 电势与电场的环路定律	239
18.2.3 总结	240
18.3 静电能	241
18.3.1 静电势能	241
18.3.2 自能与相互作用能	242
18.3.3 电场能	244
18.4 电荷体系	245
18.4.1 电偶极子	245
18.4.2 电荷密度	248
18.4.3 极化强度 *	249
18.4.4 若干对称带电体系的电场	251
第十九章 导体与介质	255
19.1 导体与静电平衡	255
19.1.1 绝缘体与导体	255
19.1.2 导体的特点	256
19.1.3 常见简单体系	260
19.2 电像法	262
19.2.1 半无限大空间的电像法	263
19.2.2 球面外与球面内的电像法	264
19.2.3 立体谐函数 *	266
19.3 电介质	270
19.3.1 微观角度理解极化	270
19.3.2 宏观角度理解极化	273
19.3.3 微观与宏观的联系	277
19.4 再议静电能	278
第二十章 稳恒电流	281
20.1 稳恒电流描述与形成	281
20.1.1 德鲁特模型	281

20.1.2 费米气观点 *	282
20.1.3 能带论 *	284
20.1.4 惯性, 阻尼与回复力	286
20.1.5 稳恒电流与形成条件	287
20.2 电路与电路方程	289
20.3 电路分析基础	292
20.3.1 电路的整体结构与拓扑学结论	292
20.3.2 电路问题的求解套路	295
20.4 电路分析方法	297
第二十一章 静磁学	299
21.1 电流与磁场	299
21.1.1 磁场地位与电流分布	299
21.1.2 毕奥-萨伐尔定律	301
21.2 两个定律与矢势	303
21.2.1 磁场的环路定律	303
21.2.2 矢势与磁场的高斯定律	305
21.3 电流体系	307
21.3.1 磁偶极子	308
21.3.2 磁化强度	311
21.3.3 若干对称体系的磁场	311
21.4 磁介质与磁能	312
21.4.1 微观角度理解磁化	312
21.4.2 宏观角度理解磁化	313
21.4.3 磁场能量	314
第二十二章 磁生电	315
22.1 电磁感应	315
22.1.1 动生电动势	316
22.1.2 感生电动势	317
22.2 自感与互感	320
第二十三章 麦克斯韦方程组	321
23.1 麦克斯韦方程组	321
23.1.1 电生磁	321
23.1.2 麦克斯韦方程组	323
23.1.3 电荷在电磁场中的运动	327
23.2 平面电磁波	328
23.2.1 真空中的电磁波	328
23.2.2 介质中电磁波的传播	329
23.3 电磁场能量与动量	331
23.4 电磁波辐射	331
23.4.1 电磁辐射概论	331
23.4.2 偶极辐射	332

第二十四章 交流电路	335
24.1 相量表示	335
24.1.1 拟稳条件与交流元件	335
24.1.2 电阻, 电容, 电感特性	335
24.2 常见电路	337
24.2.1 谐振电路	337
第二十五章 相对论力学	339
25.1 相对论运动学	339
25.2 相对论动力学	340
25.3 相对论连续物质	342
第二十六章 光的干涉	345
26.1 标量波理论	345
26.2 分波面干涉	349
26.3 分振幅干涉	350
26.4 偏振干涉	351
26.5 相干性	353
第二十七章 光的衍射	355
27.1 光栅与波带片	355
27.2 衍射积分公式	356
第二十八章 物理光学	357
28.1 经典型色散理论	357
28.1.1 复波矢与复折射率	357
28.1.2 经典电子论的解释	360
28.1.3 * 辐射阻尼与散射截面	363
第二十九章 量子论	367
29.1 黑体辐射	367
29.2 光粒子性	367
29.3 玻尔原子	367
29.4 物质波与波函数	368
第三十章 物理学尺度	369
30.1 宇观	369
30.2 宏观	369
30.3 介观	369
30.4 微观原子	369
30.5 微观亚原子	369

第一章 序

1.1 说在一切前面

There are clues everywhere,
all around us.

The wrong interpretation of the clues,
we call our world.

— David Lynch

现代人的生活,变换莫测,日新月异.不同人用着不同的方式去理解变化背后体现出来的规律.优秀 的理解方式互相交流与学习,学科就从中诞生.物理学(physics)就是其中的一门学科,它是自然科学(nature science)的主要构成成分之一.今天也许很难找到依据为物理学和自然科学的范围明确划分界限,但是物理学秉承了自然科学的一些重要方法论特征,是可以帮助我们明显的与其他学科做区分的,它们包括:

- 经验主义.把物理学和文学,艺术,乃至亲缘关系近的数学比对,就容易发现后者们的核心在于创新而非做回顾.而物理很多时候,etc
- 实验物理.
- 可证伪性.
- 与数学水乳交融:描述与诠释.
- 有效预言.

1.2 本书编排

1.3 预备知识

1.3.1 力学

1.3.2 电磁学

1.3.3 热学

1.3.4 光学

1.3.5 近代物理

1.3.6 数学

第二章 运动学

2.1 时空与物质

物理学,从刚开始成为实验性的科学的伽利略时期开始,到近半个世纪年来理论物理学家对额外维度的探讨,都给予了时空(spacetime)最核心的地位.牛顿的理论,分析力学,经典场论,相对论这些理论最基本的图像都是时空与物质(matter)的分立性.时空是装备了一个能体现出物理物理意义的度量(metric)的3+1维对象¹.而物质是在每个时空点处的某种结构.下面会分别介绍不同物质体系的描述方法.在这之前我们来看看不同的时空观:

2.1.1 时空观,坐标系

牛顿力学理论体系基于伽利略时空观(Galilean spacetime picture),也就是绝对时空观(absolute spacetime picture).在这里空间是三维的平直空间.设想一个人站在该空间的某个空间点,他的胸前,头顶和右手平举的三个方向就是相互垂直的方向.如果以自己的臂长为标准长度,他能够定义空间中每两个点之间的空间间隔.事实上,这个人能够以一种正确的方式为每个空间点定义一个坐标,那么所有三维空间点的集合与任意两个点之间的距离为:

$$A(x, y, z) \in \mathbb{R}^3 ; \quad x, y, z \in \mathbb{R}$$

$$A(x_1, y_1, z_1), B(x_2, y_2, z_2) : l^2 = \overline{AB}^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$$

因为空间不同于时间,我们为以上数字赋予特殊的物理含义,也就是加上量纲(dimension).如果两点坐标为(0,0,0)和(1,1,1),那么 $l = \sqrt{3}\text{m}$,而不是 $l = \sqrt{3}$.其单位m一方面表示了这个物理量的属性,另一方面指定了某个实际物理体系确定下来的固有长度大小.现行的(2018年1月1日,下文同)国际单位制对1m的定义如下²:

1m是光在1/299792458 s内在真空中行进的距离.

这就是我们的三维平直空间(3-dimensional flat space).注意空间点具有物理实际意义,它可以脱离坐标系而单独存在.事实上坐标系的原点可以建立在空间中的任意点处,朝向也可以是任意方向,两个空间点之间的距离 l 不会依赖于坐标系的选取,但两个点的坐标会因坐标系不同而改变.如果我们选取的坐标系总是下述使得两个相距很近的点之间的微元距离公式成立:

$$dl^2 = dx^2 + dy^2 + dz^2$$

那么建立的坐标系就是一个笛卡尔坐标系(Cartesian coordinate system),即空间直角坐标系(3D orthogonal coordinate system).但是以空间直角坐标系为基础,我们又经常建立球坐标(spherical coordinate)和柱坐标(cylindrical coordinate)系统.通常取 x 轴为幅轴(azimuth axis),点在 $x-y$ 平面上的投影与原点的连线相对 x 轴转过的角度为 φ ,即幅角(azimuth angle).而 z 轴为极轴(polar axis),而点与原点连线与极轴的夹角 θ 为极角(polar angle).天文观测用球坐标就很方便,它是以描述的空间点到原点之间的距离 r ,也

¹数学上有一套严格的说法,把这种连续,光滑的四维对象称为伪黎曼流形(pseudo-Riemannian manifold).

²一方面,它依赖于狭义相对论的正确性,目前极少理论物理工作者会质疑它.另一方面,应该要先定义下文的1s,再来定义1m.

称矢径(radius), 和两个描述角位置的极角幅角来构成三个坐标 (r, θ, φ) 的. 而理论物理里也常用到的柱坐标是以 (ρ, φ, z) 为描述空间点的坐标, ρ 是空间点到 z 轴的距离.

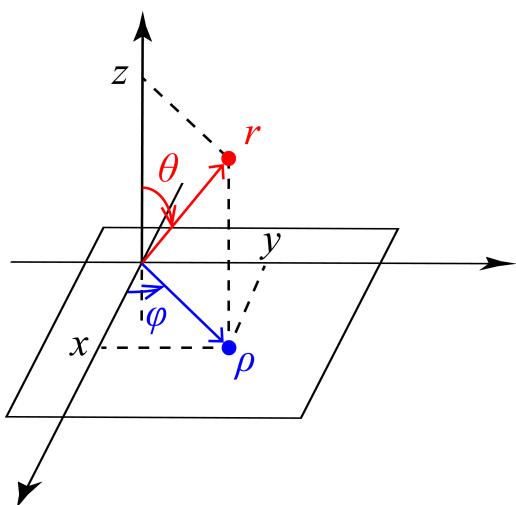


图 2.1: 三种坐标

需要注意, 我们对以上 θ, φ 角度定义比较微妙, θ 一般限制其在 $[0, \pi]$ 范围内, 在端点 $\theta = 0, \pi$ 处 φ 的不同取值代表同一个空间点. 但 φ 的取值我们不加以任何限制, 也就是原则上 $\varphi \in \mathbb{R}$. 而任意 φ 相差 $2n\pi$ 的坐标实际上代表同一个空间点. 这么做是因为我们最好不要用静态的几何观点去看待角度这一概念. 而是用运动, 变换的观点去看待角度. 作为连续的运动的点的幅角变化也应是连续的, 为了使点绕 z 轴一圈后幅角不至于突然改变, 就必须默认每一点的幅角可以有多种取值, 在具体的运动中应该灵活选择其具体取值大小.

之后经常会说到各种对称性, 在本系列教材中我们采取如下说法: **球对称**(spherical symmetric) 仅仅代表某个函数 $f(r, \theta, \varphi)$ 与 φ 无关. 而**柱对称**(cylindrical symmetric) 代表的是 $f(\rho, \varphi, z)$ 与 z 无关. 与 φ 和 θ 都无关的

$f(r, \theta, \varphi) = f(r, \forall\theta, \forall\varphi)$ 被称为**各向同性**(isotropic). 最后**中心对称**(centrosymmetric) 是一个很弱的对称性, 它仅仅代表函数在**中心反演**(space inversion) 下的对称性:

$$f(x, y, z) = f(-x, -y, -z)$$

绝对时空观中的时间则是一种完全与空间独立的属性. 实际上, 每一个空间点处的人都能感受到时间的流逝. 十分抽象的把这些不同的时刻画在一根时间轴上, 便是一维的时间坐标:

$$A(t) \in \mathbb{R} ; t \in \mathbb{R}$$

而时空又是两个独立的概念. 任意一个坐标点处都有时间轴, 而任意一个时刻都有一个三维空间切片. 事实上, 我们的时空是一个 3+1 维的结构, 合理地选取坐标后, 实际上可以把时空结构写成四维坐标:

$$A(x, y, z, t) \in \mathbb{R}^4 ; x, y, z, t \in \mathbb{R}$$

而对于任意两个时空点, 存在绝对的时间, 也就是可以找到两个事件的绝对时间差:

$$\tau = |t_2 - t_1|$$

但空间距离却具有相对性. 我们要求在同一时刻的同一空间切片上定义空间的度量:

$$t_1 = t_2 : l = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

不同时刻的空间, 被线性地组织在一起, 保证每一刻都平直的空间既不会随时间膨胀收缩, 也不会旋转加速. 这种时空结构被叫做**牛顿-嘉当几何**(Newton-Cartan geometry)³.

时间是一个新的量纲, 其国际单位制对 1s 的定义为:

1s 是铯-133 原子基态的两个超精细能级之间跃迁所对应的辐射周期时长的 9192631770 倍.

牛顿-嘉当几何的显著特点是物理量的定义没有所谓的**协变性**(covariance). 固然, 在三维空间坐标框架的旋转下, 任何物理过程中的事件所发生的位置坐标发生旋转变换, 而几乎所有标量(比如质量, 能量)都不

³用更为严格的语言来说, 我们需要一个装备了一阶形式“钟” c_ν 和二阶逆半正定对称张量类空间度规 $s^{\mu\nu}$ 的四维微分流形 \mathcal{M} :

$$\mathcal{M} : s^{\mu\nu} c_\nu = 0$$

变,几乎所有矢量(比如动量,角动量)都按照类似坐标变换的法则发生变换.但是,对于更普遍的一列伽利略变换(Galilean transformation),也就是相对匀速运动的参考系之间的变换,时空坐标的形式却不能简单地推广到能量动量角动量上.

不同于绝对时空观中时间与空间成为相互独立的量纲的特点,狭义相对论改变了对基本物理量的看法.狭义相对论把时空看成为可以相互转化的不可分割的新的3+1维整体,同时性已然破缺.两个时空点之间无法定义绝对的时间间隔.在相对论时空观(relativistic spacetime picture)下,时间和空间可以用同一把尺子去丈量:这是由于光速的不变性:

$$c = 299792458 \text{ m/s}$$

而量出来的长度叫做时空间隔(spacetime interval):

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2$$

这赋予时空以截然不同的结构,最关键的一点是,时间的绝对性被取消了,不同事件发生时间先后的比较不总是可行,相对论的有限速度因果律在这里取代了经典观点的时序因果律.我们将在电磁学理论的最后介绍相对论理论.

2.1.2 物质

时空是物理过程发生的舞台.那么将物质引入后舞台上所演出的便是事件(event).事件这个物理概念由这样的数学工具来描述:它的第一个要素自然是时空坐标(r, t),是事件在时空上的外化.第二个要素是事件内禀的属性.它取决于所引入的物质种类,还取决于我们所关心问题的层次.一般用标量,矢量,乃至张量这样的数学工具来描述它.举例,电磁场物质在经典物理中用电场磁场来描述,但在量子力学(quantum mechanics, QM)中这不够了,需要用矢势和标势来描述才是完整的.在更深的量子场论(quantum field theory, QFT)中甚至这也是不够的.还需要量子化为光子才合适.又比如,电子这个很有意思的现象,最简单的模型是质点模型,内禀的属性是质量,动量与能量⁴.然而与电磁场的经典相互作用强度告诉我们还有一项内禀属性叫做电荷量.近代人们惊奇地发现原来电子还固有磁矩,与之相应的电子具有自旋.最后狄拉克等人发展出量子电动力学(quantum electrodynamics, QED),统一地用一个四分量的旋量波函数和它的方程中的若干参数来完整地描述所有发现的电子内禀属性.

经典物理学⁵(classical physics)中涉及到的物质主要有:

1. 质点 (point mass, particle)

不得不承认质点是牛顿力学的根基,一切可观的结论的出发点.质点所对应的事件集合为时空中的一条世界线(world line).每一个时间仅仅有可能只有一个事件发生.实际上质点的运动用运动学方程(kinematic equation)来描述:

$$\mathbf{r} = \mathbf{r}(t) = (x(t), y(t), z(t))$$

而质量(mass)是质点必要的内禀属性.它将作为参数出现在下一节介绍的动力学方程中.它反应物质受到同样大小的相互作用运动状态改变的难易程度.国际单位制从1889年至2018年末对1kg的定义如下:

1kg是保存在法国巴黎布勒特伊宫的国际计量局实验室的约47立方厘米立式铂铱合金小圆柱的质量.当然,出于实用考虑,也是很多它的官方复制体的质量.

这个定义今已经有一百多年了,历史远长于其他六个国际基本单位.但目前这一古老的定义方式已经被废除,新的定义⁶为:

⁴不要认为动量能量不是内禀的而是由时空运动所决定的,读者可以思考在水中一个气泡的上升,它为体系带来的动量方向如何?

⁵本书的经典物理学指牛顿力学,分析力学,经典电磁学,经典统计力学与狭义相对论的范畴.

⁶一方面,它依赖于狭义相对论和量子力学的正确性,目前极少理论物理工作者会质疑它.另一方面,应该要先定义上文的1s和1m,再来定义1kg.

1kg 被这样定义: 取普朗克常数的固定数值在单位 $\text{kg} \cdot \text{m}^2 \cdot \text{s}^{-1}$ 下为 $6.62607015 \times 10^{-34}$.

长度, 时间和质量为三大力学量纲, 量纲是物理量的属性, 物理量的表示方法为数值加单位, 每个量纲都有自己独特的一套单位, 不同单位间差一个纯数的倍率. 在物理量的加减时量纲必须相同而且结果保持量纲不变. 但不同量纲物理量可以进行乘除而生成新的量纲. 除了简单的加减乘除的以上规则以外, 其他特殊函数必须只能作用在无量纲的纯数上. 这叫做**量纲法则**(dimensional rules).

2. 质点系 (particle system)

质点系是对质点的一种自然扩展. 讨论质点的运动时质点受到的相互作用来自何方? 在牛顿力学的框架下必然来自施力物体. 故把受力物体和施力物体都作为简单的质点时, 两个质点就可以作为整个体系讨论, 把相互作用区分为内力与外力而推广之前的结论, 在这个过程中我们能体会到哪些结论是一脉相承的普遍规律而哪些需要重新审视与修改. 从原理上质点系不过是多个质点同时存在的情况而已. 而**经典统计力学**(classical statistic mechanics) 是这种观点的提炼与延伸, 它着眼于一个宏观大分子数的体系的长时间平均下的行为, 理论上常常结合概率论的做法, 对等概率的体系代表点构成的系统做平均. 从中提取统计力学下不平凡的物理量.

3. 连续介质 (continuous media)

一方面, 连续介质是从纯粹的理论过渡到实际问题的至关重要一步. 另一方面, 在处理连续介质的问题中形成了场论, 它作为一种与质点截然不同的物理图像形成了自己的一整套理论. 更重要的, 量子场论认为由于不确定原理, 质点与世界线的观点实际上是作为场的传播的一种近似. 也就是说, 场论相比质点的力学更具有兼容性和普适性. 对于简单的连续介质, 内禀的属性由每一个点处的质量密度 ρ 和速度 v 描述, 这实际上形成了一个标量场和矢量场:

$$\rho = \rho(\mathbf{r}, t) ; \quad \mathbf{v} = \mathbf{v}(\mathbf{r}, t)$$

而动力学方程的列法又强烈依赖于介质内的相互作用. 于是这个大的问题又分出很有代表性的弹性力学和流体力学, 还有介于两者之间的粘弹性塑性模型等等. 还有从热力学平衡态出发考虑的近平衡态统计力学方法.

4. 场 (field)

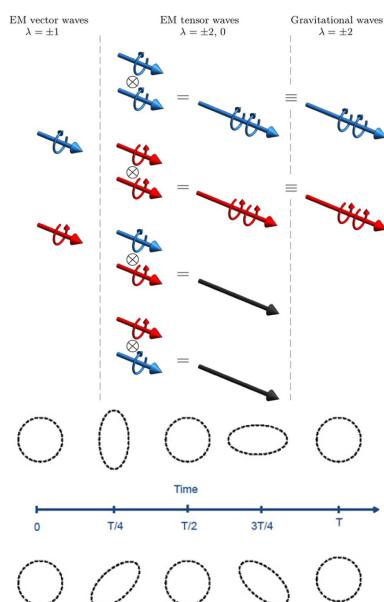


图 2.2: 引力波与电磁波的区别

关于场的理解历史上经历了两个过程. 第一个阶段是正如上面的连续介质的做法, 场是作为某种事先已经被研究的足够清楚的粒子体系与它们之间的相互作用的描述方法与数学工具. 而从牛顿时代开始人们就默认了超距作用的存在, 而场数学工具的第二个作用就是描述电荷之间, 质量之间的超距作用. 但是逐渐人们意识到相互作用传播速度的有限性, 于是场开始变成某种与物质其他耦合的物质而单独存在, 这就是第二阶段, 人们意识到了场本身就是一种物质, 电子与电子之间不能直接相互作用, 电子改变了真空, 稳定地与电磁场相互作用而产生库仑场, 然后在对第二个电子产生相互作用力. 将场作为某种先验地存在于时空中的物质形式进行描述与研究的学科就叫做**场论**(field theory). 有哪些可能的场形式? 场的物理量如何书写? 场与粒子, 场与场之间如何相互作用便是场论的研究内容. 比如说, 动力学所关心的内容, 场的质量本身没有定义, 可以定义的一般是场的定域动量与波矢, 能量与角频率, 平面波的色散关系等等. 的确, 场论最基本的做法是傅里叶分解, 它认为平面波是不受任何相互作用时的场的惯性运

动方式,而任何场的运动,每一时刻都可以分解成一系列平面波的叠加.就好像任何复杂体系的运动本质上都是质点系的运动,而质点的运动每时每刻都可以看出匀速运动那样.

场论根据其描述方法分为标量场论,矢量场论,张量场论和旋量场论等.电磁场论是矢量场论,它用四势矢量 (φ, \mathbf{A}) 来描述.引力场论在广义相对论弱场近似下是张量场论,但传播的引力波和电磁波类似只能有两种偏振,其模式为 + 型和 × 型,这又是不同于电磁波的.在更现代的观点中,场是真空的某种激发,十分类似于一张二维的拉紧的薄膜,没有任何振动时它已经固有能量了,场物质反应为膜的各种振动形式.而一个石子受到外力而压弯了膜,就可以理解为粒子与场之间的相互作用.

2.1.3 参考系,物理规律与其不变性

坐标系(coordinate system) 是一套用来描述时空点位置的坐标系统,而**参考系**(reference system) 则是下文将引申的一类坐标系的统称.

在伽利略时空观⁷下参考系的定义相对简单.我们只需要任意指定一个观察者,观察者可以在 $3+1$ 维的时空中作任意运动,每一个时刻对应一个测量的坐标系统:观察者分别以右手边,胸前和头顶三个方向建立坐标系,坐标系的刻度就是其代表的位置到原点观察者处的空间距离.这样便可以表示任意事件的时空坐标.而且构成一个三维空间直角坐标系.叫做观察者的**固有参考系**(proper reference system).

物理规律的本质是什么?在一个特定的参考系中,我们进行一些物理量的测量,测量得到的值之间在特定的实验条件下会满足特定的关系而不依赖于某些具体的参数,一般都是等式.这样的普遍成立的等式就可以反映某些物理规律.从**动理学**(kinetics) 的角度理解,体系的演化是**状态**(state) 随时间的演化,而体系的状态由一组完备的**态参量**(state variable) 描述,态参量可以包含一些广义坐标 q_α 和他们对时间的导数即广义速度 $v_\alpha = \dot{q}_\alpha$.而以下**决定论**(determinism) 则对物理规律的形式做出了规定:

动理学系统:

已知某时刻 $t=0$ 的初态 $\{q_\alpha(0), \dot{q}_\alpha(0)\}$ 后,应能预言任意将来的状态 $\{q_\alpha(t), \dot{q}_\alpha(t)\}$.

事实上,不光是将来的状态,在一般经典物理的框架下,过去的状态也是可以被“预言”的.这是因为我们进一步要求动理学系统的规律是以下形式的二阶方程:

$$f_i(\ddot{q}_\alpha, \dot{q}_\alpha, q_\alpha, t) = 0$$

而且广义坐标的个数应该和方程的数目一致,方程是至多二阶的.微分方程的描述在数学上完全导致了体系彻底的决定性,只要知道初始条件便可以往前往后推理所有发生的物理现象⁸.

这一物理学科研究范围内的有关决定性的原理很强大,强大到历史上很多物理学家和哲学家都质疑过它的真实性和适用性.最著名的莫过于拉普拉斯提出的**拉普拉斯妖**(Laplace's demon) 的论点:如果存在一个全知的拉普拉斯妖,能够知道在某一时刻宇宙中所有原子的精确位置与动量,那么任意时刻的状态也能被预言.这种决定论观点总是被拿来与**自由意志论**(free will) 去归谬,如果过去与未来都已经被谱写,那么人类所有的行为的意义和动机也就不能被解释.这直到今天也是科学与哲学未能获得满意的回答的问题.

但是科学的一个目的就是在于能够对未发生的事情进行预言,所以本着对决定论的认可与尊重,我们才能继续我们的科学的研究.

除了需要符合决定性原理,物理规律还需要反映出一定的对称性.整个牛顿的**经典力学**(classical mechanics) 体系都要遵从的一个大的对称性,我们可以称之为**伽利略对称性**(Galilean symmetry),也是伽利略时空观的一个重要组成部分.这个对称性的含义如下阐释:

⁷本章仅限讨论非相对论,相对论见后

⁸更有甚者,经典物理也难以解释耗散现象,在忽略耗散时,方程应该还具有平移对称性(不显含 t) 和时间反演对称性(方程对广义速度是偶函数).此时对将来和对过去的预言在广义速度取相反数时是完全对称的,而且预言只依赖于时间差不依赖于初始时刻.

首先我们重拾“绝对时空观”的说法. 其含义是存在一个特殊的观察者, 它建立的坐标系 $O - xyz$ (对位于原点的观察者来说, 是参考系) 是默认的第一个使得我们以上提出的物理规律成立的坐标系. 自身具有 $3 + 1$ 维的结构:

$$(x, y, z, t) \in \mathbb{R}^3 \times \mathbb{R}, \quad \mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$$

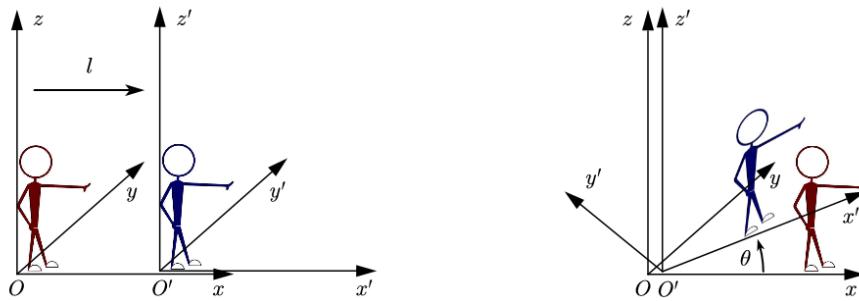


图 2.3: 固定平移与转动

让我们来阐明参考系与坐标系之间的区别. 我们设想另外一个观察者, 它与原观察者保持相对静止. 但是其建立的坐标系 $O' - x'y'z'$ 依然可能有很多种可能性: 默认观察者自己为原点, 右手边为 x 轴, 胸前为 y 轴, 头顶为 z 轴. 一种可能性是, 新观察者相对原来的观察者仅仅是做时空平移(translation), 而其轴的方向保持平行.

2.2 运动的描述

2.2.1 质点的运动

质点的运动是复杂的连续体的运动的基础. 所谓运动(motion) 就是每一个时间都有一个位置. 就是说位置是时间的函数:

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

我们使用实数上的矢量空间和矢量来刻画真实的物理学上的二维或三维空间和其中的点. 也即:

$$\mathbf{r} \in \mathbb{R}^2 \text{ or } \mathbb{R}^3$$

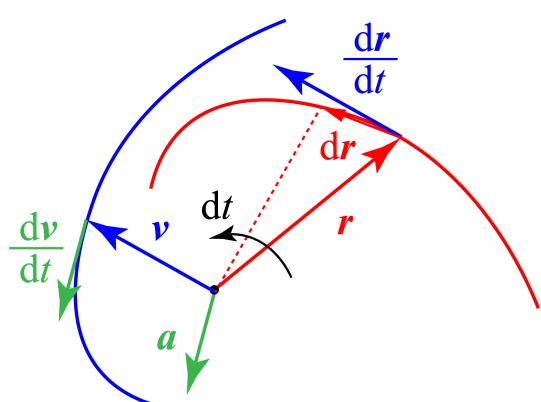


图 2.4: 质点的位置, 速度与加速度

前者对应二维运动, 后者对应三维运动. 我们从中可以发现, 函数是描述运动的数学工具, 对矢量与函数的分析学知识便是运动学和物理学的基础. 而运动学本身是可以看成是数学的一个分支的. 对质点的位置矢量求导数便得到速度矢量, 再求导数或是求二阶导数便是加速度矢量:

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t}$$

$$\mathbf{a} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t}$$

$$\mathbf{a} = \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \lim_{\Delta t \rightarrow 0} \frac{[\mathbf{r}(t + \Delta t) - \mathbf{r}(t)] - [\mathbf{r}(t) - \mathbf{r}(t - \Delta t)]}{\Delta t^2}$$

正如上图所示, 我们可以把每一个瞬时的速度矢量共起点画出来, 并把端点连做曲线, 此时背景的空间往往叫做速度空间(velocity space), 它的更普遍的对应是动量空间, 这在近代物理的理论值具有重要地位. 引入速度空间后质点的运动便可以等价地在速度空间中进行研究. 比如在均匀重力场中的抛体运动在速度空间中

就是简单的竖直向下的匀速直线运动. 而平方反比力下的天体运动(开普勒问题)实际上在速度空间中可以证明将变为变速偏心圆周运动. 这使得通过速度求加速度就好像通过位置求速度那样方便. 具体求法接下来还得进一步展开.

加速度的导数偶尔也会进入物理研究的视野, 它被叫做**急动度**(jerk). 即:

$$\mathbf{j} = \frac{d\mathbf{a}}{dt} = \frac{d^3\mathbf{r}}{dt^3}$$

矢量函数概念就是实数到矢量的映射, 但是具体能写出函数形式的例子在物理中却不是那么常见. 比如抛体运动的运动对应的函数, 即**运动方程**(equation of motion)⁹为:

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}_0 t + \frac{1}{2} \mathbf{g} t^2$$

它是牛顿定律下以下微分方程结论的解:

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2} = \mathbf{g}$$

但是在一般情况下, 要表示出一个既有大小又有方向的矢量函数, 我们得采用分解的思想, 研究能够完全确定这个矢量值的参数. 通常有以下做法:

1. 直角坐标系下的分解

在一个欧几里得空间里的一个参考系中研究问题, 很常见的一种做法是建立直角坐标系来表示一个点的位置. 以更普遍的三维运动为例. 既然物体的位置由坐标 (x, y, z) 来表示. 我们就能认为表示运动的任务由三个标量函数 $x(t), y(t), z(t)$ 来承担. 其实这与上面介绍的用矢量来表示位置是非常一致的, 因为实际上点的坐标通常也被理解为原点指向这个点的矢量在三个基矢的坐标:

$$\mathbf{r} = (x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

出于哈密顿(Hamilton)爵士关于把三维空间三个方向基矢视作四元数代数运算的旋转等效的奇妙观点, 我们分别把 x, y, z 方向的基矢按照惯例记做 $\mathbf{i}, \mathbf{j}, \mathbf{k}$. 也常常因为它们是单位矢量(长度为 1), 而记做 $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$. 这样便能把 \mathbf{r} 分解出 x, y, z 来.

分解以后, 重新合成便能得到需要的物理量. 这其中的意思是: 比如我们为了求一个曲线运动速度或者加速度的 x 分量. 也就是求下列表达式:

$$\begin{aligned}\mathbf{v} &= v_x \mathbf{i} + v_y \mathbf{j} + v_z \mathbf{k} = \frac{d\mathbf{r}}{dt} \\ \mathbf{a} &= a_x \mathbf{i} + a_y \mathbf{j} + a_z \mathbf{k} = \frac{d^2\mathbf{r}}{dt^2}\end{aligned}$$

中的分量 v_x, a_x . 其求法非常直接, 便是:

$$v_x = \frac{d}{dt} x(t) \quad , \quad a_x = \frac{d^2}{dt^2} x(t)$$

这相当于是说, 求导数和向某固定方向投影是可以交换前后次序的:

$$\mathbf{i} \cdot \frac{d}{dt} \mathbf{r} = \frac{d}{dt} (\mathbf{i} \cdot \mathbf{r}) \quad , \quad \mathbf{i} \cdot \frac{d}{dt} \mathbf{v} = \frac{d}{dt} (\mathbf{i} \cdot \mathbf{v})$$

这是当然的, 因为矢量点乘的导数的法则也符合与标量乘积求导数类似的莱布尼茨法则, 而常矢量 \mathbf{i} 不随着运动改变. 所以导数只有对 \mathbf{r}, \mathbf{v} 求导的项.

在数学上, 曲线段也经常被描述为是与某实数上区间同胚(homeomorphic)的几何图形, 这其实也就是说, 存在自变量 t , 不失一般性地, 我们让其取值范围为 $(-\infty, +\infty)$. 而存在一个从这个自变量到三维空间点 (x, y, z) 的光滑映射:

$$x = x(t) \quad , \quad y = y(t) \quad , \quad z = z(t)$$

⁹指运动学运动方程, 动力学运动方程就是反应运动的规律的, 能够最终解出来运动学运动方程的微分方程.

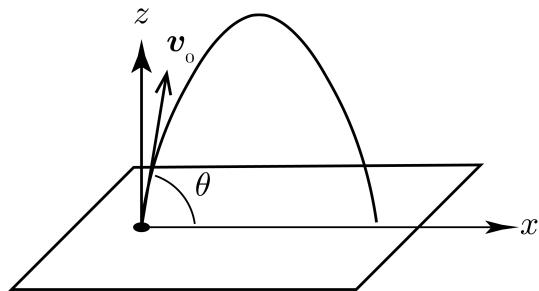


图 2.5: 抛体运动

在数学上看,这其实就是曲线方程:其参数方程的唯一给法.这样的式子就给出了几何学上的光滑曲线的定义.而不同于以往我们对曲线是 $y = f(x)$ 的图像的刻板认识.但在这里,我们恰巧发现如果把参数 t 理解为时间的话,这恰好就对应到了一个粒子在三维空间中的运动,出于这样的原因,我们把反应了粒子位置随着时间变化的方程称为运动方程.它是**轨迹方程**(function of trajectory) 的一种特殊情况.对于轨迹方程,我们只需要通过符合方程的点 (x, y, z) 构成粒子所有运动中经过的位置构成的集合即可.例如,在地面 $z = 0$ 出发以仰角 φ 斜抛一个物体.那就以出发点为原点建立坐标系,竖直向上为 z 轴,抛的方向向前为 x 轴.那么运动方程为:

$$\begin{cases} x = v_0 t \cos \varphi \\ z = v_0 t \sin \varphi - \frac{1}{2} g t^2 \end{cases}$$

但是轨迹方程不仅可以写成以上形式,还可以消去 t 得到:

$$z = x \tan \varphi - \frac{g x^2}{2 v_0^2} (1 + \tan^2 \varphi)$$

2. 平面极坐标系下的分解

如果质点做平面运动,那么我们问题得到了简化.此时往往有些情形下我们关心从原点的观察者出发看质点的运动时,质点离观察者的距离和质点相对观察者方位.故引入**极坐标系**(polar coordinate system) 是自然而然的做法.其步骤为:确定观察者位置为**极点**(pole),引出一条原点出发的射线为**极轴**(polar axis);测量质点位置到极点距离 r 为**矢径**(radius);从极点指向质点位置形成矢量矢径 \mathbf{r} ,测量极轴绕极点沿指定的正方向旋转到与矢径重合时旋转的角度 θ 为**极角**(polar angle).描述质点的位置时, (r, θ) 即构成极坐标.

仍然,极坐标可以用来表示质点的唯一位置:只需要给出 (r, θ) 就可以确定位置.所以只要知道函数 $r(t), \theta(t)$ 就能够确定其运动,这便是极坐标下的运动方程.同理如果消去 t 就能得到另一类轨迹方程.只要物体运动轨迹不经过极点,我们总是取 $r(t), \theta(t)$ 为光滑的函数. θ 的取值为 $(-\infty, +\infty)$ 以保证运动的连续性.不过 θ 相差 $2k\pi, k \in \mathbb{Z}$ 的所有坐标表示运动过程中的同一位置.

例如,在之前的抛体运动中,如果取 z 轴为极轴,那么其极坐标下的运动方程为:

$$\begin{cases} r^2 = (\mathbf{v}_0 t + \frac{1}{2} \mathbf{g} t^2)^2 = \frac{1}{4} g^2 t^4 - g v_0 t^3 \sin \varphi + v_0^2 t^2 \\ \tan \theta = \frac{2 v_0 \cos \varphi}{2 v_0 \sin \varphi - g t} \end{cases}$$

这会有利于我们分析一些问题.例如如果关心以何种倾角抛出一个物体,物体离我们的距离会越来越远?就只需要让 r 为关于 t 的增函数,除了对 r 求导的做法,一种更加明智的做法可能是注意到关系式:

$$\frac{dr}{dt} > 0 \Leftrightarrow \frac{d}{dt} \left(\frac{1}{2} r^2 \right) > 0$$

$$\begin{aligned} \Leftrightarrow \frac{d}{dt} \left(\frac{1}{2} r^2 \right) &= \frac{d}{dt} \left(\frac{1}{2} \mathbf{r} \cdot \mathbf{r} \right) = \mathbf{r} \cdot \mathbf{v} \\ &= (\mathbf{v}_0 t + \frac{1}{2} \mathbf{g} t^2) \cdot (\mathbf{v}_0 + \mathbf{g} t) > 0 \end{aligned}$$

$$\Leftrightarrow (\mathbf{v}_0 + \frac{1}{2}g\mathbf{t}) \cdot (\mathbf{v}_0 + g\mathbf{t}) = \frac{1}{2}g^2t^2 - \frac{3}{2}gv_0t \sin \varphi + v_0^2 > 0$$

这就化为求关于 t 的二次函数恒大于零的系数条件问题, 结论是初等的: 判别式小于零:

$$\left(\frac{3}{2}gv_0 \sin \varphi\right)^2 - 4\left(\frac{1}{2}g^2\right)(v_0^2) < 0 \Rightarrow \sin \varphi < \frac{2\sqrt{2}}{3}$$

采用极坐标系描述抛体运动也许是费力而不讨好的. 但是描述其他一些运动又比直角坐标系存在天然的优势. 比如一类经典追击问题的表述如下: 四个平面运动的质点初始组成正方形. 这个环形的结构中每一个质点都向着前一个质点运动发生追击. 其中运动速度矢量具有变化的方向, 即向着前方质点. 但具有不变的速率 v . 确定物体的运动方程和轨迹.

对于这个问题, 预先的关于物体在时间为 t 时的位置信息是几乎完全空缺的, 但对于速度矢量我们知道一个信息: 其大小为 v . 如果考虑直角坐标系, 这个条件表示为:

$$|\mathbf{v}| = \sqrt{\dot{x}^2 + \dot{y}^2} = v$$

但是对于位置和速度, 我们还知道一个信息, 就是它们夹钝角而且为 $3\pi/4$. 用直角坐标系的公式仔细化简的话也可以得到:

$$\tan \phi_{\mathbf{r} \wedge \mathbf{v}} = -1 \Rightarrow \frac{\mathbf{r} \cdot \mathbf{v}}{|\mathbf{r} \times \mathbf{v}|} = \frac{x\dot{x} + y\dot{y}}{x\dot{y} - y\dot{x}} = -1$$

但这就较难求解¹⁰. 那么如果采用极坐标系会为我们带来何种新的思路呢?

那就是活动标架(moving frame) 的引入. 无论粒子在何处, 粒子的 \mathbf{r} 一定是作为 r 和 θ 的函数的. 那么可以在粒子所在的点沿两个方向建立局部的 $R - \Theta$ 直角坐标系. 这两个方向怎么确定呢? 其实就是基矢的方向:

$$\mathbf{e}_r := \frac{\partial \mathbf{r}}{\partial r} / \left| \frac{\partial \mathbf{r}}{\partial r} \right| = \frac{\partial \mathbf{r}}{\partial r}, \quad \mathbf{e}_\theta := \frac{\partial \mathbf{r}}{\partial \theta} / \left| \frac{\partial \mathbf{r}}{\partial \theta} \right| = \frac{1}{r} \frac{\partial \mathbf{r}}{\partial \theta}$$

上式中后一个偏导数的模为 r 是显然的. 只需要考虑到圆的弧长和半径的关系即可.

局部的标架给我们什么方便? 我们由偏导数和全微分的法则可以描述局部物体的微小位移. 本质上, 这相当于 dr 和 $d\theta$ 导致的位矢 \mathbf{r} 改变 $d\mathbf{r}$:

$$\begin{aligned} d\mathbf{r}_{(dr, d\theta)} &= d\mathbf{r}_{(dr, 0)} + d\mathbf{r}_{(0, d\theta)} = dR\mathbf{e}_r + d\Theta\mathbf{e}_\theta \\ d\mathbf{r}_{(dr, d\theta)} &= \frac{\partial \mathbf{r}}{\partial r} dr + \frac{\partial \mathbf{r}}{\partial \theta} d\theta = dr\mathbf{e}_r + rd\theta\mathbf{e}_\theta \\ \Rightarrow dR &= dr, \quad d\Theta = rd\theta \end{aligned}$$

可见在曲线坐标情形下坐标的改变量 $dr, d\theta$ 与在对应质点位移方向走过的弧长 $dR, d\Theta$ 的值一般是不一样的. 它们之前的比值 (弧长/坐标变化) 称作拉梅系数(Lamé coefficient).

既然叫做活动标架法, 那么显然随着粒子的运动, 两个基矢 $\mathbf{e}_r, \mathbf{e}_\theta$ 的方向也在不断改变中. 我们用极坐标, 就是把一个物理量 \mathbf{A} 分解到粒子运动的瞬时局部坐标系中:

$$\mathbf{A} = A_r \mathbf{e}_r + A_\theta \mathbf{e}_\theta$$

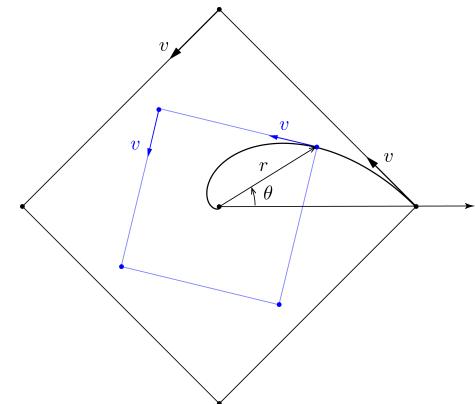


图 2.6: 首尾相追

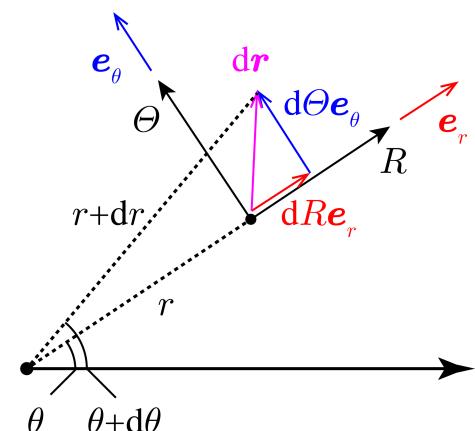


图 2.7: 极坐标系的活动框架

¹⁰有兴趣的读者可以挑战一下.

那么在粒子发生运动后,以上矢量的改变应当由四部分组合而成:

$$d\mathbf{A} = d'\mathbf{A} + \tilde{d}\mathbf{A}$$

$$d'\mathbf{A} := dA_r \mathbf{e}_r + dA_\theta \mathbf{e}_\theta$$

$$\tilde{d}\mathbf{A} := A_r d\mathbf{e}_r + A_\theta d\mathbf{e}_\theta$$

其中 d' 表示相对变化,类似于选取跟随旋转的活动标架而旋转的参考系后,矢量的变化量,故基矢自己的变化是被无视了的.而原来的 d 才表示绝对的变化.而 \tilde{d} 所带来的变化称作牵连变化或随体变化.它是即使前面的相对变化为零也会带来的,由于标架的“拉拽”导致的旋转带来的改变.事实上很容易证明其中的:

$$d\mathbf{e}_r = d\boldsymbol{\theta} \times \mathbf{e}_r = +d\theta \mathbf{e}_\theta$$

$$d\mathbf{e}_\theta = d\boldsymbol{\theta} \times \mathbf{e}_\theta = -d\theta \mathbf{e}_r$$

其中 $d\boldsymbol{\theta}$ 是方向垂直于纸面向外的,大小即极坐标下物体坐标之角坐标的增量 $d\theta$.

这就足以给出:

$$(d\mathbf{A})_r = dA_r - A_\theta d\theta$$

$$(d\mathbf{A})_\theta = dA_\theta + A_r d\theta$$

两边再同时除以时间,我们得到粒子运动过程中一个矢量导数的两个分量与这矢量两个分量的导数之间的关系:

$$(\dot{\mathbf{A}})_r = \dot{A}_r - \dot{\theta} A_\theta$$

$$(\dot{\mathbf{A}})_\theta = \dot{A}_\theta + \dot{\theta} A_r$$

导数的分量不等于分量的导数,这也是坐标系的弯曲所带来的.数学上,这意味着算符 ∂_t 与算符 $\mathbf{e} \cdot$ 的不对易性(non-commutative):

$$\mathbf{e} \cdot (\partial_t \mathbf{A}) \neq \partial_t (\mathbf{e} \cdot \mathbf{A})$$

通过这个原理,我们可以推理得到:

$$\mathbf{r} = r\mathbf{e}_r + 0\mathbf{e}_\theta \Rightarrow r_r = r, r_\theta = 0$$

$$\mathbf{v} = \dot{\mathbf{r}} \Rightarrow v_r = \dot{r}_r - \dot{\theta} r_\theta = \dot{r}, v_\theta = \dot{r}_\theta + \dot{\theta} r_r = \dot{\theta} r$$

$$\mathbf{a} = \ddot{\mathbf{r}} \Rightarrow a_r = \ddot{v}_r - \dot{\theta} v_\theta = \ddot{r} - \dot{\theta}^2 r, a_\theta = \ddot{v}_\theta + \dot{\theta} v_r = \ddot{\theta} r + 2\dot{\theta} \dot{r}$$

这个操作甚至可以继续下去,计算急动度等:

$$\mathbf{j} = \dddot{\mathbf{r}} \Rightarrow j_r = \ddot{a}_r - \dot{\theta} a_\theta = \dddot{r} - 3\dot{\theta} \ddot{\theta} r - 3\dot{\theta}^2 \dot{r}, j_\theta = \ddot{a}_\theta + \dot{\theta} a_r = \ddot{\theta} \ddot{r} + 3\dot{\theta} \dot{r} + 3\dot{\theta}^2 r - \dot{\theta}^3 r$$

...

极坐标下速度,加速度的公式是非常常用的,它们是:

$$\mathbf{v} = (\dot{r}, \dot{\theta} r) \quad \mathbf{a} = (\ddot{r} - \dot{\theta}^2 r, \ddot{\theta} r + 2\dot{\theta} \dot{r})$$

故,在之前的首尾相追问题2.6中以正方形中心为原点,以初始位置为极轴方向建立极坐标系.那么两个条件:速度大小恒定为 v 和位矢速度夹角恒定 $3\pi/4$ 表述为:

$$v_r^2 + v_\theta^2 = v^2, \quad \frac{v_\theta}{v_r} = -1$$

相当容易解出来两个分速度的值,再结合之前的速度公式:

$$\dot{r} = -\frac{\sqrt{2}}{2} v, \quad \dot{\theta} r = \frac{\sqrt{2}}{2} v$$

这样,便可以轻松解出运动方程来,结合初始条件,不妨设初始矢径为 r_0 ,幅角为 0:

$$\dot{r} = -\frac{v}{\sqrt{2}} \Rightarrow r = r_0 - \frac{vt}{\sqrt{2}}$$

$$\dot{\theta} = \frac{v}{\sqrt{2}r_0 - vt} \Rightarrow \theta = -\ln\left(1 - \frac{vt}{\sqrt{2}r_0}\right)$$

轨迹方程通过上述方程消参即可.但也可以通过之前的微分方程相除得到新微分方程求解,曲线为对数螺线(logarithmic spiral):

$$\frac{dr}{d\theta} = -r \Rightarrow r = r_0 e^{-\theta}$$

平面的极坐标系在三维空间的推广为柱坐标系(cylindrical coordinate system)和球坐标系(spherical coordinate system).类似公式的核心也与平面极坐标相差无几,都在于活动标架的基矢微分.请读者自行推导.

3. 自然坐标系下的分解

设想仍然采用活动标架法的思路,但是,标架的方向不再由粒子的位置相对事先建立的背景的坐标系的关系而决定.而是紧密地由粒子运动的性质:瞬时的速度方向,轨迹曲线的弯曲方向等决定.那么我们就得到了特殊的标架:自然坐标系(natural coordinates).

自然坐标系的建立是派生于粒子运动的轨迹的那条曲线的性质的.借助牛顿等人创立的微积分与笛卡尔等人创立的解析几何的帮助,对于光滑曲线曲面的性质的研究构成了古典微分几何(classical differential geometry)的主要课题.

我们考虑三维空间问题.对于一条空间曲线.它总是一些点的集合,而如果规定某个起始点以后,每一个点都会有一个 s 坐标,它就是从起始点到这个点的弧长.而其实我们可以把 s 视作一个自变量,因为 s 改变而改变的是不同点的位置 r .也就是说 $r(s)$ 作为 s 的函数.那么这个曲线在某一点处的切矢量就可以被自然地定义为:

$$\tau = \frac{dr}{ds}$$

以上矢量模自然为 1.因为分子 dr 的长度就是分母的 ds .方向则是在曲线的切线(tangent line)方向.切线是这样一种线:曲线上有一点 P .如果在 P 两侧任取一个点 Q 并连接 PQ ,使 Q 无限接近 P 即 $ds \rightarrow 0$,那么 PQ 所在的直线的极限就是 P 点的切线.这又引申出第二个概念:如果在 P 所有各取一个点 Q, R .那么 PQR 三点确定的圆也将随着 Q, R 同时无限接近于 P 而具有极限.这个圆叫做密切圆或曲率圆(osculating circle).其所在的平面叫做密切平面(osculating plane).而圆心 C 为曲率中心(center of curvature),最后,实际上 P 到 C 的方向就是法(normal)方向,该方向的单位矢量叫做法矢量 n .而副法矢量 $\tau \times n = \beta$ 所决定的方向为副法(binormal)方向.这样三个矢量便构成弗雷内框架(Frenet frame)或称 TNB 框架.

事实上,如果计算曲线在过程中的转向,它可以用三个矢量方向的变化来表示.如果计算三个矢量对 ds 的变化率,一定能写:

$$\frac{d\tau}{ds} = +0\tau + k n + \varepsilon \beta$$

$$\frac{dn}{ds} = -k\tau + 0n + \gamma \beta$$

$$\frac{d\beta}{ds} = -\varepsilon\tau - \gamma n + 0\beta$$

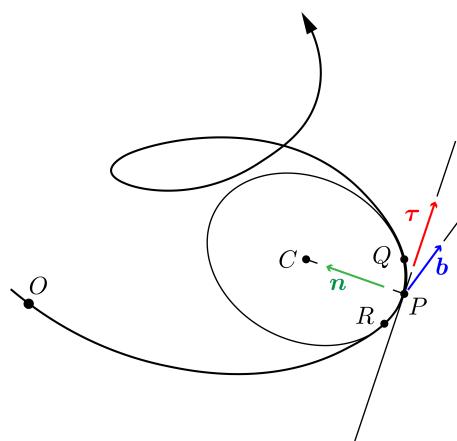


图 2.8: Frenet 框架

这是因为框架一定要具有正交归一的一套基矢量,而且这个性质在曲线上任何一点都要成立,指的是以下六个式子:

$$\begin{aligned}\boldsymbol{\tau} \cdot \boldsymbol{\tau} &= 1, \quad \boldsymbol{n} \cdot \boldsymbol{n} = 1, \quad \boldsymbol{\beta} \cdot \boldsymbol{\beta} = 1 \\ \boldsymbol{\tau} \cdot \boldsymbol{n} &= 0, \quad \boldsymbol{n} \cdot \boldsymbol{\beta} = 0, \quad \boldsymbol{\beta} \cdot \boldsymbol{\tau} = 0\end{aligned}$$

对六个式子分别求导数就看出来了,一个单位矢量自己的导数必然垂直于自身,从而三个基矢导数在自己上的投影都必须为零.而两个不同基矢的导数互相在对方方向的投影则会正负相消,这样才能保证两个矢量始终正交.就能理解之前的9个系数为什么要写成这个形式.

最后我们指出,系数 ε 应当是0.这是因为作为法线方向的定义本身,它就应当在 $\boldsymbol{\tau}$ 的变化量 $d\boldsymbol{\tau}$ 所在的方向上.从而第一个式子不会具有 $\boldsymbol{\beta}$ 方向的分量.综上所述,我们得到了:

$$\begin{aligned}\frac{d\boldsymbol{\tau}}{ds} &= +k\boldsymbol{n} \\ \frac{d\boldsymbol{n}}{ds} &= -k\boldsymbol{\tau} + \gamma\boldsymbol{\beta} \\ \frac{d\boldsymbol{\beta}}{ds} &= -\gamma\boldsymbol{n}\end{aligned}$$

这就是弗雷内公式(Frenet formulae).其中 k 称作曲线的曲率(curvature).它的意义是衡量了曲线方向转动的快慢,事实上由于曲线切线的无穷小转动角度其实在竖直上恰好会等于无穷小矢量 $d\boldsymbol{\tau}$ 的长度,从而有:

$$k = \left| \frac{d\boldsymbol{\tau}}{ds} \right| = \frac{d\theta}{ds}$$

即,曲率为:质点沿曲线走单位长度所转过的角度.曲率半径实际上是曲率的倒数.它是曲率圆的半径 $\rho = 1/k$.

而 γ 则被称作挠率(torsion).我们通过弗雷内公式的第三式可以发现,挠率反映了副法矢量,或者说被复法矢量作为法平面的密切平面,绕曲线此时的切线的旋转快慢.它使得粒子瞬时运动的那个平面在改变着.下面我们会发现,实际上挠率对运动学中直到加速度的表达式都是没有任何贡献的.

如何用自然法描述粒子的运动?事实上这就相当于把粒子运动的路程 s 和三个矢量 $\boldsymbol{\tau}, \boldsymbol{n}, \boldsymbol{\beta}$ 都表示为时间的函数:

$$s(t), \boldsymbol{\tau}(t), \boldsymbol{n}(t), \boldsymbol{\beta}(t)$$

但事实上,曲线本身轨迹必须先存在才能让我们事先确定好框架,而这样一个曲线其实本身由 $\boldsymbol{r}(s)$ 给出,而且满足:

$$\frac{d\boldsymbol{r}}{ds} = \boldsymbol{\tau}, \quad \frac{d^2\boldsymbol{r}}{ds^2} = k\boldsymbol{n}, \quad \frac{d^3\boldsymbol{r}}{ds^3} = -k^2\boldsymbol{\tau} + \frac{dk}{ds}\boldsymbol{n} + k\gamma\boldsymbol{\beta}$$

然后我们唯一要做的就是给出 $s(t)$ 函数,这一定可以帮助我们确定运动的各个量.

事实上,由于泰勒展开公式,近似到三阶:

$$\Delta\boldsymbol{r} \simeq \frac{d\boldsymbol{r}}{ds}ds + \frac{1}{2}\frac{d^2\boldsymbol{r}}{ds^2}ds^2 + \frac{1}{6}\frac{d^3\boldsymbol{r}}{ds^3}ds^3 = (kds - \frac{k^2}{6}ds^3)\boldsymbol{\tau} + (\frac{k}{2}ds^2 + \frac{1}{6}\frac{dk}{ds}ds^3)\boldsymbol{n} + \frac{k\gamma}{6}ds^3\boldsymbol{\beta}$$

如果只保留领头项,忽略高阶小量.我们就得到,在粒子运动的局域坐标系中,位移 ds 长度到达的点三个坐标 T, N, B 分别近似为:

$$T = ds, \quad N = \frac{k}{2}ds^2, \quad B = \frac{k\gamma}{6}ds^3$$

仍然考虑之前的三阶近似下的严格 $\Delta\boldsymbol{r}$ 与 ds 的关系,这一次代入 $s(t)$ 的三阶泰勒展开:

$$ds \simeq \dot{s}dt + \frac{\ddot{s}}{2}dt^2 + \frac{\dddot{s}}{6}dt^3$$

我们整理 dt 的各阶项,最终得到:

$$\Delta\boldsymbol{r} \simeq \boldsymbol{v}dt + \frac{1}{2}\boldsymbol{a}dt^2 + \frac{1}{6}\boldsymbol{j}dt^3$$

其中速度与加速度的公式为:

$$\mathbf{v} = \dot{s}\boldsymbol{\tau}$$

$$\mathbf{a} = \ddot{s}\boldsymbol{\tau} + k\dot{s}^2\mathbf{n}$$

速度方向必然沿切向, 大小上就等于 \dot{s} , 即单位时间走过的弧长. 而加速度则有两个分量. 一是由于速率在增减导致的切向加速度, 大小等于 $a_\tau = \ddot{s}$; 第二项则是由于曲线的弯曲导致的法向加速度. 它的值为 $a_n = \dot{s}^2/\rho$.

如果要计算急动度等, 可以从之前的泰勒展示中获得, 其实还可以直接对加速度求导数, 此时表达式会突然产生很多项:

$$\mathbf{j} = (\ddot{s} - k^2\dot{s}^3)\boldsymbol{\tau} + (3k\dot{s}\ddot{s} + \dot{k}\dot{s}^3)\mathbf{n} + k\gamma\dot{s}^3\boldsymbol{\beta}$$

利用自然坐标下的分解方式, 如果已经直到特定曲线的曲率, 就可以根据质点在曲线上的运动方式来计算运动的加速度. 但是反过来利用运动学公式来确定曲线的曲率半径也不失为一种极其简单的做法. 我们注意到, 利用加速度在垂直于速度方向的那个分量即为 v^2/ρ 的结论, 可以得到:

$$\rho = \frac{\mathbf{v}^3}{|\mathbf{a} \times \mathbf{v}|}$$

例如, 如果已知二维空间中的曲线方程 $y = f(x)$. 那么如果令一质点在曲线上运动且满足 $x = t$, $y = f(t)$, 那么速度为 $\mathbf{v} = (1, f')$, 加速度为 $\mathbf{a} = (0, f'')$. 利用上式可得:

$$\rho = \frac{(1 + f'^2)^{\frac{3}{2}}}{|f''|}$$

再例如. 之前的首尾相追问题中, 曲线对数螺线写作:

$$r = a e^{k\theta}$$

如何计算曲率半径? 我们让粒子在曲线上绕极点做匀速旋转, 也就是:

$$\theta = \omega t, \quad r = a e^{k\omega t}$$

那么借助极坐标的加速度公式得:

$$\mathbf{v} = (\dot{r}, \dot{\theta}r) = (k\omega a e^{k\omega t}, \omega a e^{k\omega t}) = (k\omega r, \omega r)$$

$$\mathbf{a} = (\ddot{r} - \dot{\theta}^2 r, \ddot{\theta}r + 2\dot{\theta}\dot{r}) = ((k^2 - 1)\omega^2 a e^{k\omega t}, 2k\omega^2 a e^{k\omega t}) = ((k^2 - 1)\omega^2 r, 2k\omega^2 r)$$

代入之前的曲率公式便可以得到:

$$\rho = \sqrt{1 + k^2 r}$$

2.2.2 刚体的运动

关于刚体的运动, 它的定义我们其实将在之后的刚体专门的章节又一次做更进一步的探讨. 初步我们认为所谓刚体就是这样一种物体: 存在一个特殊的参考系(平动或转动)能够使得该物体上每一个点都每时每刻相对静止. 但是这样我们又涉及到下一节才讲的参考系的概念. 在此我们又不得不早于预期提及.

为此我们换一种更加直观的方式引入这些概念而不是去严格定义刚体与参考系的概念. 设想我们有一个正交的刚性框架, 即三条射线状坐标轴构成的物体, 下称 ξ, η, ζ 轴, 该框架就会自然造成四个矢量. 首先是框架原点 C 的位置矢量 \mathbf{r} . 其次是

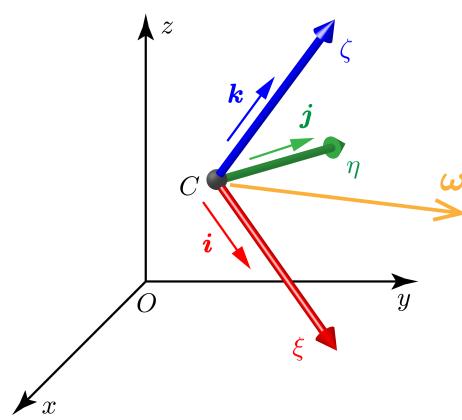


图 2.9: 刚体的本征框架

三坐标轴方向的三个基矢¹¹ $\mathbf{i}, \mathbf{j}, \mathbf{k}$. 它们都会随着时间变化, 可以用四个导数表征. 其中 $v = \dot{r}, a = \ddot{r}$. 但是, 对于三个基矢导数有没有更加简单的描述方法呢? 如果我们把三个基矢导数就向三个基矢本来的方向分解¹²:

$$\ddot{\mathbf{i}} = +\omega_{11}\mathbf{i} + \omega_{12}\mathbf{j} + \omega_{13}\mathbf{k}$$

$$\ddot{\mathbf{j}} = +\omega_{21}\mathbf{i} + \omega_{22}\mathbf{j} + \omega_{23}\mathbf{k}$$

$$\dot{\mathbf{k}} = +\omega_{31}\mathbf{i} + \omega_{32}\mathbf{j} + \omega_{33}\mathbf{k}$$

依然, 我们注意到出于以下正交归一的结构不会改变, 在求导下为零, 所以上述表达式中的九个系数反对称:

$$\frac{d}{dt} \begin{cases} 1 = \mathbf{i} \cdot \mathbf{i} & \mathbf{j} \cdot \mathbf{j} & \mathbf{k} \cdot \mathbf{k} \\ 0 = \mathbf{i} \cdot \mathbf{j} & \mathbf{j} \cdot \mathbf{k} & \mathbf{k} \cdot \mathbf{i} \end{cases} = 0 \Rightarrow \omega_{ij} + \omega_{ji} = 0$$

出于某种符号约定与对称的原因, 我们把导数重新写作:

$$\ddot{\mathbf{i}} = +\omega_3\mathbf{j} - \omega_2\mathbf{k}$$

$$\ddot{\mathbf{j}} = -\omega_3\mathbf{i} + \omega_1\mathbf{k}$$

$$\dot{\mathbf{k}} = +\omega_2\mathbf{i} - \omega_1\mathbf{j}$$

这样, 如果我们定义一个矢量:

$$\boldsymbol{\omega} = \omega_1\mathbf{i} + \omega_2\mathbf{j} + \omega_3\mathbf{k}$$

就能够造成:

$$\ddot{\mathbf{i}} = \boldsymbol{\omega} \times \mathbf{i}, \quad \ddot{\mathbf{j}} = \boldsymbol{\omega} \times \mathbf{j}, \quad \dot{\mathbf{k}} = \boldsymbol{\omega} \times \mathbf{k}$$

这就是刚体的角速度(angular velocity) 矢量. 用它可以计算刚体中的各种转动的物理量.

具体来说, 什么叫做刚体? 事实上以上三个坐标轴本身就构成了刚体, 或者是某刚体的一部分. 如果有一个空间点每时每刻向杆坐标系的三个坐标面引三垂线, 三个坐标都始终保持常数, 那么就说这个空间点是固连(fixed) 在该坐标系中的. 而如果一个物体上的每一个点都固连带该坐标系中, 那么相对这个坐标系该物体就无法发生形变, 从而这就是一个刚体. 这种情况下, 刚体与坐标系就存在一个相互关系, 这个坐标系是由刚体所对应的固连坐标系, 而坐标系本身又可以视作某种不能变形的刚体.

我们在坐标系中找到固连在坐标系上的两个点, 或者说在刚体上找到固连在刚体上的两个点 (以后不加以区分). 从一个点指向另一个点形成矢量 \mathbf{R} , 由于固连在坐标系中, 这个矢量对坐标系造成的三个分量都是固定的, 也就是 $\mathbf{R} = R_1\mathbf{i} + R_2\mathbf{j} + R_3\mathbf{k}$ 中的 R_1, R_2, R_3 都是常数. 那么这个矢量跟随坐标系而变化的导数就是:

$$\dot{\mathbf{R}} = R_1\ddot{\mathbf{i}} + R_2\ddot{\mathbf{j}} + R_3\dot{\mathbf{k}} = \boldsymbol{\omega} \times (R_1\mathbf{i} + R_2\mathbf{j} + R_3\mathbf{k}) = \boldsymbol{\omega} \times \mathbf{R}$$

在物理上, 这就是说, 任何固连在转动坐标系中的矢量 \mathbf{R} 都将以角速度 $\boldsymbol{\omega}$ 跟随坐标系旋转, $\boldsymbol{\omega}$ 的方向就是瞬时转动轴, $\boldsymbol{\omega}$ 的大小就是旋转角速度大小. 注意两点, 一是这个角速度 $\boldsymbol{\omega}$ 是全局的, 无论矢量相对坐标系在任何位置, 无论距离最初选取的坐标系中心有多远, 只要固连在坐标系中, 就会以完全一样的角速度矢量 $\boldsymbol{\omega}$ 去改变 $\dot{\mathbf{R}} = \boldsymbol{\omega} \times \mathbf{R}$. 二是这个角速度作用在任何量纲上. 无论是两个空间点之间的那个 (位移) 矢

¹¹原来的基矢方向记做 $e_{x, y, z}$.

¹²由于字母 i, j 上本来就有点, 所以在符号上约定遇到要在这两个字母上加点时改为横向 i, j . 偶尔也见到反而去掉点的 i, j . 出于这样无法表示高阶导数, 我们不采用这种模式.

量,还是速度矢量,力矢量等等,任意量纲的矢量 \mathbf{A} 只要相对旋转坐标系具有固定的方向,在地面系中就会跟着坐标系一同旋转而具有导数:

$$\dot{\mathbf{A}} = \boldsymbol{\omega} \times \mathbf{A} \quad \text{for } \mathbf{A} \text{ fixed in frame } C - \xi\eta\zeta$$

以后我们可以脱离坐标系而讨论刚体.为了研究刚体的运动,我们只需要在刚体上取一个合适的固连点 C 作为中心(类似于坐标系的原点),这个点叫做基点(cardinal point).那么,不失一般性地,刚体的运动将用基点 C 的位置 $\mathbf{r}(t)$ 和角速度 $\boldsymbol{\omega}(t)$ 描述.譬如,为了求得某一时刻刚体上一点 P 的速度.我们可以找到从 C 指向 P 的矢量 \mathbf{R} .那么,由于 $\mathbf{r}_{OC} + \mathbf{r}_{CP} = \mathbf{r}_{OP}$.这就是说:

$$\mathbf{v}_P = \dot{\mathbf{r}}_{OP} = \dot{\mathbf{r}}_{OC} + \dot{\mathbf{r}}_{CP} = \mathbf{v}_C + \dot{\mathbf{R}}$$

最后再由之前的用角速度算导数的方法,得到著名的基点法速度表达式:

$$\mathbf{v}_P = \mathbf{v}_C + \boldsymbol{\omega} \times \mathbf{R}$$

即:刚体上一点速度等于跟随基点平动的速度 \mathbf{v}_C 加上绕基点以 $\boldsymbol{\omega}$ 转动的速度 $\boldsymbol{\omega} \times \mathbf{R}$ 的和.

至于加速度,只需要对上式再求一次导数 $\dot{\mathbf{v}}_P = \ddot{\mathbf{v}}_C + \dot{\boldsymbol{\omega}} \times \mathbf{R} + \boldsymbol{\omega} \times \dot{\mathbf{R}}$:

$$\mathbf{a}_P = \mathbf{a}_C + \beta \times \mathbf{R} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{R})$$

即:刚体上一点速度等于跟随基点平动的加速度 \mathbf{a}_C 加上绕基点以 β 转动的角向加速度 $\beta \times \mathbf{R}$,最后再加上由于速度 $\boldsymbol{\omega} \times \mathbf{R}$ 本身也在绕轴以 $\boldsymbol{\omega}$ 旋转带来的向轴加速度 $\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{R})$ 的和.

2.3 参考系变换

2.3.1 点变换

事实上,对旋转坐标系的分析正如对刚体的分析那样,一个旋转坐标系的描述正由两点构成,一是这个参考系原点 C 的运动方程.给出它我们就可以求出任意时刻另一个参考原点的速度与加速度:

$$\mathbf{v}_C = \dot{\mathbf{r}}_C, \quad \mathbf{a}_C = \ddot{\mathbf{r}}_C$$

第二点就是要给出新参考系的旋转角速度 $\boldsymbol{\omega}$,角加速度记作 $\beta = \dot{\boldsymbol{\omega}}$.通过上一节我们知道,如果选取固连在新参考系的点 P ,并且把 CP 记作 \mathbf{r}' .那么 P 点速度加速度(现在记作 $\mathbf{v}_c, \mathbf{a}_c$)为:

$$\mathbf{v}_c = \mathbf{v}_C + \boldsymbol{\omega} \times \mathbf{r}'$$

$$\mathbf{a}_c = \mathbf{a}_C + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}') + \beta \times \mathbf{r}'$$

我们现在研究什么问题呢?如果我们在 $O-xyz$ 系中给出一个质点,也用 P 表示,的运动方程 $x(t), y(t), z(t)$.其实也就给出了在 $C-\xi\eta\zeta$ 系中的运动方程 $\xi(t), \eta(t), \zeta(t)$.那么把位矢 OP 记作 \mathbf{r} , CP 记作 \mathbf{r}' .再把 OC 记作 \mathbf{r}_C .于是得到恒等式:

$$\mathbf{r} = \mathbf{r}_C + \mathbf{r}'$$

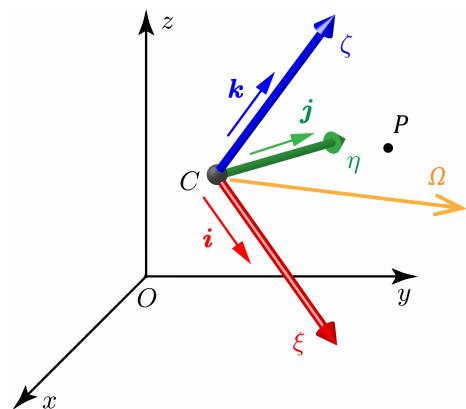


图 2.10: 参考系变换

但这不代表 $\mathbf{v} = \mathbf{v}_C + \mathbf{v}'$.这里我们要注意有一点已经悄悄发生了变化,那就是导数,作为一个算符,变成了一个有多种自由定义余地的概念.只要想想这一点:从运动方程求出速度的过程叫做求导¹³,但是同一个运动方程对应的运动是同一个运动,但是在不同参考系中依然可以有多个不同的取值.这就足以说明有多种不同的求导方法.具体来说,这里的 \mathbf{r}' 可能有三种情况:a. 固连在 C 系中,相对 C 系大小方向不变.b. 相对地面系 O 系大小方向都不变,即 P 跟随 C 相对地面系平移.c. 两者都不是.那么相应地定义三种导数:

¹³ 导数就是从运动方程到速度矢量的映射

1. 绝对导数: 即平常的 $\frac{d}{dt}$, 可以简单地用加点来代替. 我们还创造一个新的符号 D 来表示求导, 例如对于情形 b, 这个导数就是零. 求导数的方法是: 找到一个矢量在 O 系中每时每刻向三个坐标轴方向的投影分量并分别求导最后组合为矢量. 用这个导数能定义 P 点在 O 系中的速度与加速度:

$$\mathbf{v} = \text{D}\mathbf{r} (= \frac{d}{dt}\mathbf{r} = \dot{\mathbf{r}}) , \quad \mathbf{a} = \text{D}\mathbf{v}$$

2. 相对导数: 这就是要换相对 C 系的视角来研究一个矢量的导数了. 我们记作 D'. 例如对于情形 a, 那么相对导数就是零. 求导数的方法是: 找到一个矢量在 C 系中每时每刻向三个坐标轴方向的投影分量并分别求导最后组合为矢量. 这也就是说它用来定义相对速度和相对加速度是合适的:

$$\mathbf{v}' = \text{D}'\mathbf{r}' = r'_1\mathbf{i} + r'_2\mathbf{j} + r'_3\mathbf{k} , \quad \mathbf{a}' = \text{D}'\mathbf{v}' = \ddot{r}_1\mathbf{i} + \ddot{r}_2\mathbf{j} + \ddot{r}_3\mathbf{k}$$

3. 随体导数: 这是矢量由于跟随 C 系转动而带来的相对 O 系的变化率. 我们记作 \tilde{D} . 例如在情形 a 中, \mathbf{r}' 随体导数就是 $\omega \times \mathbf{r}'$. 这个算符其实就等价于 $\omega \times$, 这是因为他就定义为:

$$\tilde{D}\mathbf{A} = A_1\ddot{\mathbf{i}} + A_2\ddot{\mathbf{j}} + A_3\ddot{\mathbf{k}} = A_1\omega \times \mathbf{i} + A_2\omega \times \mathbf{j} + A_3\omega \times \mathbf{k} = \omega \times (A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k}) = \omega \times \mathbf{A}$$

值得注意的是, 这三个导数算符之间有如下关系:

$$D = \tilde{D} + D'$$

这很容易证明, 对一个矢量求导数, 一看分量变不变, 而看基矢量本身有没有旋转. 那么就会产生两项. 实际上就是莱布尼茨法则.

于是这样才能推理参考系变换的公式. 我们对 $\mathbf{r} = \mathbf{r}_C + \mathbf{r}'$ 两边求绝对导数:

$$\mathbf{v} = \text{D}\mathbf{r} = \text{D}\mathbf{r}_C + \text{D}\mathbf{r}' = \mathbf{v}_C + (\tilde{D} + D')\mathbf{r}' = \mathbf{v}_C + \omega \times \mathbf{r}' + \mathbf{v}'$$

再次对上式两边求绝对导数:

$$\begin{aligned} \mathbf{a} &= \text{D}\mathbf{v} = \text{D}\mathbf{v}_C + \text{D}\omega \times \mathbf{r}' + \omega \times \text{D}\mathbf{r}' + \text{D}\mathbf{v}' \\ &= \mathbf{a}_C + \beta \times \mathbf{r}' + \omega \times (\tilde{D} + D')\mathbf{r}' + (\tilde{D} + D')\mathbf{v}' \\ &= \mathbf{a}_C + \beta \times \mathbf{r}' + \omega \times (\omega \times \mathbf{r}') + \omega \times \mathbf{v}' + \omega \times \mathbf{v}' + \mathbf{a}' \\ &= \mathbf{a}_C + \beta \times \mathbf{r}' + \omega \times (\omega \times \mathbf{r}') + 2\omega \times \mathbf{v}' + \mathbf{a}' \end{aligned}$$

这两个公式就是参考系变换下的变换公式:

$$\mathbf{v} = \mathbf{v}_C + \omega \times \mathbf{r}' + \mathbf{v}'$$

$$\mathbf{a} = \mathbf{a}_C + \beta \times \mathbf{r}' + \omega \times (\omega \times \mathbf{r}') + 2\omega \times \mathbf{v}' + \mathbf{a}'$$

如何理解以上的公式? 事实上结构也很简单, 我们把加速度中的最独特的一项单独拿出来叫做柯里奥利加速度(Coriolis acceleration):

$$a'_c = 2\omega \times \mathbf{v}'$$

那么以上两公式具有以下结构:

$$\text{绝对} = \text{牵连} (+\text{柯氏}) + \text{相对}$$

柯氏即柯里奥利. 相对速度与相对加速度就是 \mathbf{v}' , \mathbf{a}' . 言下之意就是我们定义了所谓的牵连速度(convected velocity) 与牵连加速度(convected acceleration):

$$\mathbf{v}_c = \mathbf{v}_C + \omega \times \mathbf{r}'$$

$$\mathbf{a}_c = \mathbf{a}_C + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}') + \boldsymbol{\beta} \times \mathbf{r}'$$

细心的读者已经发现了,这两个式子和本节最开始推导出来的固连在 C 系中的点的绝对速度,加速度公式是一致的. 牵连速度(加速度),顾名思义,被转动参考系“带着走”的速度(加速度). 也正是以上规律启发我们总结出以下的动点动系法来理解看似复杂的速度加速度合成公式:

1. 确定定系, 动系, 动点, 定点. 定系往往就是地面系; 动系一般是相对某刚体或人为构造的刚性框架相对静止的参考系¹⁴; 动点就是选做研究对象的相对两个系一般都有运动的点. 最后, 定点指的是这样的概念: 在研究的一瞬间与动点重合, 但是始终保持与动系固连的特殊参考点.
2. 在定系考虑动点的运动. 算出绝对量. 在动系考虑动点的运动. 算出相对量. 在定系中考虑定点的运动, 它的量就是牵连量.
3. 建立绝对相对牵连之间的等量关系.
4. 唯一要注意的是: 只有当转动系是真转动的 $\boldsymbol{\omega} \neq \mathbf{0}$, 而且相对动系动点也是真运动的 $\mathbf{v}' \neq \mathbf{0}$ 时, 我们才有必要在加速度中添加一项柯氏加速度 $2\boldsymbol{\omega} \times \mathbf{v}'$. 从推导过程来看, 一是即使固连在动系中的速度矢量也会跟随动系转动而带来加速度, 二是动系中的相对位矢产生变化而带来在变化的牵连速度从而产生加速度, 是造成柯氏加速度的两个因素. 这一项不是因为在动系中的某特定位置而造成的, 事实上与 \mathbf{r}' 无关, 决定性因素是“相对动系有速度”.

最后值得注意, 如果 $\boldsymbol{\omega}$ 和 $\boldsymbol{\beta}$ 恒为零, 这也并不少见, 那么这样的动系就是平动系(translational frame). 此时公式极为简单, 没有柯氏加速度, 牵连量就是动系原点或任意固连点的量:

$$\mathbf{v} = \mathbf{v}_C + \mathbf{v}' , \quad \mathbf{a} = \mathbf{a}_C + \mathbf{a}'$$

2.3.2 刚体变换

刚体的变换是点变换的直接推论. 同样地, 首先已知动系相对定系具有原点速度 \mathbf{v}_C 和加速度 \mathbf{a}_C 以及角速度 $\boldsymbol{\omega}$ 和加速度 $\boldsymbol{\beta}$, 又已知定系中刚体基点速度加速度, 现在记作 \mathbf{V}, \mathbf{A} . 事实上我们还要给出刚体自己的角速度角加速度才算是对刚体的完整描述, 它们为 $\boldsymbol{\Omega}, \mathbf{B}$. 那么绝对的和相对的各个量之间的关系为:

$$\mathbf{V} = \mathbf{v}_C + \boldsymbol{\omega} \times \mathbf{R}' + \mathbf{V}' , \quad \boldsymbol{\Omega} = \boldsymbol{\omega} + \boldsymbol{\Omega}'$$

$$\mathbf{A} = \mathbf{a}_C + \boldsymbol{\beta} \times \mathbf{R}' + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{R}') + 2\boldsymbol{\omega} \times \mathbf{V}' + \mathbf{A}' , \quad \mathbf{B} = \boldsymbol{\beta} + \boldsymbol{\omega} \times \boldsymbol{\Omega}' + \mathbf{B}'$$

证明过程留作练习由读者自己思考.

2.4 运动的牵连

最后本节旨在枚举一些在运动学解题中十分常见的初等模型与结论.

2.4.1 接触系

两个几何图形具有一个交点是一件非常常见的事情. 交点将同时与两个几何图形保持结合关系. 动态地看待这个问题, 我们可以分别利用两个几何图形自身的运动和交点相对两者的运动合起来写出交点的速度矢量与加速度矢量来, 但是作为同一个研究对象两种写法的结果理应相等. 这构成了一组方程.

尤其是, 在上述“相交系”中我们还有一类十分独特的情形, 就是两个几何图形相切. 两个平面形刚体如果始终保持相互接触, 其轮廓之间就是这样的关系. 此时要知道客观地这形除了一种约束: 不是让两个刚

¹⁴注意, 以下整个过程都可以脱离具体的坐标系而计算, 这就为坐标系的选取(种类是直角坐标系, 极坐标系还是自然坐标系, 方向如何)彻底提供了自由.

体做任意运动都可以让两者保持相切的. 所以必然存在所谓的约束方程(equation of constraint). 而这个方程的列法其实就是基于之前说的接触点的速度的一致性.

让我们看一个例子, 它不仅帮助我们理解接触系, 还可以帮助我们理解纯滚动. 就是经典的在粗糙的地面上做匀速纯滚动的圆盘. 首先应当注意到, 涉及到所谓的“瞬时接触点”时可能指的是以下三个点之一:

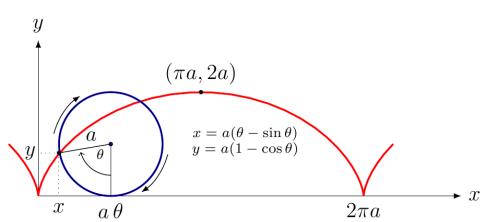


图 2.11: 圆盘在地面上做纯滚动
鼠做匀速直线运动.

1. 刚体 1 上的定点 P_1 . 在这里取盘为刚体 1, 此时这个点是做旋轮线运动的那个点, 此时恰好位于最低位置.

2. 刚体 2 上的定点 P_2 . 在这里取地面为刚体 2, 那么这个点就是一个静止的点.

3. 动点 P_3 . 就是每时每刻的切点形成的运动对象. 在这里我们可以把圆盘变为圆筒, 并且让一只松鼠在圆筒内部跑动以保持在最低位置, 也就是每时每刻的切点. 那么松

我们有什么结论呢? 如果把 P_1 的瞬时速度在法线方向与切线方向的分速度与加速度写出来, 分别是 $v_{1n}, v_{1\tau}, a_{1n}, a_{1\tau}$ 并约定法向以向刚体 1 内侧为正, 切向可以以切线的任意一个方向为正. P_2 同理为 $v_{2n}, v_{2\tau}, a_{2n}, a_{2\tau}$. 那么让我们考虑一下 P_3 的速度与加速度. 这就不难发现, 如果采用动点动系法, 就有:

速度: 无论相对刚体 1 还是 2, P_3 的速度都是牵连加相对. 牵连速度就是定点的速度, 但相对速度由于点的轨迹就是刚体的轮廓, 故方向必然在切线方向. 然后由于速度的唯一性, 这可以给出:

$$v_{1n} = -v_{2n}$$

加速度: 道理是相同的, 但需要注意两点, 一是相对加速度具有了法线方向的分量. 二是如果严格换一个刚体的转动系往往带来柯氏加速度. 具体解题时这仍然不失为一种常见的解题思路, 但是难以总结现成的美妙, 下一小节通过添加纯滚动这一更强的结论我们可以得到相应的方程. 但是这个方程肯定是存在且可列式的, 毕竟:

以上两个方程构成对接触系的约束的刻画.

2.4.2 纯滚系

上面的圆盘在地面上发生的运动实际上还是纯滚动. 我们将首先给出纯滚动的定义, 再来讨论纯滚动时的更强的附加条件, 毕竟纯滚动相比简单的接触系还多一个约束.

事实上考虑相对运动就知道了, 接触点在刚体 1 或 2 上相对运动的弧长就是相互滚动过的接触弧长, 在两个刚体上应当具有公共值, 这便是纯滚动的定义. 这就是说:

$$\mathbf{v}'_1 = \mathbf{v}'_2 = \dot{s}\boldsymbol{\tau}$$

最后再利用 $\mathbf{v} = \mathbf{v}_{1c} + \mathbf{v}'_1 = \mathbf{v}_{2c} + \mathbf{v}'_2$. 便得到了附加条件:

$$v_{1\tau} = v_{2\tau}$$

这就是常见的纯滚条件一, 也就是说只有当相互接触的两个刚体定点切向速度一致才构成纯滚动. 否则就是纯滚动的反面: 必有相对滑动, 俗称“连滚带滑”.

考虑加速度我们将得出一个不可思议但容易忽视的结论. 首先注意到, 如果两个刚体角速度分别为 ω_1 和 ω_2 , 作为平面问题写成统一正方向的标量. 而再设两刚体在接触点处轮廓的曲率为 k_1 和 k_2 . 那么由于相对两刚体点在轮廓上的运动造成的单位时间切线方向的转角为:

$$\frac{d\theta_1}{dt} = \frac{d\theta_1}{ds} \frac{ds}{dt} = k_1 \dot{s}$$

$$\frac{d\theta_2}{dt} = \frac{d\theta_2}{ds} \frac{ds}{dt} = k_2 \dot{s}$$

而每当点成为切点两切线也会重合成为公切线,那么两个转角的和就是单位时间内两个刚体的相对转角:

$$\frac{d\theta_1}{dt} + \frac{d\theta_2}{dt} = |\omega_1 - \omega_2|$$

最后,列出对刚体 1 和 2 牵连 + 柯氏 + 相对应当一致的方程:

$$a_{1n} - 2\omega_1 \dot{s} + k_1 \dot{s}^2 = -(a_{2n} - 2\omega_2 \dot{s} + k_2 \dot{s}^2)$$

整理得到纯滚条件二:

$$a_{1n} + a_{2n} = \frac{(\omega_1 - \omega_2)^2}{k_1 + k_2}$$

这是一个必然大于零的表达式¹⁵,它意在,两个相接触的点必然后来会分离,这个分离加速度只与两刚体瞬时转动快慢和接触点弯曲程度有关,与平动转动加速度都没有关系.

2.4.3 约束系

约束是一种普遍存在且可以普遍讨论而有很广的适用范围的物理现象.本节并非要详细讨论约束,这一点在之后第三章静力学中再加以详细讨论.这里只讨论两个问题:绳系和杆系.

对于不可伸长的轻绳模型,如果它也被拉直(因为内部受力不为零又没有外力).那么在运动学层面上就等价于不可伸长的轻杆.设杆长为 l ,把两端点 1 和 2 速度加速度沿平行与杆与垂直与杆方向分解,具有关系:

$$v_{1\perp} - v_{2\perp} = \omega l$$

$$v_{1\parallel} - v_{2\parallel} = 0$$

$$a_{1\perp} - a_{2\perp} = \beta l$$

$$a_{1\parallel} - a_{2\parallel} = -\omega^2 l$$

其中第二个关系就是著名的端点沿杆方向分速度必须是常数.对于绳子,即使绕过了定滑轮,显然沿定滑轮的切向速度也一直不变,因为绳子不可伸长.

最后,如果绳子允许弯曲但不允许伸长,其运动学约束方程该如何列?那就要求绳子上每一点在“局部”沿切向方向没有相对速度,表示为公式就是:

$$\frac{d\mathbf{v}}{ds} \cdot \boldsymbol{\tau} = 0$$

注意这个条件不可以写为:

$$\frac{dv_\tau}{ds} = \frac{d(\mathbf{v} \cdot \boldsymbol{\tau})}{ds} = 0$$

¹⁵注意 k 本身可以小于零,但两个 k 的和必然大于零.

第三章 动力学

3.1 牛顿定律

3.1.1 概述

中世纪到近代以来物理学的发展,可谓沧海桑田,旧的观点被否定,几十年以后突然又以全新的面貌出现在物理学理论中. 其中,实验中观察到各种令人惊诧的结果,无论物理学所依托的数学形式一变再变,要说亘古不变的主题,恐怕只有自然界的神奇与深邃,和我等凡人难以参透其中奥妙却仍为之孜孜不倦研究的决心. 物理是一门研究自然界存在的运动并提出解释与理解方法的学科. 运动形式多种多样,变化万千,但很多都被我们或多或少地解释,其实总结来看,从古至今我们也仅仅拥有过两类解释:一类由牛顿提出,一类则来自爱因斯坦. 前前后后经历了大概以下发展:

亚里士多德: 天体运行,自由落体是自然 (nature) 的运动; “树欲静而风不止” 是强迫 (violent) 的运动. 但错误地认为“物体越重,下落越快”

伽利略: 辩证地看待亚里士多德的观点,开创性地提出相对性原理,成功论证与形成惯性与常见运动的正确认识.

笛卡尔,惠更斯等: 成功得到大量动力学公式,包括后来的牛顿第二定律,动能的表达式与能量守恒定律,

牛顿开创的相互作用观: 开创性的提出物质间存在相互作用,相互作用是改变物质运动状态的原因. 并很好地把天体的运动归结到互相之间产生的相互作用上. 后人建立了场的理论以后,超距的粒子间相互作用就不再被理论学家们采纳了,而要理解为场与粒子的相互作用,再晚一点,场和粒子都被赋予了波粒二象性,粒子也要被理解为场,但所有场都要被分解为平面波,平面波又要被量子化. 如果没有相互作用,那么这些平面波的运动状态就不会发生改变,只有相互作用才会改变其运动,或传播方向发生偏折 (发生散射),或强度发生改变 (粒子的产生与湮灭). 牛顿的观点是以相互作用 (interaction) 为核心的观点.

拉格朗日,哈密顿等: 将牛顿的结果与观点上升到分析力学 (analytic mechanics) 的高度. 将复杂体系的结构浓缩在统一的一个能量函数或作用量泛函数中. 为后来场论建立和相对论,统计力学,量子力学的发展都提供了理论支持.

马赫: 批判牛顿的过于绝对的时空观. 预言质量分布会对惯性本身产生影响.

爱因斯坦开创的几何观: 仿佛又回到了亚里士多德,不仅仅是无相互作用下的匀速直线运动,一切在万有引力下的天体运动其实也是“自然”的运动. 看上去复杂的曲线运动不过也是沿对被引力所扭曲的弯曲时空几何下的天然“直线”---测地线 (geodesic) 运动而已. 对引力问题提出开创性的解释. 率先描绘了 20 世纪建立的若干理论的物理图像.

天体物理,粒子物理: 爱因斯坦以后的理论研究中很多问题还处于开放状态. 时而采用牛顿的相互作用观,时而采用爱因斯坦的几何观. 是一种两个图像的有机结合.

我们今天说**经典力学**(classical mechanics) 大多数情况即指**牛顿力学**(Newtonian mechanics), 指的是由伽利略最初引领, 笛卡尔, 惠更斯等奠基, 牛顿集大成的一套理论. 它以伽利略的**两种新科学的对话与数学展示**(Discorsi e dimostrazioni matematiche intorno a due nuove scienze) 和牛顿的**自然哲学的数学原理**(Philosophiæ Naturalis Principia Mathematica) 两篇文献为核心. 逻辑上以质点的牛顿三大定律为基础. 不包括后来上升到的分析力学层次, 作为绝对时空观适用伽利略变换, 不兼容相对论.

牛顿定律总结如下:

-
1. 第一定律: 在惯性系中, 一个质点应静止或匀速直线运动. 除非有外力施加于物体.
 2. 第二定律: 在惯性系中, 质点受力的矢量和 \mathbf{F} 等于其质量 m 与加速度 \mathbf{a} 的乘积.
 3. 第三定律: 一个质点对另一个质点施加作用力时, 另一个质点同时将施加反作用力于第一个质点, 作用力与反作用力等大反向共线.
-

三个定律之间的联系是很值得讨论的. 对三个基本定律的设置反映了很多物理思想¹. 分析如下:

3.1.2 牛顿第一定律

牛顿第一定律旨在阐述时空的平直性.

也正基于以上这一点, 我们不能仍为牛顿第一定律可以由牛顿第二定律出发推导. 它不是表面上的牛顿第二定律中 $\mathbf{F} = \mathbf{0}, \mathbf{a} = \mathbf{0}$ 的特例. 恰恰相反, 即使从形式上看, 也应当先由牛顿第一定律给出**惯性系**(inertial frame) 的概念才能适用牛顿第二定律. 那么怎样的参考系是惯性系呢? 牛顿第一定律是暗含认可**绝对惯性系**(absolute inertial frame) 的存在性的. 在这样第一个系中牛顿第一定律才可以成立. 这一个系无非就是由某种特殊运动状态的观察者(绝对观察者)建立的三维欧几里得空间坐标系外带一维时间坐标. 而只要质点不受力(力是对相互作用的描述, 只要远离其它所有物质就会使质点成为**孤立**(isolated)的), 就会静止或匀速直线运动, 所谓匀速直线运动是指:

$$\begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} = \begin{pmatrix} x(0) \\ y(0) \\ z(0) \end{pmatrix} + \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \cdot t$$

其中三个速度 v_x, v_y, v_z 都是常数. 但是我们发现, 相对这个系做匀速直线运动的另一个无自转观察者建立的系中, 原来做匀速直线运动的质点现在依然做匀速直线运动. 所以这样的系也叫做**惯性系**, 可以叫做**相对惯性系**. 即: 惯性系就是任何相对绝对惯性系做匀速直线平动的参考系. 而惯性系就是牛顿定律成立的前提条件.

所以牛顿的经典力学体系的第一大特点是承认绝对惯性系的存在性. 实际上承认这一点也几乎等价于承认了时空的平直性. 只不过在众多彼此等价(任何物理实验都不足以判断这些系的不等价性)的系中, 牛顿认为还应当有一个第一性的, “最初”的惯性系.

3.1.3 牛顿第二定律

牛顿第二定律是牛顿定律的核心, 旨在描述相互作用对运动的影响.

相互作用(interaction) 是牛顿经典力学后经过了几百年发展而浓缩出来的更普遍的概念. 如果没有相互作用, 物质基本的运动模式就是“惯性”的, 如匀速运动的粒子或传播的平面波. 任何使物质偏离这种“惯性”运动的模式的原因就都在相互作用上. 相互作用的模式是多种多样的, 除了牛顿指出的下面仔细讨论的

¹例如爱因斯坦狭义相对论的两个假设在某种意义上与牛顿第一第二定律的逻辑完全一致. 都是先定义时空背景再阐述物理规律的共性.

情形,光会因为与物质相互作用而产生吸收和散射,两个惰性气体原子靠近时两个电子云也会产生量子的诱导极化而使原子间产生范德瓦耳斯吸引.这些都是广义的相互作用,但是为了描述它们,前者需要借助波来描述光这种物质,后者还需要借助波函数来描述电子云,相互作用更是要依赖于薛定谔方程等来理解.所以经典物理理论有它的适用性,这一点在下面我们需要记住.

这也就是牛顿经典力学体系的第二大特点了:物质的基本模型一律视作质点(mass point)²,相互作用的基本模型一律视作力(force),质点具有惯性这一基本属性,用质量(mass)来描述.

从牛顿第二定律 $F = ma$ 形式上看,难免会由于数学思维得出“用力来定义质量的多少”或者是“用质量定义力的大小”这样的观点,而进入力与质量循环定义的误区.但实际上谁也没有定义谁.因为正常的物理思维是先要提前定义质量和力:

首先质量是受力物体所固有的属性之一,它具有广延性:如果把质量为 m_1 的质点与质量为 m_2 的质点复合为一个大质点,其质量变为 $m_1 + m_2$.可比性:任意两个质点的质量都可以依照在同样大小的力下的行为来比较其质量的比例关系³.和固有性:只要质点没有变性,质量就不变,所以在不同参考系去看同一个质点,在运动的不同时刻质量都是严格不变的,也称“质量守恒”,我们将避免这一称呼,因为一般的守恒量只是针对一个参考系中的过程中的各个状态,而换一个系看守恒量往往要变,质量却显然不变,这是其一;另外相对论下质量就不再守恒了,它不普遍,这是其二.

力也是对固有的相互作用的强度做的描述.它也是固有的,应当有一种合理的公式去描述施力物体的状态如何影响施加的力(这个力也与受力物体的状态有关).而且如果换一个参考系考虑,力的强度是不变的.根据力对质点产生的效果,力也应当具有矢量性,不同两个力是可以比较的,很多情况下甚至可以叠加:多个施力物体互不影响各自施加的力,同时作用于同一个受力物体,在生活中这种现象非常常见,不难得出这样的总结.

于是才有了牛顿的著名论断:

$$\mathbf{F} = m\mathbf{a}$$

这样一个定律内容是丰富的,它首先是兼容于伽利略相对性原理从而自洽:在另外的惯性系中考察, F 和 a 都是不变的.但是又不是兼容于相对性原理的唯一解决方案,狭义相对论,就提供了另一套可能性,在低速极限 $v \ll c$ 时狭义相对论的动力学方程表达式和这并无二致;其次质量 m 作为系数联系了力(原因)与加速度(结果)这并不是偶然,它也是为了和叠加原理相协调:一份的力作用在一份质点上产生这样的加速度,那么把两份力各自作用在两份质点上那就还是产生同样的加速度.所以力除以质量决定运动学加速度⁴那是天经地义.至于为什么两份力作用在原来的一份质点上会产生加速度加倍的效果,这就是经典力学中的类似于公理性质的存在了.

3.1.4 牛顿第三定律

牛顿第三定律是对“相互作用”中的“相互性”的刻画,暗示了动量与角动量守恒.

任何物理理论,如果妄想在短短几条式子中道尽万千世界发展与变化的规律,都是玄学而不是科学.科学的精神是要基于事实,进行总结与归纳,借助数学进行基本的抽象与再创造.所以现实就是,世界的运行本就是复杂的,不可预测的.物理学的发展史也就充满了惊人的转折:每当人们以为在某一个大的领域有简单而清晰的图像共识的时候,总是会有意外发生告诉人们特殊情况下理论不再适用.原则上,没有任何命题能够成为物理学上的公理:它能无条件适用于从宏观到微观的方方面面.

牛顿清楚地意识到了物质与物质之间相互作用的复杂性,是不可能用几个简单的公式写出所有情况下他的理论中两个任意质点之间的作用力的.所以退而求其次,用牛顿第三定律来描述了所有情形下这样的力需要满足的条件.的确合情合理.但是又看牛顿第三定律的提法本身是否合理?

²牛顿甚至认为光也适用于质点模型

³事实上用碰撞实验可以给出一个更简单的比较质量的方法,让待测质量的质点以一个单位的速度去正碰撞一个以 x 个单位速度运动的单位质量的质点,若恰好停下就说明待测质量为 x 个单位质量,若没停下就改变 x 的值直到停下.

⁴即 $F/m = f(v, a)$

实际上我们先注意到动量守恒和角动量守恒的条件比牛顿第三定律的条件要强(也更本质):我们可以由前者推导出后者.以动量为例,动量是用来描述一个体系的动力学运动趋势的.动量守恒就意味着,如果一个孤立的体系的一部分“平均地”在初始时刻具有向前运动的趋势,比如一部分向前运动而另一部分静止.那么之后就绝对不可能演化为向后运动.或者还可以这么理解:如果看到一个质点在做一个非匀速直线运动:比如忽快忽慢的直线运动,那么显然单独看质点本身动量并不守恒.但是这必然意味着质点不是孤立的,必须和别的体系一同看作一个整体,质点增加或减少的动量必须由体系的另一部分的动量减少或增加来抵消,来满足整体动量守恒的需求.如果体系仅仅由两个质点构成,再把动量的增加或减少的快慢理解为受力(动量定理),我们就有了牛顿第三定律中关于相互作用力等大,反向的部分.

所以动量守恒可以用来“证明”牛顿第三定律中的一对相互作用力.同理角动量守恒则可以用来证明牛顿第三定律中质点间相互作用力共线的部分.最后读者可以验证牛顿第三定律的数学形式为⁵:

$$\mathbf{F}_{12} + \mathbf{F}_{21} = \mathbf{0} \quad , \quad \mathbf{r}_2 \times \mathbf{F}_{12} + \mathbf{r}_1 \times \mathbf{F}_{21} = \mathbf{0}$$

之所以这么理解是因为,动量守恒和角动量守恒显得更普适.事实上,它们是体系物理规律在空间平移对称性和旋转对称性下具有协变形式的直接推论.这一点是由近代著名的[诺尔特定理](#)(Noether Theorem)这一数学定理精确证明了的.

也正是牛顿第三定律暗示了牛顿经典力学体系的缺点.因为我们无法解释牛顿第三定律的“自然性”,取消牛顿第三定律所构成的体系无非是具有外力的非孤立系(动量不守恒),牛顿第三定律为体系增添的“结构性要求”并没有很好地由理论解释与包含.所以才有了后来建立在对守恒量,尤其是能量进行强调的分析力学理论的提出.

3.1.5 质点系与它的牛顿定律

对于复杂体系,牛顿提出一种可能的等效观点,就是把任意这样的体系看做是一种离散的质点系的推广.质点系由质点集合 $\{(m_i, \mathbf{r}_i)\}$, ($i = 1, 2 \dots n$) 组成.我们只要找到从质点理论到有限质点构成的质点系理论中那些不变的结论,就有理由相信这些结论可以推广到复杂的真实体系.因为真实体系被期待可以用一个足够大数目的质点系来近似.对于复杂的质点系首先需要引入[质心](#)(center of mass, CM) 的概念:

$$\overline{\{(m_i, \mathbf{r}_i)\}} = (m_C, \mathbf{r}_C)$$

我们的做法是:直接把质心理解为一个新的质点,它具有新的质量 m_C 和位置 \mathbf{r}_C ,并满足表达式:

$$m_C = \sum_i m_i \quad ; \quad \mathbf{r}_C = \frac{\sum_i m_i \mathbf{r}_i}{\sum_i m_i}$$

可以把上式读作:质心的质量等于质点系的总质量,质心的位矢是各个质点位矢关于质量的加权平均.

质心的求法是符合交换律与结合律的.这在质量上是显然的,因为质心质量就是各个质点质量构成的连加式.对于质心位矢,考虑总质量为 m_1 的第一个质点系 $(m_{1i}, \mathbf{r}_{1i})$ 和总质量为 m_2 的第二个质点系 $(m_{2j}, \mathbf{r}_{2j})$ 复合而成的整个质点系,如果直接算两个体系质心形成的质心:

$$\frac{m_1 \mathbf{r}_{1C} + m_2 \mathbf{r}_{2C}}{m_1 + m_2} = \frac{\sum_i m_{1i} \cdot \frac{\sum_i m_{1i} \mathbf{r}_{1i}}{\sum_i m_{1i}} + \sum_j m_{2j} \cdot \frac{\sum_j m_{2j} \mathbf{r}_{2j}}{\sum_j m_{2j}}}{\sum_i m_{1i} + \sum_j m_{2j}} = \frac{\sum_k m_k \mathbf{r}_k}{\sum_k m_k}$$

可以发现这就是整个体系的质心.这种方法就是“组合法”,比如求几根棍子的质心,可以把棍子替换为在质心的质点再求.由此还可以引申出“割补法”或者“负质量法”,不过就是把上式反过来理解:如果我们

⁵ \mathbf{F}_{12} 表示 1 给 2 的力.

要求第一个体系的质心, 不难验证可以这样求:

$$\mathbf{r}_{1C} = \frac{m\mathbf{r}_C - m_2\mathbf{r}_{2C}}{m - m_2}$$

动态地看待求质心的式子, 还可以发现, 求导以后, 我们得到了质心作为一个在运动的质点, 其速度, 加速度也是可以直接从原质点系来做加权平均的:

$$\mathbf{v}_C = \frac{\sum_i m_i \mathbf{v}_i}{\sum_i m_i} ; \quad \mathbf{a}_C = \frac{\sum_i m_i \mathbf{a}_i}{\sum_i m_i}$$

为了推导出质点系的类似的牛顿定律, 我们发现牛顿三大定律是缺一不可的. 当然原则上如果没有牛顿第三定律, 我们就把作用在每一个质点 m_i 上的合力, 不管施力物体是谁, 记作 ${}^t\mathbf{F}_i$, “t” for “total”. 原则上我们也可以得到:

$${}^t\mathbf{F} = \sum_i m_i \mathbf{a}_i \quad {}^t\mathbf{F} = \sum_i {}^t\mathbf{F}_i$$

这一点很容易证明, 只需要把每个质点的牛顿定律加起来即可. 但是注意到牛顿第三定律既然正确, 我们就可以做进一步的研究: 把任何质点受力分解为内力(internal force) 和外力(external force):

$${}^t\mathbf{F}_i = {}^{in}\mathbf{F}_i + {}^{ex}\mathbf{F}_i$$

内力指 $i = 1, 2 \dots n$ 个质点之间一对一对存在的力, 所以实际上不如写明:

$${}^{in}\mathbf{F}_i = \sum_{j \neq i} \mathbf{F}_{ji}$$

最后单独的那个真正的外力以后直接简写为 ${}^{ex}\mathbf{F}_i = \mathbf{F}_i$.

经过这么一番折腾完以后, 我们还是把所有质点的牛顿定律来求和, 但是注意到由于牛顿第三定律关于相互作用力等大反向的部分, 内力相互抵消, 得到:

$$\mathbf{F} = \sum_i m_i \mathbf{a}_i \quad \mathbf{F} = \sum_i \mathbf{F}_i$$

这就是质点系的牛顿第二定律, 相比上面那个也对的直接把每个牛二加起来的表达式, 能引起注意的变化只有一点, 那就是: 不需要考虑内力! 但正是这一点极大地简化了很多实际问题的分析.

值得一提, 我们顺理成章地要考虑另外两个层面的规律, 一是“对质心”, 二是“相对质心”. 前者值的是, 把质心作为一个质点来研究问题. 那么问题也很简单, 利用之前证明过的:

$$\mathbf{a}_C = \frac{\sum_i m_i \mathbf{a}_i}{\sum_i m_i} \Rightarrow m_C \mathbf{a}_C = \sum_i m_i \mathbf{a}_i$$

就可以得到:

$$\mathbf{F} = m_C \mathbf{a}_C$$

这就是著名的质心运动定律(motion law of CM). 特殊的, 如果合外力 $\mathbf{F} = \mathbf{0}$. 那么退化到质心守恒: 质心做匀速直线运动或者静止.

最后我们提出质心系的概念. 质心系(frame relative to CM) 指: 原点在质心的平动欧几里得坐标系. 结合我们已经学过的和将来要学的所有内容我们把其性质罗列如下, 请读者能够给出证明的直接证明, 给不出的带着问题往下学习:

- 质心系虽然有非惯性力, 但非惯性力与原力系主矢和为零.
- 质心系虽然有非惯性力, 但非惯性力系对质心的主矩为零.
- 质心系是零动量系.
- 质心系虽然有非惯性力, 但非惯性力在过程中总功和为零.
- 取质心为原点能使质点系二阶矩最小, 故刚体相对所有平行轴中以过质心的轴转动惯量最小.
- 只有取质心才能写出柯尼希定理.

3.1.6 非惯性系的处理

虽然牛顿定律只在惯性系中成立.但是如果我们选取一个相对惯性系的转动系,其原点速度加速度为 v_C 与 a_C .而角速度与角加速度分别为 ω 与 β .本来在原来的惯性系中研究一个质点,正确的式子是:

$$\mathbf{F} = m\mathbf{a}$$

现在换了参考系以后,将不能仿照上式写出一个依然正确的式子.但是加速度变换总是对的:

$$\mathbf{a} = \mathbf{a}_C + \boldsymbol{\beta} \times \mathbf{r}' + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}') + 2\boldsymbol{\omega} \times \mathbf{v}' + \mathbf{a}'$$

而我们也永远认为力矢量在参考系变化下是大小与方向都不变的:

$$\mathbf{F} = \mathbf{F}'$$

代入就足以得到依然正确的,仅仅用新的非惯性系中的各个参量表示的类牛顿定律.我们以后将直接叫做非惯性系中的牛顿定律:

$$\mathbf{F}' + (-m\mathbf{a}_C) + (-m\boldsymbol{\beta} \times \mathbf{r}') + [-m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}')] + (-2m\boldsymbol{\omega} \times \mathbf{v}') = m\mathbf{a}'$$

可以发现,这类似于需要为质点运动虚构一些力,它们全都称作**惯性力**(force of inertial).包括:

- 平动惯性力 $-m\mathbf{a}_C$,典型的情况就是平动而非转动的非惯性系中的惯性力.特点是一个恒力,而与质点的运动状态无关.
- 切向惯性力 $-m\boldsymbol{\beta} \times \mathbf{r}'$ 与惯性离心力 $-m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}')$,典型的情况是绕轴做旋转的参考系.特点是这两个力仅仅依赖于质点的位置,离轴越远力越大,在轴上力就消失了.
- 柯里奥利力 $-2m\boldsymbol{\omega} \times \mathbf{v}'$,典型的情况是相对转动系还有运动的情况.特点是给出一个与位置无关,但是正比与速度大小,方向始终垂直于速度方向的力.

如果我们的研究对象是一个刚体而非质点,那么相当于每一个质点上都要受到不同的惯性力.此时如果考虑这些所有力作用在刚体上的等效合力和力矩,那么将是一个非常复杂的情况,尤其是,各个柯里奥利力将对刚体整体造成所谓的“回转力矩”,用这个观点可以解释很多第七章刚体中的现象,感兴趣的读者可以参考相关资料.

3.2 动量定律

这四节我们将罗列质点与质点系动力学的核心内容.读者应当具有初等的牛顿理论的基础.这一部分与三个牛顿定律的关系如何?首先它们是牛顿定律的自然延伸,属于并占据经典力学的主要部分.其次这些结论与相关推论适用性很强.比如三个守恒律:动量能量与角动量的守恒几乎是在所有物理理论中普适的.从内容上看关于质点系的定律自然就包含了单质点的相应特殊情况.而写法总是有三种:相对地面系的,相对质心系的和把质心作为质点的.联系三种写法的是质心系独有的各个柯尼希定理.最后,能量比较特殊我们单独列一小节多加阐述,具体的碰撞问题加一节,最后再补充一个新的位力定律.便构成了一 3×3 条定律为主干的本章的内容.实际上也就是牛顿经典力学的核心部分.

3.2.1 质点的动量

质点的**动量**(momentum) 定义为:

$$\mathbf{p} = m\mathbf{v}$$

很直接地可以推导出质点的动量定理(law of momentum) 来, 用 \mathbf{F} 表示合力:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}$$

证明过程就是把牛顿定律右侧拿来等量代换.

质点动量守恒并不是守恒律最常见的提法. 但是不管怎样, 我们能看出来, 如果在某方向⁶分力为零, 那么该方向分动量导数为零, 从而是一个不变量.

3.2.2 质点系的动量

依据上一节的符号约定, 质点系 $\{m_i, \mathbf{r}_i\}$ 中第 i 个质点的牛顿定律写作:

$$\mathbf{F}_i + \sum_{j \neq i} \mathbf{F}_{ji} = m_i \mathbf{a}_i$$

所谓质点系的动量定理, 无非就是把上式拿去对每一个质点求和, 然后等号右侧化为动量的形式. 过程中我们发现需要定义:

质点系的总动量:

$$\mathbf{P} = \sum_i \mathbf{p}_i = \sum_i m_i \mathbf{v}_i$$

外力矢量和:

$$\mathbf{F} = \sum_i \mathbf{F}_i$$

过程中内力全然抵消, 这就导致了最后表达式简单地为:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}$$

接下来就是研究质心的两个层次, 一是把质心作为质点研究等号左侧的“作用” 和等号右侧的“动力学量”; 二是考虑相对质心系的引入惯性力以后的新写法.

如果在地面系中质点系受到一个力 \mathbf{F}_α , 那么把这个力移动到质心上以后, 我们认为这个力是不需要改变的. 但是如果考虑相对质心系, 固然原来的力也是存在且不变, 但是还得考虑惯性力. 那么总的合力推导出来应当是:

$$\mathbf{F}_C = \mathbf{F}, \quad \mathbf{F}' = \mathbf{0}$$

后者是因为惯性力的合力 $\sum_i -m_i \mathbf{a}_C = \sum_i -m_i \mathbf{a}_i = -\mathbf{F}$. 我们发现有:

$$\mathbf{F} = \mathbf{F}_C + \mathbf{F}'$$

形如上式的式子都可以叫做柯尼希定理(König theorem). 再来看动量, 质心动量就应当是质点系总动量, 而相对质心系动量必然为零:

$$\mathbf{p}_C = \mathbf{p}, \quad \mathbf{p}' = \mathbf{0}$$

前者由于质心的定义很容易看出来, 后者一般被叫做: 质心系是零动量系, 两者之和符合:

$$\mathbf{p} = \mathbf{p}_C + \mathbf{p}'$$

这就不难看出:

$$\begin{aligned} \mathbf{F}_C &= \frac{d\mathbf{p}_C}{dt} \\ \mathbf{F}' &= \frac{d\mathbf{p}'}{dt} \end{aligned}$$

⁶必须是直角坐标! 比如极坐标下 $F_\theta = 0$ 但 p_θ 不守恒, $L = rp_\theta$ 才是对应的守恒量, 个中缘由涉及到分析力学.

最后我们讨论普遍的动量守恒(conservation of momentum).

第一,动量守恒是普遍的,虽然表面上看 $\mathbf{F} = \frac{d\mathbf{p}}{dt}$ 一个体系的动量只要外力不为零就会有非零的导数.但是根据动量守恒的精神,一个体系的动量增必然以另一个体系动量的减为代价,也就是需要注意到既然存在外力 \mathbf{F} ,那就需要考虑力作用的相互性.只要把所有施力物体都归入同一个体系使相互作用封闭.那么等号左侧只能是零,那么等号右侧就会有普遍的动量守恒.

第二,有些场合动量不以质点为载体而是以场为载体.动量依然守恒,毕竟我们将认为场给质点的力与“质点给场的力”构成反作用力符合牛顿第三定律.事实上只要可能涉及到超距作用就一定是要用场来处理的,不用场处理可能导致不精确甚至谬误的结果,这是因为牛顿第三定律都不一定适用.只要想想一个质点位置突然改变,由于相互作用传递速度必然有限导致相隔一定距离的第二个质点受力变化将会延迟,就不能使得这个力等于同一时刻的反作用力,那么两质点的动量和就不再守恒.另一种典型的情况是两个电流元之间的磁场安培力也不符合牛顿第三定律,全都是因为有场的动量未计入计算的缘故.所以这些场合,如果牛顿第三定律“近似”适用,我们才说质点系的动量“近似”守恒.严格的动量守恒设计场论的知识,有兴趣的读者可以进行相关学习.

3.3 角动量定律

角动量的逻辑与上一节是完全一致的.只不过在牛顿定律的数学处理上要多一步罢了.

3.3.1 质点的角动量

选取一个参考系的原点 O 以后便可以写出质点的位矢 \mathbf{r} .那么质点受力对 O 的力矩(moment of force, moment, torque) 定义为:

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

而为质点则定义一个动力学量也就是角动量(angular momentum):

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = \mathbf{r} \times m\mathbf{v}$$

之所以这么做,一种想法是对牛顿定律 $\mathbf{F} = m\mathbf{a}$ 两侧用 \mathbf{r} 叉乘的一种可能的构造思路⁷.这就能给出角动量定律(law of angular momentum):

$$\mathbf{M} = \frac{d\mathbf{L}}{dt}$$

证明过程只消注意到右边莱布尼茨法则出来两项的另一项 $\dot{\mathbf{r}} \times \mathbf{v} = \mathbf{v} \times \mathbf{v} = \mathbf{0}$.

角动量定律有一个值得注意的点,那就是矩(moment)的中心的选取的任意性.角动量(动量的矩)和力矩都因为是 O 点计算的才具有位矢 \mathbf{r} ,这个点叫做矩心(center of moment).而选取不同的矩心的结果就会产生不一样的定律的表达式形式,它们都是对的.但是解题时何以选取?这个复杂的问题我们下一章再做简要介绍.

另一个点是原则上必须选取不动的点作为矩心.选取动点暗含两层意思,一是,可能需要把每一个粒子对动点的动量理解为相对速度对应的动量.二是,可能在求角动量的导数时要求的是跟随动点的角动量的导数.无论哪一点都将使得相应结论不再正确.当然,如果两种可能的更改都没有发生,即,仅仅是这一时刻对这一点列瞬时的角动量定律,下一时刻换下一个点列.那么这就是构成一个时时刻刻都成立的等式,只有在这种意义下角动量定律才可以原封不动地应用.否则需要重新从牛顿定律出发加以推导.

质点的角动量守恒也不是一个典型的结论.但是有一种典型的情形:就是平面上一个有心力作用下的质点运动是角动量总是守恒的.因为此时 $\mathbf{F} \parallel \mathbf{r} \Rightarrow \mathbf{M} = \mathbf{0}$.

⁷Noether 定理才是严谨的构造思维.

3.3.2 质点系的角动量

定义质点系的总角动量:

$$\mathbf{L} = \sum_i \mathbf{L}_i = \sum_i \mathbf{r}_i \times m_i \mathbf{v}_i$$

外力力矩和(不需要包括内力):

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

每一个质点的牛顿定律,现在我们用每一个质点的位矢做叉乘:

$$\mathbf{r}_i \times \mathbf{F}_i + \mathbf{r}_i \times \sum_{j \neq i} \mathbf{F}_{ji} = \mathbf{r}_i \times m_i \mathbf{a}_i$$

然后再来求和,注意到牛顿第三定律中的“共线”性 $\mathbf{r}_i \times \mathbf{F}_{ji} + \mathbf{r}_j \times \mathbf{F}_{ij} = \mathbf{0}$ 让第二项内力也一对一对抵消了.于是得到质点系的角动量定律:

$$\mathbf{M} = \frac{d\mathbf{L}}{dt}$$

同理,考虑把外力移动到质心形成的总力矩和相对质心系的力矩,把从质心出发的各个质点的位矢记作 \mathbf{r}'_i ,就有:

$$\mathbf{M}_C = \mathbf{r}_C \times \mathbf{F} \quad , \quad \mathbf{M}' = \sum_i \mathbf{r}'_i \times \mathbf{F}_i$$

后一个式子隐藏着一个关键信息:质心系中惯性力以质心为矩心时力矩和为零,这可以如下推导:

$$\sum_i \mathbf{r}'_i \times (-m_i \mathbf{a}_C) = - \left(\sum_i m_i \mathbf{r}'_i \right) \times \mathbf{a}_C = -\mathbf{0} \times \mathbf{a}_C$$

而质心的角动量与相对质心的总角动量为:

$$\mathbf{L}_C = \mathbf{r}_C \times m_C \mathbf{v}_C \quad , \quad \mathbf{L}' = \sum_i \mathbf{r}'_i \times m_i \mathbf{v}'_i$$

注意后一式中由于换系,相对质心速度应当为 $\mathbf{v}'_i = \mathbf{v}_i - \mathbf{v}_C$.如此定义以后依然可以符合柯尼希定理:

$$\mathbf{M} = \mathbf{M}_C + \mathbf{M}'$$

$$\mathbf{L} = \mathbf{L}_C + \mathbf{L}'$$

对于后者我们简略写出证明过程:

$$\begin{aligned} \mathbf{L} &= \sum_i \mathbf{r}_i \times m_i \mathbf{v}_i = \sum_i (\mathbf{r}_C + \mathbf{r}'_i) \times m_i (\mathbf{v}_C + \mathbf{v}'_i) \\ &= \mathbf{r}_C \times \left(\sum_i m_i \right) \mathbf{v}_C + \left(\sum_i m_i \mathbf{r}'_i \right) \times \mathbf{v}_C + \mathbf{r}_C \times \left(\sum_i m_i \mathbf{v}'_i \right) + \sum_i \mathbf{r}'_i \times m_i \mathbf{v}'_i \\ &= \mathbf{L}_C + \mathbf{0} + \mathbf{0} + \mathbf{L}' \end{aligned}$$

于是也就有质心的角动量定律和相对质心的角动量定律:

$$\mathbf{M}_C = \frac{d\mathbf{L}_C}{dt}$$

$$\mathbf{M}' = \frac{d\mathbf{L}'}{dt}$$

容易看出角动量守恒(conservation of angular momentum)也是普遍的.而场的角动量也是可能涉及超距作用时需要考虑的因素.

3.4 能量定律

能量的情况就有一些不一样了. 在动能定理的层面上逻辑相似, 但是内力不再可以忽略, 而且动能往往不守恒, 即使守恒也不普遍, 能量普遍地需要守恒, 但这又不能在经典力学框架下加以证明. 纯属特殊情况下的特殊结论或是一种令人信服的归纳假设.

3.4.1 质点的动能

力作用在质点上对质点做功(work) 的功率(power) 定义为:

$$P = \mathbf{F} \cdot \mathbf{v}$$

一个具体的过程才可以说做功, 它等于功率的时间积分, 也可以按照力点乘位移来做积分:

$$W_{1 \rightarrow 2} = \int_{t_1}^{t_2} P dt = \int_{t_1}^{t_2} \mathbf{F} \cdot \mathbf{v} dt = \int_{t_1}^{t_2} \mathbf{F} \cdot d\mathbf{r}$$

质点的动能(kinetic energy) 则定义为:

$$T = \frac{1}{2}mv^2$$

这么定义的出发点也可以是从 $\mathbf{F} = m\mathbf{a}$ 出发, 两变同时点乘 \mathbf{v} 得到所谓的动能定律(law of kinetic energy):

$$P = \frac{dT}{dt}$$

证明过程只需要注意到 $d(\mathbf{v} \cdot \mathbf{v}) = d(v^2) = 2\mathbf{v} \cdot \mathbf{v} = 2v dv$ 为恒等式.

质点动能守恒也不是一个典型的结论. 如果质点受的合力始终垂直于质点速度, 那么便有动能守恒. 比如把质点约束在一个光滑的椭球面上做惯性运动, 或者是让电荷在一个任意的不随时间变化的磁场中运动都属于这种情况, 动能是不变量.

3.4.2 质点系的动能

定义质点系的总动能:

$$T = \sum_i T_i = \sum_i \frac{1}{2}m_i v_i^2$$

但是, 单独求外力的功率和是没有用的, 因为内力也有它的功率, 且不为零. 所以我们写出整个体系的总作用力的功率:

$$P = \sum_i P_i + \sum_{\{ij\}} P_{\{ij\}} = \sum_i \mathbf{F}_i \cdot \mathbf{v}_i + \sum_{\{ij\}} (\mathbf{F}_{ij} \cdot \mathbf{v}_j + \mathbf{F}_{ji} \cdot \mathbf{v}_i)$$

顾名思义, 求和 $\sum_{\{ij\}}$ 意为对每一对力涉及的后面的向进行求和. 每一项必然涉及到两个质点, 选取好

每一对质点以后就可以写出项来而不在乎选取的顺序. 故如果共有 n 个质点的话一共最多会有 $\frac{n(n-1)}{2}$ 对, 事实上既然 ij 交换要求项本身又不会变, 总是有:

$$\sum_{\{ij\}} = \sum_{i=2}^n \sum_{j=1}^{i-1}$$

这样对每一个质点的牛顿定律点乘质点的速度:

$$\mathbf{F}_i \cdot \mathbf{v}_i + \sum_{j \neq i} \mathbf{F}_{ji} \cdot \mathbf{v}_i = m_i \mathbf{a}_i \cdot \mathbf{v}_i$$

同理, 求和可得质点系的动能定律:

$$P = \frac{dT}{dt}$$

同理,考虑外力移动到质心后对质心运动做功的功率和相对质心系的系统所有力的功率和:

$$P_C = \mathbf{F} \cdot \mathbf{v}_C , \quad P' = \sum_i \mathbf{F}_i \cdot \mathbf{v}'_i + \sum_{\{ij\}} (\mathbf{F}_{ij} \cdot \mathbf{v}'_j + \mathbf{F}_{ji} \cdot \mathbf{v}_i)'$$

这里也隐藏着两点,一是作用在质心上的力又只需要考虑外力了,因为内力的和本来就是零.二是质心系中的惯性力又因为做功总是零而不需要写在第二个式子中了,也就是:

$$\sum_i (-m_i \mathbf{a}_C) \cdot \mathbf{v}'_i = - \left(\sum_i m_i \mathbf{v}'_i \right) \cdot \mathbf{a}_C = -\mathbf{0} \cdot \mathbf{a}_C$$

而质心的动能和相对质心的总动能为:

$$T_C = \frac{1}{2} m_C v_C^2 , \quad T' = \sum_i \frac{1}{2} m_i v'_i^2$$

如此定义以后将符合柯尼希定理:

$$P = P_C + P'$$

$$T = T_C + T'$$

证明过程请读者自己完成.同时,也有对质心的动能定律和相对质心的动能定律:

$$P_C = \frac{dT_C}{dt}$$

$$P' = \frac{dT'}{dt}$$

导致能量与角动量和动量不同的原因是,动能守恒绝对不是什么普遍结论,不仅不普遍而且还很罕见.能量才是普遍守恒的.我们细细比较其中区别:

3.4.3 其它能量形式

从表面上看,很容易观察出来,即使整个质点系已经囊括了全部的施力物体了,相互作用是封闭的,也就是每一个质点上都没有外力只有内力:

$$\mathbf{F}_i = \mathbf{0}$$

此时依然总功率可以不等于零:

$$P = \sum_{\{ij\}} P_{\{ij\}} = \frac{dT}{dt} \neq 0$$

从而还和动量,角动量那样的守恒律已经成为了妄谈.

从实质上看,我们对于动量定律角动量定律还有一种说法:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} , \quad \mathbf{M} = \frac{d\mathbf{L}}{dt}$$

力,带来动量的转移;力矩,带来角动量的转移.但是同样的说法换到做功上就不合适了:的确,下式成立:

$$P = \frac{dT}{dt}$$

但是功率不能认为是动能的转移.的确,功率是能量的转移;也的确,功率就是单位时间变成的动能.但是它不一定来自动能.只需要设想两个带正电的小球相互排斥而获得速度,那么显然把其间的力视作相互作用力时,两小球的动能并不是互相转移的关系,而是都因为力做功而获得动能,这个动能来自下面要讲的势能,或者场的能量.

所以能量守恒这一论调,不像另外两个守恒律,至少是不能从牛顿定律中单独推导.但是只要上升到分析力学立刻就可以发现能量守恒是完整体系相互作用不显含时间(稳定体系)的结果.在经典层面上通过补充下面三个独立而正确的说法便可以使得能量守恒可以被证明出来.

保守力

保守力意为由势能函数描述的力. 势能函数是至关重要的概念, 为了确定一个体系的运动, 原则上只需要给出势能函数和初始状态便可以完全求解了. 复杂的体系之所以复杂, 一是粒子数大, 二是初始状态随机. 但是各式各样的整体行为, 往往就蕴藏在任意两个粒子的相互作用的形式中. 而这个相互作用我们相信它总是保守的.

就是说, 存在二体相互作用函数, 也就是**势能**(potential energy) 函数 V , 如果两个粒子分别在 \mathbf{r}_1 和 \mathbf{r}_2 . 势能应当只与位置有关 (与速度加速度无关), 也不能含有时间:

$$V = V(\mathbf{r}_1, \mathbf{r}_2)$$

而势能决定作用力:

$$\mathbf{F}_{12} = -\nabla_2 V, \quad \mathbf{F}_{21} = -\nabla_1 V$$

如果能够找到这样的势能函数, 那么就把这样的力叫做**保守力**(conservative force).

如何理解以上的形式? 它的等价的写法是:

$$F_{12x} = -\frac{\partial V}{\partial x_2}, \quad F_{12y} = -\frac{\partial V}{\partial y_2}, \quad F_{12z} = -\frac{\partial V}{\partial z_2}$$

$$F_{21x} = -\frac{\partial V}{\partial x_1}, \quad F_{21y} = -\frac{\partial V}{\partial y_1}, \quad F_{21z} = -\frac{\partial V}{\partial z_1}$$

$$dV = V(\mathbf{r}_1 + d\mathbf{r}_1, \mathbf{r}_2 + d\mathbf{r}_2) - V(\mathbf{r}_1, \mathbf{r}_2) = -\mathbf{F}_{12} \cdot d\mathbf{r}_2 - \mathbf{F}_{21} \cdot d\mathbf{r}_1$$

那么首先这就提出了一个要求, 由于相互作用力总是应当满足牛顿第三定律的, 而且体系应当也具有对称性, 那么势能函数必然只取决于两个质点的相对位矢, 且与两个质点的相对位矢的方向没有关系, 只能和其大小有关系, 也就是说:

$$V = V(R), \quad R = |\mathbf{r}_1 - \mathbf{r}_2| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

这自然就可以计算出:

$$\mathbf{F}_{12} = -\frac{dV}{dR} \mathbf{e}_{12}, \quad \mathbf{F}_{21} = -\frac{dV}{dR} \mathbf{e}_{21}$$

其中 \mathbf{e}_{12} 和 \mathbf{e}_{21} 分别代表 1 指向 2 和 2 指向 1 的单位矢量.

第二点是, 在任一个元过程中, 这一对相互作用内力做的功的和, 一方面由于动能定律等于转化为 1 和 2 的总动能 dT , 另一方面还等于势能的减少:

$$\mathbf{F}_{12} \cdot d\mathbf{r}_2 + \mathbf{F}_{21} \cdot d\mathbf{r}_1 = -dV$$

这就造成了能量函数 $H = T + V$ 的守恒:

$$dT = -dV \Rightarrow H|_{t=0} = H|_{t=0} = E$$

在更多情况下, 我们让施力物体静止在固定点, 但是施力物体亦是一个质点系而非质点, 研究单个质点作为受力物体的运动. 那么显然此时质点符合 $\mathbf{F} = m\mathbf{a}$. 但是质点依然受到所的合力还是一个保守力, 只需要把它与各个施力物体的势能函数相加, 整体作为这个质点位矢的一元函数即可:

$$\mathbf{F}(\mathbf{r}) = -\nabla V(\mathbf{r})$$

总之, 原则上可以用势能函数描述的力就是保守力. 虽然从中文还是英文的名字来看, 似乎暗示了只要能量守恒就应当视作保守力, 但这是不恰当的. 例如一个不随时间变化的磁场中一个电荷的运动, 显然动能不随时间变化是个常数, 但这绝对不意味着体系的势能函数为 $V(\mathbf{r}) = 0$ 或常数. 的确, 这样能够使得“能量函数” $H = T + V$ 守恒到一个常数 $E = T|_{t=0}$, 但是这样的势能函数不能用来刻画磁场给质点的力:

$$q\mathbf{v} \times \mathbf{B} \neq -\nabla V$$

所以这个力不能用势能来描述,从而并不是保守力⁸.

约束力

从微观相互作用到宏观发生了很复杂的变化.但是很多情况下体系依然存在一个势函数.例如单摆的动能和势能就可以写为:

$$T = \frac{1}{2}ml^2\dot{\theta}^2$$

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

$$H(\theta, \dot{\theta}) = T + V$$

显然,这里用的不再是摆球的位矢 \mathbf{r} 作为势能的坐标.而是换成了 θ 这样的广义坐标(generalized coordinates),对应的广义力 $-\frac{\partial V}{\partial \theta}$ 实际上是一个力矩,把广义力乘广义坐标的微分实际上就是释放的势能.我们发现小球受到的力是重力和绳子的拉力,这个势能函数实际上就是只考虑了重力势能.为什么可以这么做?这是因为绳子的拉力就是一个十分典型的约束力.从守恒律的角度来看,它不会影响其他力造成的动能和势能和的守恒.这一点也十分自然,因为这样的绳只的确没有任何机制能够存储能量.

这也就是说,对约束力(constraint force)的要求主要就体现在它不能做功上.显然,上一例中的绳子拉力并不对摆球做功.容易验证,光滑的面接触,不可伸长的轻绳轻杆连接的两个物体,其约束力对约束的单个物体多整个体系做的总功总会是零.它们都是典型的约束力.更详细的讨论见下一章.我们最后以一个例子来看待对约束力的要求.

如果一个体系的两个广义坐标 q_1, q_2 对应的物体之间有一个约束 $F(q_1, q_2) = 0$,如两个用长为 l 的不可伸长的轻杆连接的质点的 x 坐标就必须满足:

$$F(x_1, x_2) = (x_1 - x_2)^2 - [l^2 - (y_1 - y_2)^2 - (z_1 - z_2)^2] = 0$$

那么在任何过程中两个坐标的变化量 dq_1 和 dq_2 就必须满足:

$$\frac{\partial F}{\partial q_1} dq_1 + \frac{\partial F}{\partial q_2} dq_2 = 0$$

同时作为约束力,对两个物体做功的和必须为零:

$$Q_1 dq_1 + Q_2 dq_2 = 0$$

这就给出了约束力的条件.约束力应当具有形式:

$$Q_1 = \lambda_1 \frac{\partial F}{\partial q_1}, Q_2 = \lambda_2 \frac{\partial F}{\partial q_2} \quad \text{s.t.} \quad \lambda_1 = \lambda_2$$

两个 λ 就代表了约束力的强度.这个式子的一种最简单情况是,同一个不可伸长的轻杆对两侧两质点的作用力必然是等大反向的.总之,不做总功就是约束力的核心特性.

耗散力

从微观到宏观,一个很重要的不同就是对内能(inner energy)的忽视.设想一个静止的小球,用力一掷小球便飞走.在静止时小球内部依然有大量运动的粒子.就纵使设想小球是一个质点系.考虑到内部的势能,加上其总动能便是此时的内能 U :

$$U = V + \sum_i \frac{1}{2}m_i v_i^2$$

⁸其实在分析力学中如果引入磁矢势 \mathbf{A} 那么这个力可以作为 $U = -q\mathbf{v} \cdot \mathbf{A}$ 的负梯度,这中含有速度的势能叫做广义势函数.但是即使可以定义势函数,习惯上依然不把这个力叫做保守力.因为可以验证此时动能与这个势能的和又不再守恒了.

我们还要求这样的质点系不能具有动量,也就是目前就在这个质点系的质心系(其实是瞬时与质心共速的惯性系)中,才能认为这个小球是“静止”的:

$$\mathbf{0} = \sum_i m_i \mathbf{v}_i$$

但是如果小球如果以速度 \mathbf{u} 被抛出去,当作质点 $m = \sum_i m_i$ 其动量和动能自然是:

$$\mathbf{p} = m\mathbf{u}, T = \frac{1}{2}m\mathbf{u}^2, \quad T = \frac{p^2}{2m}$$

即,动能是由于动量而产生的能量.这种说法是贴切的.因为柯尼希定理足以告诉我们,作为质点的动量就是抛出小球作为质点系的总动量 $\sum_i m_i(\mathbf{v}_i + \mathbf{u}_i) = m\mathbf{u}$.而抛出小球的总能量,势能部分是不变的,再把动能部分用柯尼希定理:

$$H = V + \sum_i \frac{1}{2}m_i v_i^2 + \frac{1}{2}m\mathbf{u}^2 = U + T$$

从而动能是由于运动而具有的能量这一点就不言而喻了.如果速度 $u = 0$ 那么体系就只有内能.

正是基于这一点,我们考虑一个宏观“质点”系的能量时,“质点”并不是完全理想的模型,仅仅是足够小的质点系的一种“打包”.那么体系的总的的能量就可以由“质点”的动能,“质点”间相互作用的势能,以及“质点”的内能三部分协同构成.前两部分是宏观的能量,习惯上还是合称为“机械能”.后一部分则是微观形式的能量.而耗散力是这样一种力,在认可微观过程总是能量守恒的前提下,就只剩下能量转移的方向了.根据热力学第二定律,自发进行的过程具有普遍的特性:总是趋向于宏观形式的能量减少,无规则热运动形式的能量增加.那么耗散力(dissipative force)就是期间起到搬运能量作用的那个力.也就是说,对于质点系耗散力总是做负功,使得体系动能减少而转变为内能(也称发热):

$$|W| = -W = \Delta U$$

而体系动能增加还是减少还依赖于两点,一是势能的释放可能会补偿由于耗散减少的动能.二是,我们一直默认体系封闭没有外力.但是如果有外力,外力做功也会使得体系的动能发生改变.

综上所述,对一个宏观的质点系,往往其能量转化与守恒写为以下这个形式:

$$\sum_i \mathbf{F}_i \cdot \mathbf{v}_i - \sum_{\{ij\}} ({}^d\mathbf{F}_{ij} \cdot \mathbf{v}_j + {}^d\mathbf{F}_{ji} \cdot \mathbf{v}_i) = \frac{d}{dt}(T + V)$$

左侧第一项就是外力做功输入的能量的功率.第二项是内力中的耗散力(已用左上角标 d 表示了耗散)向内能转化的能量的功率.右边就是因为这两个原因而增加的机械能的功率.这就是普遍形式的功能原理(work-energy principle).它仅仅是换着法子写了动能定律而已.但是这种写法是为了用能量守恒(conservation of energy)的角度看问题.

3.5 位力定律 *

位力定律从来都在动力学理论的大框架的边缘.主要原因可能是因为并不对应任何有意义的守恒律.但在有的场合却能够发挥奇效.

3.5.1 质点的位力

不同于力矩,位力(virial),顾名思义就是位矢点乘力,而不是叉乘:

$$\mathcal{V} = \mathbf{r} \cdot \mathbf{F}$$

而我们也为质点运动新定义一个动力学量,也把角动量的叉积变为点积,叫做“G”:

$$G = \mathbf{r} \cdot \mathbf{p}$$

可以发现 G 本身是 $I = \frac{1}{2}mr^2$ 的导数:

$$\frac{d}{dt} \left(\frac{1}{2}mr \cdot r \right) = r \cdot mv$$

但是仿照角动量定理写的以下式子并不成立:

$$\mathcal{V} = \frac{dG}{dt}$$

事实上:

$$\frac{dG}{dt} = \frac{1}{2}m \frac{d^2}{dt^2} (r \cdot r) = \frac{1}{2}m(r \cdot \ddot{r} + 2\dot{r} \cdot \dot{r} + \ddot{r} \cdot r) = r \cdot m\mathbf{a} + 2 \cdot \frac{1}{2}mv^2 = \mathcal{V} + 2T$$

或者说, 位力为粒子的“ G ”的变化率与瞬时动能两倍的差, 这构成位力定律(law of virial):

$$\mathcal{V} = \frac{dG}{dt} - 2T$$

有什么场合作用呢?一个典型的场合就是爱因斯坦对布朗运动的研究. 我们建立一个这样的模型不妨设想花粉粒子受到一个随时间随机变化的力 $\mathbf{f}(t)$. 一个可以通过给一个恒力测出定向漂移运动而测出来的, 正比于速度大小的阻尼力 $-\gamma v$. 这也就是说位力为:

$$\mathcal{V} = \mathbf{r} \cdot (\mathbf{f}(t) - \gamma \mathbf{v}) = \mathbf{r} \cdot \mathbf{f}(t) - \frac{\gamma}{2} \frac{d}{dt} r^2$$

带入位力定律并在很长的一段时间 τ 内对两边进行积分:

$$\int_0^\tau \mathbf{r} \cdot \mathbf{f}(t) dt - \frac{\gamma}{2} r^2 \Big|_0^\tau = G \Big|_0^\tau - 2 \int_0^\tau T dt$$

最后如果再在两边同时处以趋于无限大的 τ . 这对于第一项和第四项这就相当于对积分的内容求过程中的平均, 前者显然为零, 因为 $\mathbf{f}(t)$ 只依赖于时间且完全随机, 平均值就是零. 后者需要注意到动能是恒大于零的, 不管如何取平均总不会为零, 事实上热学上考虑这应当等于花粉与容器中的分子构成热力学平衡时的平均动能 $\frac{3}{2}kT$. 从而不除 τ 这一项就应该写为 $-3kT\tau$. 第三项的 G 也总是有限的, 除以无限长的 τ 自然趋于零. 最后第二项不妨让初始时刻的 $r(0) = 0$. 最后得到:

$$r^2(\tau) = \frac{6kT}{\gamma} \tau$$

这就和三维扩散方程 $r^2(t) = 3x^2(t) = 6Dt$ 不谋而合. 我们发现描述分子力涨落导致无规行走的扩散系数 D 与描述耗散的阻尼系数 γ 之间具有如下的爱因斯坦-斯莫卢霍夫斯基关系(Einstein-Smoluchowski relation):

$$D = \frac{kT}{\gamma}$$

这个式子是更普遍的涨落-耗散定律(law of fluctuation-dissipation) 的原型. 这个定律有力说明了耗散力的本质与热力学涨落的统一性.

3.5.2 质点系的位力

同理, 把位力定律进行多质点的求和, 完全类似于角动量定律的情况, 但是又类似动能定律推导内力没能抵消, 我们得到:

$$\mathcal{V} = \sum_i \mathbf{r}_i \cdot \mathbf{F}_i + \sum_{\{ij\}} (\mathbf{r}_j \cdot \mathbf{F}_{ij} + \mathbf{r}_i \cdot \mathbf{F}_{ji}) = \frac{dG}{dt} - 2T$$

其中“ G ”自然指质点系的“ G ”的和, 动能 T 也是总的动量. 第二项中求和的每一项由于牛顿第三定律, 可以进一步化简为:

$$\mathbf{r}_j \cdot \mathbf{F}_{ij} + \mathbf{r}_i \cdot \mathbf{F}_{ji} = -\mathbf{r}_{ij} \cdot \mathbf{F}_{ij}$$

一般,上面这个式子用来处理稳定的,束缚态的体系之间的量的平均值.在这个意义上,等好右侧的 $\frac{dG}{dt}$ 就平均为零.实际上 G 作为 I 的导数本身就已经平均为零呢.例如我们可以利用这个式子来证明理想气体的物态方程,此时就几乎有:

$$I = \sum_i \frac{1}{2} m_i r_i^2 = \int \frac{1}{2} \rho r^2 dV = \text{Const.}$$

毕竟宏观气体的热平衡下密度分布是几乎均匀的.而最后一项,作为平动动能的平均就应当是:

$$T = \frac{3}{2} N k T = \frac{3}{2} \nu R T$$

最后左侧第二项的内力在理想气体中也认为不存在.就只剩下体系受到的外力项了.而外力实际上就是在气体容器上气体压强对容器压力的反作用力:

$$\sum_i \mathbf{r}_i \cdot \mathbf{F}_i = \oint \mathbf{r} \cdot (-p d\mathbf{S})$$

其中 p 又是个常数.最后这个积分可以从散度定理出发证明,或者注意到 $\frac{1}{3} \mathbf{r} \cdot d\mathbf{S}$ 就是从原点出发连接面元 \mathbf{S} 构成的各个圆锥的体积.这给出了:

$$\sum_i \mathbf{r}_i \cdot \mathbf{F}_i = -3pV$$

最后代入以上结论就得到了:

$$pV = \nu R T$$

这就是用位力定理得到理想气体物态方程的简单思路.历史上著名的昂内斯方程(Onnes equation),就是用位力展开给出了有内部相互作用时的物态方程的形式:

$$p = RT \left(\frac{1}{v} + \frac{B_2}{v^2} + \frac{B_3}{v^3} + \dots \right)$$

其中 v 为摩尔体积.而 $B_2, B_3 \dots$ 称为第二,第三……位力系数⁹.可以用经典的位力定理来推证这个式子,但其实统计物理采取的相互作用图像才是对这个式子的最佳诠释.

3.6 碰撞问题

本小节不讨论连续体与可能带来的动量传输问题.仅仅限于质点,刚体及在理想约束下发生的碰撞过程的讨论.

3.6.1 二质点正碰

碰撞(collision),表面上看很复杂,理论上它被抽象一个持续时间无限短,作用力无限大的理想过程.却总是要伴随或者不伴随着能量损失.这是因为它所对应的实际发生的过程总是一个复杂的多体问题:它涉及到物质的弹性,往下涉及晶格的结构,往上涉及材料的特性.但是和你的模型总是让问题变得简单,这使得我们可以从现象出发建立唯象理论.

一维问题中,质量为 m_1 的质点以速度 v_1 取碰撞一个质量为 m_2 的速度为 v_2 的质点.关心的是碰撞后的末速度 v'_1, v'_2 .传统的观点似乎总是以守恒律为重:我们不出意外地列出:

$$m_1 v'_1 + m_2 v'_2 = m_1 v_1 + m_2 v_2$$

$$\frac{1}{2} m_1 v'^2_1 + \frac{1}{2} m_2 v'^2_2 = \frac{1}{2} m_1 v^2_1 + \frac{1}{2} m_2 v^2_2$$

⁹ 已经默认第一位力系数 $B_1 = 1$

似乎这样就能求出一个结果来. 但这个讨论是非常不完全的. 首先结论并不对. 事实上碰撞的结果本就是不确定的, 其变化的范围一般从完全非弹性(completely inelastic) 到弹性(elastic) 的都有可能, 必须得取决于碰撞物材料的弹性. 以上仅仅是一个弹性碰撞的解. 其次方法也不妥, 这是因为用守恒律固然方便, 但这就等于忽视了相互作用发生的细节. 碰撞的过程彻底变成了一个“黑箱子”, 失掉了很多本来可以获得的结论.

所以对于碰撞的过程, 最佳选择是设出过程中相互作用的冲量来. 但是相比把力设出来, 设出冲量是否就意味着没有公式来计算过程中碰撞力对两个物体做的功? 我们先指出, 碰撞过程的做功本身也是一个有歧义的概念. 设想一个人站在光滑的水平地面上, 具有初速度向墙撞去, 但是推了一把墙使得自己停了下来, 那么墙对人的力是否做功? 这个问题回答“是”和“否”都是有道理的, 因为表述的物理概念不同: 如果是考虑把人视作质点系, 那么墙对人的力作用在手掌上, 力作用过程中手掌始终与墙面接触而没有做功, 这就给出了“否”的结果, 还能说明人减少的动能实际上最终并不会传递给墙面, 就好像其逆过程: 人推墙把自己推开人获得的动能也应当来自于人自身生物能而不是墙面. 但是如果把人视作质点, 那这就是一个人与墙之间的完全非弹性碰撞, 根据动能定理墙对人做的功, 实际上, 是对人质心做的功, 就等于人获得的质心动能. 这样也可以给出“是”的结果.

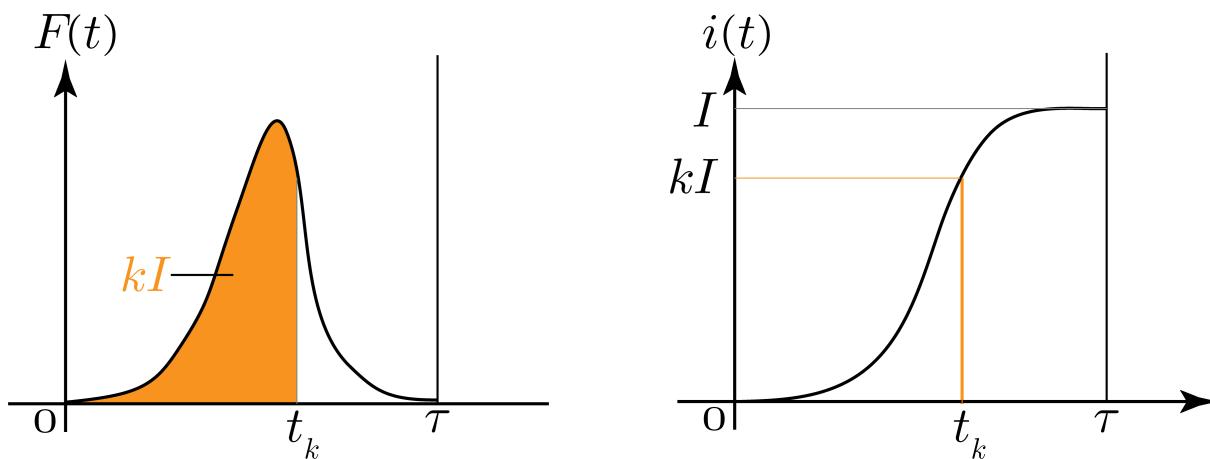


图 3.1: 力与累积的冲量

虽然这个问题倒是也分析地清, 但这足以告诉我们, 选取一个简单清晰的模型的必要性. 为此我们进一步设出碰撞过程中两质点之间持续作用的力 $F(t)$, 它从 0 时刻开始产生到 τ 时刻作用结束. 但是, 并不是只有两质点直接接触才发生作用力, 事实上我们可以人为把这两个力加在相距较远的两个质点上, 用这样的模型去等效本来可能复杂的碰撞过程.

我们再把过程中累积的冲量定义为一个函数:

$$i(t) = \int_0^t F(t') dt'$$

并且把最终造成的总作用冲量记作 I :

$$I = \int_0^\tau F(t') dt'$$

那就必然有一个时刻冲量累积到 kI , 可以把这个时刻记作 t_k , 即:

$$i(t_k) = kI$$

事实上由于 i 是严格单调的增函数, 就存在反函数 i^{-1} , 那么这个时间就是:

$$t_k = i^{-1}(kI)$$

有 $i^{-1}(0) = 0$, $i^{-1}(I) = \tau$. 这样就足够我们去计算过程中相互作用力的做功了. 由于是真实碰撞的模拟, 那么必然 $v_1 \geq v_2$ 碰撞才可以发生. 当冲量累积到 i 时, 两个质点的速度分别为:

$$u_1 = v_1 - \frac{i}{m_1}$$

$$u_2 = v_2 + \frac{i}{m_2}$$

这说明过程 1 速度不断减小, 2 速度不断增加. 现在就足够写出力对 1 和 2 做功的功率了. 分别为:

$$P_1 = -Fu_1 = -\frac{di}{dt} \left(v_1 - \frac{i}{m_1} \right)$$

$$P_2 = Fu_2 = \frac{di}{dt} \left(v_2 + \frac{i}{m_2} \right)$$

最后如果把功率对时间积分, 我们发现积分自动以 i 为自变量了:

$$W_1 = \int_0^\tau P_1(t) dt = - \int_0^I \left(v_1 - \frac{i}{m_1} \right) di$$

$$W_2 = \int_0^\tau P_2(t) dt = \int_0^I \left(v_2 + \frac{i}{m_2} \right) di$$

线性函数的积分, 就等于函数值初末值的平均值乘以积分区间的长度, 这就给出了一个冲量乘速度形式的功, 碰撞过程中相互作用力对质点 1 和质点 2 的功应当是:

$$W_1 = -\frac{1}{2} I(v_1 + v'_1)$$

$$W_2 = \frac{1}{2} I(v_2 + v'_2)$$

这并不是任何让人意外的结论, 如果带入动量定理 $I = m_1(v_1 - v'_1) = m_2(v'_2 - v_2)$ 就能发现这两个功无非就是末动能减初动能, 其实就是反映了动能定律. 不过注意到表达式推导过程中对动量定律的应用我们发现这个表达式倒不是纯粹的任意情况下都通用的功的定义. 在使用动量定律前的这个功的积分:

$$W = \int \mathbf{v} \cdot d\mathbf{I}$$

倒是的确通用的.

如果定义碰前接触速度 $v = v_1 - v_2$, 碰后分离速度 $v' = v'_2 - v'_1$. 我们可以发现:

$$W_1 + W_2 = \frac{1}{2} I(v' - v)$$

又定义恢复系数 $e = v'/v$. 碰撞就因此分为四类:

1. 完全非弹性碰撞, I 最小的情况, 作为一个实际的碰撞过程分离速度必然也至少为零 $v'_2 \geq v'_1$, 否则 1 仍然在接近 2 力就不可能会消失. 所以 I 最小就是让碰撞后共速, 这样 $e = 0$ 就是完全非弹性碰撞的决定 I 大小的条件.

2. 完全弹性碰撞. 考虑碰撞过程的能量转化, 显然, 如果 $W_1 + W_2$ 不等于零, 前后的总动能就不会是不变量. 但是如果要求碰前碰后动能不变, 这就构成了完全弹性碰撞, 简称弹性碰撞. 此时, 除了可以列动能不变来求解 I , 我们发现这等价于 $v = v'$, 也就是接触速度等于分离速度. 以后我们将总是认为弹性碰撞的等价对应条件是 $e = 1$.

3. 非弹性碰撞 (不完全的). 这是最符合实际情况的一类碰撞. 这种碰撞中 $0 < e < 1$. 也就是 $v' < v$ 的情况. 功 $W_1 + W_2$ 是小于零的, 说明碰撞过程动能有损失, 能量去向是显而易见的: 变成了两个碰撞的物体的内能, 这里的内能可能就是热学上的内能, 也可以是视作弹性体时虽然整体动量为零但局部有弹性波而产生的宏观动能与势能. 这种情况之所以普遍是因为, 碰前往往整个物体没有这样的弹性波, 能量处于最低的状态, 碰后这个能量总是会大于零. 故能量总是有损失的.

4. 超弹性碰撞. 也就是 $e > 1$. 碰撞后分离速度大于接触速度. 因碰撞而获得动能. 这样的过程也并不少见, 炸弹的爆炸, 或者刚刚的人推墙使人动起来都是这种情形. 特点是别的能量会释放出来转化为动能. 前提需要注意不能违背动量守恒.

3.6.2 自由刚体的碰撞

如果把两个质点的碰撞换为两个自由刚体相碰撞, 这里我们只考虑平面情况, 那么同理碰撞会产生一个冲量 \mathbf{I} . 但是这个冲量却有两种可能性, 如果是光滑的刚体, 那么碰撞的冲量必然沿法线 n 方向. 但是如果粗糙的刚体, 那么还可以有沿切线方向的分量. 不管何种情况我们来证明两个恒成立的结论:

一是, 1 对 2 做的功或 2 对 1 做的功总是满足:

$$W = \frac{1}{2} \mathbf{I} \cdot (\mathbf{v} + \mathbf{v}')$$

这里 \mathbf{v} 和 \mathbf{v}' 指的便是力的作用点, 即固连在刚体上的点的速度. 证明这个定理思路类似于之前的做法. 我们可以为过程中的冲量累计选取一个变量 $k(t)$. 当 $k = 0$ 时碰撞刚刚开始, 而 $k = 1$ 时碰撞就结束了. 变量取 k 时作用的冲量为:

$$\mathbf{i}(t) = k(t)\mathbf{I}$$

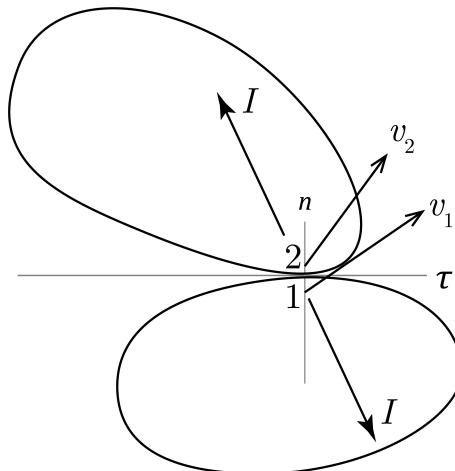


图 3.2: 两刚体碰撞

那么由于冲量对刚体质心速度和刚体旋转角速度的影响都是线性的, 而且碰撞点速度也将线性依赖于这些值. 从而积累到变量 k 时, 点具有速度:

$$\mathbf{u} = k\mathbf{v}' + (1 - k)\mathbf{v}$$

于是利用做功的积分式:

$$W = \int_0^\tau \mathbf{F} \cdot \mathbf{u} dt = \int_0^1 \mathbf{u} \cdot d\mathbf{i} = \int_0^1 [k\mathbf{v}' + (1 - k)\mathbf{v}] \cdot \mathbf{I} dk = \frac{1}{2} \mathbf{I} \cdot (\mathbf{v} + \mathbf{v}')$$

二是, 不损失机械能的条件是, 或者更特殊地, 在无摩擦的碰撞中弹性碰撞的条件是, 接触点形成的沿冲量方向的接触分速度等于该方向的分离分速度. 这个证明也是简单的. 只需要把点乘式后面的速度分解在冲量方向并把两个功相加为零即可. 也就是我们有广义的 $e = 1$ 与动能不变作为弹性碰撞的等价条件.

3.6.3 多体碰撞

如果碰撞是多体的往往有两种不同的处理思路: 一是屡次碰撞的思路, 最典型的碰撞就是牛顿摆情况, 小球与多个连续放置的小球的碰撞如果视作屡次的弹性碰撞就能解释体系形成周期性运动的原因. 因为两个质量相等的小球的碰撞总是简单的交换速度. 此时我们注意到碰撞结束后, 最左侧的两球之间从接触速度的 v 变成了 0, 而最右侧的两小球却从 0 变成了 v :

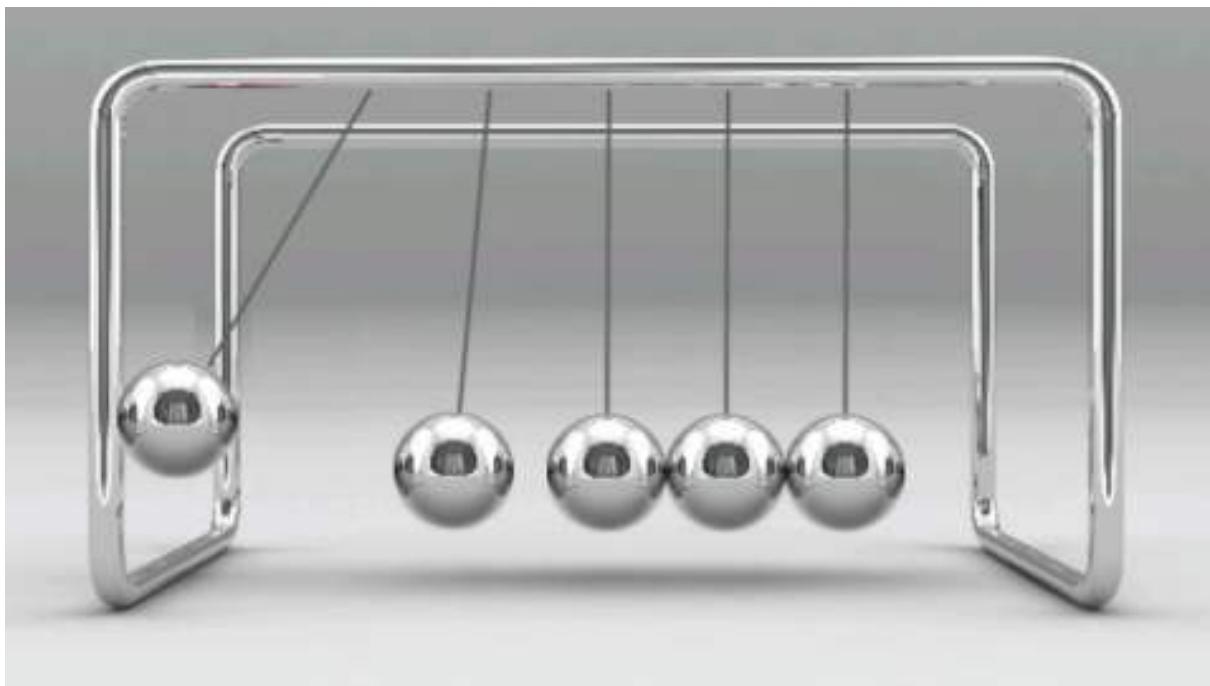


图 3.3: 牛顿摆

但是如果把最右边的小球要撞击的四个小球视作一个带约束的整体,显然既然四个小球必须作为一个整体运动,它们之间的相互作用冲量是不会做功的.同样的道理也适用于任何带约束的多体碰撞场合,虽然碰撞涉及多个物体,但是真正具有接触速度的只有一个碰撞点,那么作为约束其它点处不可能产生分离速度.从而动能不损失的条件依然是在撞击的点处法向接触速度等于分离速度.

不难理解,实际的多体碰撞总是介于两种极端情形之间的.如果是无限多个无限小的球无限密排,这就变成了一个连续的弹性介质中的波动的解的问题,波的传播速度和碰撞速度之间的大小关系就成为了一个具有决定性意义的因素.容易分析,思路一适用与撞击速度很快四个小球还来不及达到平衡的情况.而思路二适用于波的传播速度非常快,任何冲量作用在四个小球上都相当于在推动四个球构成的整体的情况.

第四章 静力学

以运动学和动力学理论为基础,时不时地,我们发现一些典型的现象,包括:

- 不管体系多么复杂,能量函数对其行为似乎起决定性的作用.
- 似乎总是能够从不同的体系中抽象出它的一个特征数字:自由度.
- 约束越多样体系约复杂,它的未知约束力越来越多,但是多到一定程度体系只剩下一个自由度了反而用能量守恒就能够解答大多数问题.

这些现象无疑是紧密联系的,值得研究的.事实上我们要做的就是以之前的所有动力学定律为依据进一步展开讨论.再从头开始建立新的理论:先讨论运动的描述,再单独研究力的特性.最后合到一起,看看这能让我们得到什么.

新理论,新思想,新图像.在本章节的学习过程中,如果感到疑惑,大抵都是复杂的数学推导掩盖了背后的物理目的.希望读者先记住两点:一是,我们要做的是把以前对质点,刚体和对质点系列的选取自然坐标(大抵是直角坐标)而列的部分定理修改为改用广义坐标来描述,过程中的所有步骤都是为这个目的来服务的.二是,本着“能量函数的形式”就决定“体系的一切结构与演化”的观点,我们就能找到一个明确的方向.无论是从结果上还是从细节上,我们将会多次参考这种想法.

4.1 约束

所谓**约束**(constraint)就是对运动的限制.有时,物体发生直接的接触,或是一个物体在另一个物体表面滑动或是滚动,这种情况约束实际上就发生了;有时,约束通过往往是轻质¹的绳,直杆或曲杆,套筒,无摩擦的铰链等等去连接两个乃至多个物体,这样那些物体之间实际上也存在约束,偶尔我们也会把连接它们的机构直接叫做约束.

之所以把它们叫做约束,是因为它们产生了两个共同的结果,一是**约束方程**(equation of constraint),二是**约束力**(force of constraint).根据前者的运动学效果²,我们把与约束分为以下类别:

4.1.1 约束分类

无论是挂在天花板下的小球,还是“鹰击长空,鱼翔浅底”,都给我们以**单边约束**(unilateral constraint)的图像.

¹后面就能理解,这样才不会占用新的广义坐标,因为它不带来任何能量

²根据后者的动力学效果还可以分为理想和非理想约束,见后面小节平衡问题:虚功原理

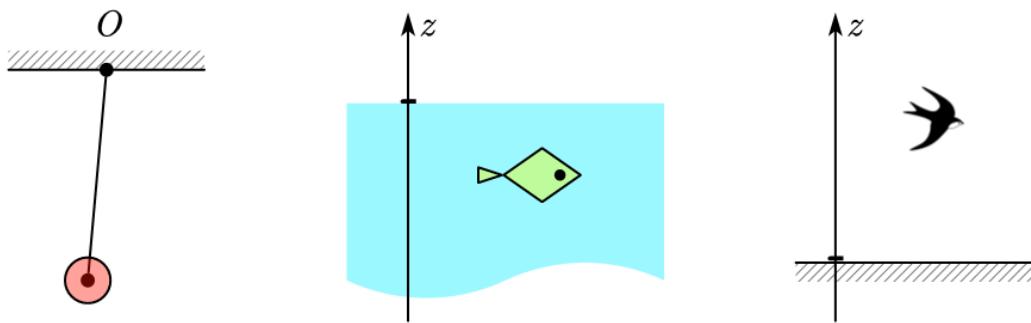


图 4.1: 可解约束

其中, 小球的位置符合 $x^2 + y^2 + z^2 \leq l^2$, 鱼的位置符合 $z \leq 0$, 鸟则是 $z \geq 0$.

更简单的例子还包括两个物体表面发生接触(滑动或滚动)的情况. 此时两物可能会脱离接触, 比如在半球面顶点上滑下来的质点. 这些所有约束的典型特点, 从约束方程上看都是一个单边不等式:

$$F \geq 0 \quad \text{or} \quad F \leq 0$$

如果不等式取等号, 那么可以认为对物体运动的限制客观存在, 从而会相伴相生约束力. 但是如果运动到不等号成立了, 比如上面三种情形中, 绳子松弛了, 鱼没有在表面游鸟也没有在地面走, 那么其实原则上也不存在任何对运动的限制. 称作约束被“解除”了. 所以单边约束也可以称作可解约束. 显然为了简单起见, 我们可以把这一类问题分解为严格符合约束和没有约束两种情形来等效. 这实际上构成了不可解的**双边约束**(bilateral constraint) 模型.

所谓双边约束就是把之前的单边的约束的不等式进行等式化, 比如 $x^2 + y^2 + z^2 = l^2$ 就对应着把可弯曲的轻绳替换为了不可弯曲的轻杆. 而后两个问题就变成了 $z = 0$, 即不能让物体陷入也不能让物体脱离的平面约束.

故, 针对这一类分类模式, 双边约束显然才是“有效”的部分. 故本书以后所有约束都指双边的.

继续根据约束的不同特点分类, 我们就产生了两种可行的依据与分法:

一是, 根据约束的可积性来分为**完整约束**(holonomic constraint) 和**非完整约束**(nonholonomic constraint).

我们发现, 一个体系的运动由若干**广义坐标**(general coordinates) 和它们的导数—**广义速度**(general velocity) 来描述. 比如一个转盘就可以由转角 φ 来描述其处在的几何位置³, 而角速度 $\dot{\varphi}$ 来描述转动快慢. 从而约束, 作为对运动的限制, 也就是应当关于这些量的方程.

但是无可辩驳的是, 在约束发生时这一类方程, 至少在最初被确定下来的形式下分为两大类, 一是不含广义速度的**几何约束**(geometric constraint), 二是含广义速度的**微分约束**(differential constraint).

举三个例子:

A. 一个圆盘状的冰球本来由盘心的 (x, y, z) 和盘面方向, 即垂直盘面的单位法向量 $(\cos \alpha, \cos \beta, \cos \gamma)$ 构成其广义坐标, 后面的描述方法用到了方向余弦. 即使盘并没有被地面约束, 三个方向余弦也要满足:

$$\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1$$

现在盘到了地面(光滑冰面)上, 更是要满足:

$$z = 0 \quad , \quad \alpha = \beta = \frac{\varphi}{2} \quad , \quad \gamma = 0$$

这些都是几何约束.

³注意, φ 与 $\varphi + 2\pi$ 描述的位置在几何上没有区别, 仅仅是在运动的不同过程中发生, 一般认为达到后一个位置就是回到了初始位置(运动经历一个周期).

B. 一个半径为 R 的轮子沿 x 方向单纯地在地面做纯滚动. 本来在没有地面之前, 轮子由水平竖直坐标 (x, z) 以及自转角 φ 描述. 但是, 在有了地面做约束以后, 就需要满足接触条件和纯滚条件 (第一章学过):

$$z = 0 \quad , \quad \dot{x} = R\dot{\varphi}$$

第一个约束是几何约束, 第二个便是微分约束.

C. 考虑一个踩着独轮车在地面上运动的小丑, 并忽略运动的一些细节, 认为小丑仅仅由其水平面上坐标 (x, y) 和面朝方向 θ 描述. 且要求轮子在地面上做纯滚动, 轮子滚动方向就是小丑面朝方向. 描述轮子还需已知自转角 φ , 轮子半径依然为 R . 这些就是描述这个体系的所有广义坐标. 那么客观地看这里有两个限制关系:

$$\frac{\dot{y}}{\dot{x}} = \frac{\sin \theta}{\cos \theta} \quad , \quad \sqrt{\dot{x}^2 + \dot{y}^2} = R\dot{\varphi}$$

两个都是实打实的微分约束.

之所以称作微分约束, 往往是因为这些约束一般都可以对广义速度齐次化, 尤其是可以一次齐次化, B 中的例子已经是关于广义速度 $\dot{x}, \dot{\varphi}$ 的一次线性组合等式了, C 的例子也可以化为:

$$\dot{x} = R \cos \theta \dot{\varphi} \quad , \quad \dot{y} = R \sin \theta \dot{\varphi}$$

这样我们就可以在等号左右同时乘以 dt , 写成对各个变量变化微元之间的“约束”或“限制”:

$$\text{Case B : } dx = R d\varphi$$

图 4.2: 独轮车

$$\text{Case C : } dx = R \cos \theta d\varphi \quad ; \quad dy = R \sin \theta d\varphi$$

这样就避免了广义速度的引入, 依然是广义坐标之间的制约关系, 只不过各个量不全是坐标本身, 还含有当下状态到下一个状态间坐标的微分.

那么约束的完整性也以上两种约束是什么关系? 我们需要注意, 在 B 中的那个约束其实是可以积分的. 约束方程的成立就意味着:

$$x - R\varphi = x(0) - R\varphi(0) = \text{Const.}$$

这样岂不就退化为几何约束了吗? 的确, 像这样的微分约束与一个几何约束“等效”. 从而我们把几何约束和可以用微分约束等效出来的几何约束一同称作完整约束. 如果一个体系内所有约束都是完整约束, 那么这样的体系称作**完整体系**(holonomic system). 后面我们就会发现, 分析力学中的很多主要结论都有体系必须是完整体系的要求.

反之, C 就不是一个完整体系了, 因为体系含有不可积分的微分约束. 任何看出这一点来的? 设想通过某种积分的方式我们得到了体系的一个几何约束:

$$F(x, y, \theta, \varphi) = 0$$

这蕴含的结果马上就会招致一个矛盾. 对于几何约束, 例如在 B 中的后一个化为的 $x - R\varphi = \text{Const.}$, 总是有经历一个复杂的过程如果 x 复原了, φ 也会随之复原. 这里也一样, 如果假定以上几何约束存在, F 至少“真正”含有 (x, y, θ, φ) 四个变量之一, 不妨设是 φ ⁴. 那么如果在一个过程中 x, y, θ 均复原, 那么前后都满足约束方程的话 φ 一般就只有有限个解, 更特殊地, 必然要求也回到初态发生复原. 然而, 联系一下实际的物理情景便可以明白, 小丑完全有能力让自己的位置回到初始位置, 面朝方向也归位, 同时轮子转过一个任意大小的角度的: 只需要沿着半径或大或小的圆周转圈即可. 从这一点上我们就可以发现这样的约束是不可能积分化为几何约束的非完整约束.



⁴其它可能性请读者自己琢磨.

进一步观察我们还能发现,在 C 问题中 (x, y, θ, φ) 四个变量具有某种“全局独立性”,我们总是有办法把任意的两个 $(x_1, y_1, \theta_1, \varphi_1), (x_2, y_2, \theta_2, \varphi_2)$ 作为初末态用过程进行连接.但是我们决不要因此而认为这几个变量之间因此就彼此独立不会相互制约,因为约束方程客观存在,也就是说,考虑任一个状态下一时刻发生的位移 $(dx, dy, d\theta, d\varphi)$,它们在“局域”看来并不具有独立性而是有相关性.这种局部处处成立(locally holds everywhere) 但全局不生效(globally invalid) 的行为是出于较深刻的拓扑学和分析学的原因⁵.需要加以留心.

第二种分类方式则更加清晰明了.它取决于约束方程本身含不含时.例如在之前的绳端系球,以及在上面的 C 中的一约束方程中:

$$x^2 + y^2 + z^2 = l^2 ; \quad dx = R \cos \theta d\varphi$$

它们都是不含时间的.从而称为稳定约束(scleronic constraints)⁶.但是它们对应着还存在类似的,非稳定约束(rheonomic constraints)⁷的版本,一种可能的情况如下:

$$x^2 + y^2 + z^2 = (l + vt)^2 ; \quad dx = (R + vt) \cos \theta d\varphi$$

它们就表示绳子还在不断变长或者轮子半径在不断膨胀的新问题.

与之前的完整非完整体系造成区别不同,也许在牛顿的力学体系中处理一个非完整约束体系和非稳定约束体系的难度是相当的:都要进行细致的受力分析,能量守恒不成立或者虽然成立但用处不大.但是在分析力学中,一个体系即使约束非稳定也是与稳定约束一样适用于大多数结论的.分析力学的方法在求解非稳定但完整的理想约束体系中是一大利器.

4.1.2 广义坐标与自由度

拥有了约束的观点,便很好理解广义坐标之间的独立性的概念.自由度(DOF, degree of freedom)也正是基于这一点而产生.它的定义是:考虑所有约束条件都成立的前提下,体系可以独立改变的坐标微分个数.

一般来说,在三维空间中(以下所有结论在二维或高维空间中都得重新讨论),在把一个体系的所有存在的约束都完全解除之后,体系可能由质点,线状刚体或者非线状刚体组成.那么直接地我们给出或者规定:质点的自由度为 3,非线状刚体的自由度为 6,线状刚体的自由度为 5.这一点待会就可以验证.而一个体系若包含 a 个质点, b 个线状刚体, c 个非线状刚体.而如果约束的引入为体系带来了 r 个独立的约束方程(几何约束或微分约束),显然根据自由度的定义,体系独立变化的广义坐标的数目降为:

$$s = 3a + 5b + 6c \longrightarrow f = s - r$$

这个 f 就是最终的自由度的个数.

质点的自由度为 3 自然是最好理解的.无论是用直角坐标,球坐标还是柱坐标,互相独立的坐标个数都是 3.而它们之间又不存在任何制约关系.

如果我们把两个质点用一个刚性棍连接,就会带来一个距离不变的约束.体系自由度变为:

$$f = 2 \cdot 3 - 1 = 5$$

这实际上就是线状刚体的自由度数.因为只要确定好线状刚体上的两个点的 6 个坐标,那么实际上就构成了对刚体的完整描述:每一个质元的位置都可以简单找到.而这六个坐标又有一个约束关系,从而自由度为 5.

但是非线状刚体就更复杂.既然非线状必然能够在刚体上找到三个不共面的固连点.只要确定好这三个点的 9 个坐标就能确定整个刚体的位形.但是这三个点之间两两距离不变,存在三个独立的约束方程,所以自由度数变为:

$$f = 9 - 3 = 6$$

⁵位形-微分流形上约束对应的矢量场间具有不为零的李导数.

⁶来自希腊语“坚硬”: σκλεός

⁷来自希腊语“流动”: φέω

为什么线状刚体不能作为非线状刚体的特例而也具有相同的 6 自由度呢? 比如一个陀螺, 作为一个刚体分析一般为认为它有三个平动自由度和三个转动自由度. 如果陀螺做定点转动, 它的三个转动自由度就被简单地隔离了出来: 它们是自转, 进动和章动. 但是我们不能指望一根线状刚体来自转: 自转进行一个角度, 对应的状态却实际上与之前的状态没有本质区别. 更重要地是反映在能量上, 自转既没有带来动能, 也没有带来势能的改变, 从而自转角根本就是一个冗余的内禀自由度(intrinsic DOF), 不会参与问题的讨论过程中.

从而以上论断也将基于一个基本事实: 质点和线状刚体是真正理想的, 现实模型: 小球, 细棍更精细的模型当然是 6 自由度的刚体, 甚至更高自由度的质点系, 连续体模型.

清楚了体系的基本组成部分的原始自由度数. 我们再把视野转向常见的约束上来.

要注意, 约束的个数这样一种说法的歧义性. 有时候我们发现体系有 r 个独立的约束方程, 从而说体系存在 r 个约束. 但是有时候我们指的是有几个机构在制约着物体的运动. 两者是有区别的, 我们保留前者的称呼, 后者的个数叫做约束物的个数.

Case1. 如果用一个面来限制质点的运动, 那么约束的个数是 1. 如果质点是一颗珠子被穿在一条空间曲线上, 那么尽管约束物只有一个, 但存在两个约束. 这是因为平面和曲线方程分别是:

$$\text{Surface : } F(x, y, z) = 0 \quad ; \quad \text{Curve : } \begin{cases} F(x, y, z) = 0 \\ G(x, y, z) = 0 \end{cases}$$

这也正构成了两种情况下的约束方程.

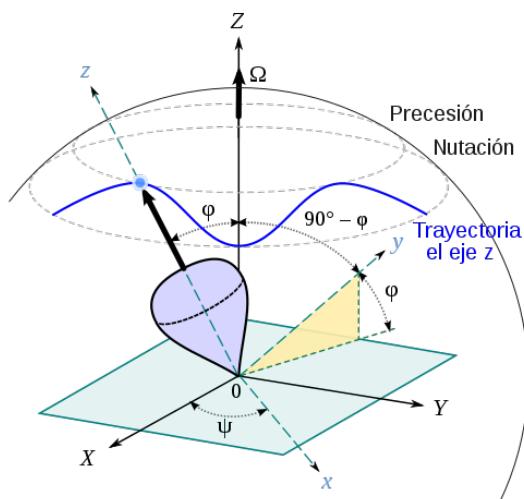


图 4.3: 刚体的三自由度转动

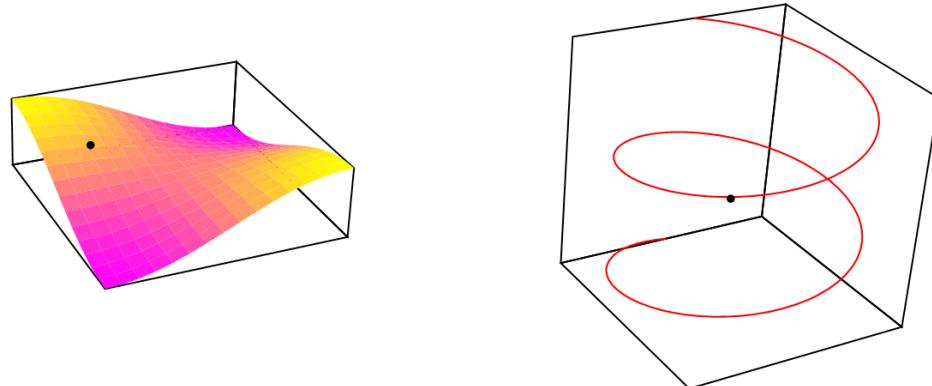


图 4.4: 面约束和线约束

Case2. 但是如果面和线是绝对粗糙的, 那么质点就动弹不得. 这样带来的约束个数, 显然两种情况都有三个.

Case3. 用绳子和杆子的两端连接两个物体, 对应的约束个数显然是 1. 就是说那两个点之间的距离为一个常数.

Case4. 如果绳子还绕过了滑轮, 那么原则上每一段绳带来一个约束减少一个自由度. 例如坐标为 (x_1, y_1, z_1) 的质点连接的绳(长 l) 绕过坐标为 (x_2, y_2, z_2) 的滑轮(视作质点) 与坐标为 (x_3, y_3, z_3) 的质点连接, 约束

方程就是:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} + \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2} = l$$

Case5. 而不同于用两段绳分别连接 12 和 23, 这样就会带来两个约束. 区别在于前者的一段绳两个部分可以交换绳长, 但是后者的两段绳则不能交换绳长.

Case6. 最后还有铰链模型. 如果仅仅是两个物体用铰链连接, 那么结论是显然的: 如果是平面问题, 那么约束个数为 2. 如果是空间问题, 那么约束个数为 3. 约束方程就是被连接的两个点具有完全一致的所有坐标.

Case7. 但如果是 n 个物体同时连接与一个铰链上, 那么我们只需要把第 1 个物体和第 2 个物体连接, 再把第 2 个物体和第 3 个物体连接... 最终把第 $n-1$ 个物体同第 n 个物体连接即可, 一共进行了 $n-1$ 次独立的连接, 故约束个数为 $2(n-1)$ (平面问题) 或 $3(n-1)$ (空间问题).

引入自由度的概念, 一个重要的原因是为了更好地描述体系的运动. 现在我们先看更复杂的情况: 如果体系是非完整体系, 那么事情比较难办: 不妨设 r 个独立约束中应当有 h 个完整约束和 k 个不完整的约束. 那么应当给予保留 $f' = s - h$ 个广义坐标. 体系的运动过程中, 这 f' 个广义坐标并不能完全相互决定, 记作 $q_1, q_2 \dots q_{f'}$. 但是体系任意状态的描述都只需要确定这 f' 个广义坐标的值和导数. 比如, 任何一个质点的 x 坐标, 还有速度的 x 分量就必然可以用之前的量来表示. 而还应当存在 k 个微分约束:

$$x = x(q_i, t) \quad , \quad \dot{x} = \frac{\partial x}{\partial t} + \sum_i \frac{\partial x}{\partial q_i} \dot{q}_i$$

$$f_j(q_i, dq_i, dt) = 0$$

$$i = 1, 2 \dots f' \quad , \quad j = 1, 2 \dots k$$

但是如果是完整体系, 问题就简单了, 自由度的个数 f 就会是体系独立的坐标的个数. 从而如果广义坐标是 $q_1, q_2 \dots q_f$, 它们之间就不存在约束关系了, 体系的一切量, 原来如果由类似于 x, \dot{x} 这样的量来表示, 现在就直接是:

$$x = x(q_i, t) \quad , \quad \dot{x} = \frac{\partial x}{\partial t} + \sum_i \frac{\partial x}{\partial q_i} \dot{q}_i \quad (i = 1, 2 \dots f)$$

尤其是稳定的完整约束体系. 那么由于约束方程不含 t , 那么上式就更是简化为:

$$x = x(q_i) \quad , \quad \dot{x} = \sum_i \frac{\partial x}{\partial q_i} \dot{q}_i \quad (i = 1, 2 \dots f)$$

4.1.3 约束力与广义力

为了简单起见, 我们把刚体也当作多个通过约束固连在一起的质点构成的系统. 这样体系便只含有以下要素: 质点, 外力, 约束, 质点间的内力, 约束力(也是内力).

正过来看, 任何一个约束都可以限制到若干质点. 从而同时对这几个物体的运动提出单个的制约关系(自由度减 1).

但是另一方面, 每一个质点又同时可能受到若干个约束物的共同限制, 每一个约束物就会对该质点施加一个或多个约束方程来把这个质点的运动和其他质点运动联系起来.

在牛顿力学情况下, 我们总是喜欢分析那个质点受到的约束物给它的力的具体特征. 这种力就叫做约束力.

约束力的最核心特征, 就在于它是被动力(passive force). 直接看是找不到这么一个力的大小与方向的. 约束力的具体值, 是两个因素共同作用的结果: 一方面就是约束本身的性质: 是光滑的还是粗糙的? 是要耗散还是甚至可以输出能量? 但是更重要的是, 约束力必然受到具体运动情况和体系动力学结构的影响. 这

一点很好理解: 因为如果把某个约束解除, 那么体系之后的运动就不一定遵从约束方程. 那么解除约束之前的体系就正是因为约束给出来的力影响了其运动而导致其运动“回归正轨”.

基于这一点, 就不难判断, 质点受到“必要”约束力的个数, 就是质点坐标参与约束方程的个数.

比如, 被穿在光滑曲线上的质点受到的约束力必然是两个, 再比如, 平面问题中, 两个用光滑铰链连接的物体会在两个方向上产生约束力, 而如果是 n 个物体铰接在一起, 那么期间我们如果设出 $2(n - 1)$ 个约束力就意味着个数设对了.

显然, 如果是粗糙的曲线上有一个滑动的质点, 曲线给质点的力应当有两个支持力和一个滑动摩擦力, 也就是说应当是三个, 而不是约束给出的两个. 再比如, 将一条长 l 的轻绳子绕过三个光滑滑轮组成一个闭合的圈. 那么明显每个物体与之相关的约束方程只有一个. 但是约束力似乎也得有两个拉力.

这两种情况也得分开讨论. 第二种情况下, 两个拉力总是相等, 从而三个质点的受力都指向三个角平分线的交点: 内心. 这个约束给三个质点的力其实应当认为是两个力的合力. 它只有沿内心与质点连线方向的分量, 没有垂直方向的, 可以算作一个约束力而不是两个. 但是前者的确应当算作三个约束力. 这是因为滑动摩擦力分量明显不是“必要的”. 它独立于之前的两个支持力. 也不同于之前两个支持力.

其中具体道理还得见后面虚功原理一节的关于理想约束的推导. 我们指出, 在上一章讨论约束力功的时候我们提出的约束力公式, 在这里可以直接推广. 实际上就能得到我们一直在说的“必要”约束力. 在一般情况下, 如果某约束对一质点的 x 坐标产生了限制:

$$f(x, \dots) = 0$$

那么与之相伴而生的作用在质点上的约束力就必然有一项 x 分量:

$$F_x = \lambda_f \frac{\partial f}{\partial x}$$

如果还含 y 那 y 方向自然也会产生一个约束力, 公式是一样的. 如果含有其他质点坐标时就换那个质点的坐标来求片导数. 注意由同一个约束方程产生的所有的约束力都共用同一个 λ_f .

对于理想约束这就是全部的约束力和它要符合的形式, 非理想约束则可能产生各种各样的其它形式的约束力.

还有一个重要概念需要先一步加以介绍. 如果我们考虑任何一个作用在实际体系的某质点上的力 \mathbf{F} , 该质点坐标为 $\mathbf{r} = (x, y, z)$. 我们引入一个新的概念. 这个力产生的在每一个广义坐标 q_i 上的广义力 Q_i : 其定义也相当简单, 就是要使得下面连等式成立:

$$dW = \mathbf{F} \cdot d\mathbf{r} = \sum_i Q_i dq_i$$

这是希望用广义力对广义位移来做功来代替以往的矢量力对质点做功的一成不变的做法. 我们默认体系都是稳定约束, 那么其实该质点的任意位移, 简单地可以用广义坐标上的位移来表示:

$$\mathbf{r} = \mathbf{r}(q_i) \Rightarrow d\mathbf{r} = \frac{\partial \mathbf{r}}{\partial q_i} dq_i$$

代入之前的要求就发现, 广义力的算法为:

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

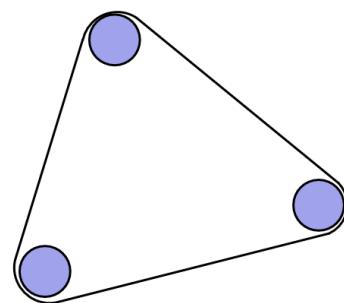


图 4.5: 三物体一约束

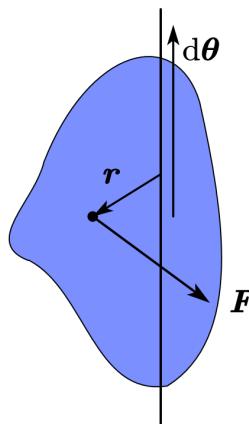


图 4.6: 力矩作为广义力

看两则应用:

一是刚体可做定轴转动,那么当刚体产生 $d\theta = d\theta e$ 的转动时,作用在 r 的力 \mathbf{F} 的作用点的实际位移为:

$$dr = d\theta \times r$$

那么力的实际做功为:

$$\begin{aligned} dW &= \mathbf{F} \cdot dr \\ &= \mathbf{F} \cdot (d\theta \times r) \\ &= (r \times \mathbf{F}) \cdot d\theta \\ &= M \cdot ed\theta \\ &= Q_\theta d\theta \end{aligned}$$

这也就是说,沿轴方向的力矩 $M \cdot e$ 就是与角位移 $d\theta$ 伴随的广义力 Q_θ .

再来看看在牛顿力学体系中比较基础的滑轮组问题.如果两个自由度分别取为:一. 2,3 挂着的动滑轮的下降与 1 上升的高度 x ;二. 相对动滑轮,2 的上升与 3 的下降高度 y .那么 1,2,3 的向下位移就分别为:

$$s_1 = -x, \quad s_2 = x - y, \quad s_3 = x + y$$

这样,三个重力在这两个自由度上累积的广义力就是:

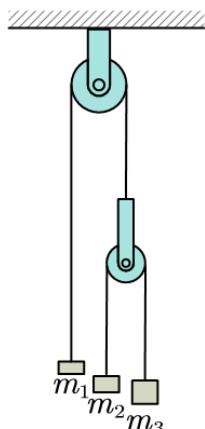


图 4.7: 滑轮组

$$Q_x = -m_1 g + (m_2 + m_3)g$$

$$Q_y = (m_3 - m_2)g$$

这有什么用?实际上我们计算动能还可以得到“广义质量”:

$$\begin{aligned} T &= \frac{1}{2}(m_1 \dot{s}_1^2 + m_2 \dot{s}_2^2 + m_3 \dot{s}_3^2) \\ &= \frac{1}{2}(m_1 + m_2 + m_3)\dot{x}^2 + \frac{1}{2}(m_2 + m_3)\dot{y}^2 + (m_3 - m_2)\dot{x}\dot{y} \end{aligned}$$

由于产生了交叉项,我们再进行一次换元:

$$z = y + \frac{m_3 - m_2}{m_3 + m_2}x$$

$$Q_x dx + Q_y dy = \left(Q_x - \frac{m_3 - m_2}{m_3 + m_2} Q_y \right) dx + Q_y dz = Q'_x dx + Q'_z dz$$

从而动能和广义力变为:

$$T = \frac{1}{2} \left(m_1 + \frac{4m_2 m_3}{m_2 + m_3} \right) \dot{x}^2 + \frac{1}{2} (m_2 + m_3) \dot{z}^2$$

$$Q'_x = \left(\frac{4m_2 m_3}{m_2 + m_3} - m_1 \right) g, \quad Q'_z = (m_3 - m_2)g$$

这就完全与两个质点独立地受到两个恒力的动能与广义力的“结构”一致了.从而两个自由度的坐标二阶导数为:

$$\ddot{x} = \frac{\frac{4m_2 m_3}{m_2 + m_3} - m_1}{\frac{4m_2 m_3}{m_2 + m_3} + m_1} g, \quad \ddot{z} = \frac{m_3 - m_2}{m_3 + m_2} g$$

从而最开始的广义坐标 y :

$$\ddot{y} = \ddot{z} - \frac{m_3 - m_2}{m_3 + m_2} \ddot{x} = \frac{2m_1 \frac{m_3 - m_2}{m_3 + m_2}}{\frac{4m_2 m_3}{m_2 + m_3} + m_1} g$$

本题用广义力方法求解的关键好处在于：本来作为矢量都同向的重力，在转化为广义力时可以因为体系结构不同而带上正负号，以适应最后的结果。

其实，广义力乘广义位移等于做功的现象我们并不少见，不管是热学中的 $dW = -pdV$ ，还是电路问题中的 $dW = UdQ$ ，抑或是电磁介质问题中的 $dW = EdP$ ， $dW = HdM$ 等等公式。全都归于统一的广义力乘广义位移这样的形式上。

4.2 力系化简

静力学中大量涉及到由质点、刚体和简单约束构成的问题。本节正是要针对这一类问题形成行之有效的强大工具。我们建立在之前的理论基础上，重新构建静力学过程中适用的新逻辑体系。

4.2.1 静力学公理体系

质点也好，刚体也好，其平衡都是较为简单讨论的。在某一个状态下，我们用**力系**(force system) 来描述它们的受力状态。比如刚体上受到的力系为 \mathfrak{F} 。一种简单的理解是，力系应当是一个场。在刚体的每一个点上都存在该点的受力情况。质点则只需要描述自己那点上的受力情况就行。而这个场，应当是一个矢量场：

$$\text{公理一} : \quad \mathfrak{F} : \mathbf{f}(r) = f_x(x, y, z)\mathbf{i} + f_y(x, y, z)\mathbf{j} + f_z(x, y, z)\mathbf{k}$$

如果 f_x, f_y, f_z 是连续函数，那么这样的力系实际运用中叫做**分布力系**(distributed force system)。 \mathbf{f} 其实在量纲和概念上为外力的力密度。但是想到质点受力的描述，我们立马发现所有力都必须集中于一个点，仿佛与分布力系略有区别：

$$\mathfrak{F} : \{\cdots(r, \mathbf{F}_i) \cdots\}$$

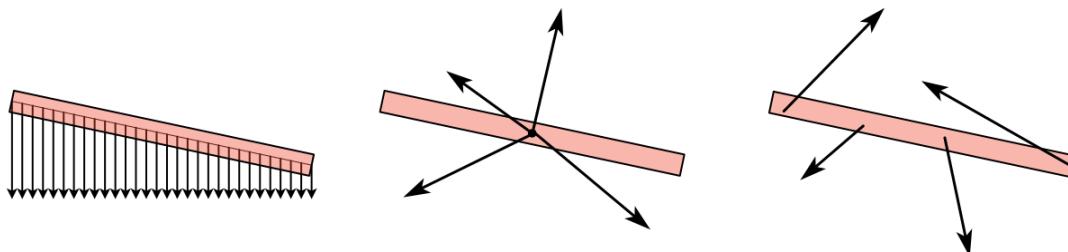


图 4.8: 分布力系, 共点力系与空间力系

其中 r 是质点的位矢。这样的力系可以叫做**共点力系**(concurrent force system)。事实上，我们也可以让刚体上若干点受到离散的作用力，这才是我们常说的**空间力系**(spacial force system)：

$$\mathfrak{F} : \{\cdots(r_i, \mathbf{F}_i) \cdots\}$$

但其实两者是统一的。统一的关键是要引入一种广义函数：**德尔塔函数**(delta function)。即：

$$\mathfrak{F}_1 : \{\cdots(r_i, \mathbf{F}_i) \cdots\} = \mathfrak{F}_2 : \sum_i \mathbf{F}_i \delta(\mathbf{r} - \mathbf{r}_i)$$

德尔塔函数的定义是这样的，如果让自变量为 Δ ，那么：

$$\delta(\Delta) = \begin{cases} 0 & (\Delta \neq 0) \\ +\infty & (\Delta = 0) \end{cases}, \quad \int_D \delta(\Delta) d\Delta = \begin{cases} 0 & (\mathbf{0} \in D) \\ 1 & (\mathbf{0} \notin D) \end{cases}$$

统一为分布力系将有利于统一运算:如果是力个数可数的离散的空间力系,可以考虑两个力系的集合并:

$$\{\cdots(\mathbf{r}_{1i}, \mathbf{F}_{1i})\cdots\} \cup \{\cdots(\mathbf{r}_{2j}, \mathbf{F}_{2j})\cdots\} = \{\cdots(\mathbf{r}_{1i}, \mathbf{F}_{1i})\cdots(\mathbf{r}_{2j}, \mathbf{F}_{2j})\cdots\}$$

但是如果力共点了又应当去求矢量和.也就是说,本来共点力系的力就是一个三维空间中的线性代数,它们理所当然地定义了加法和数乘:

$$\text{Addition: } \mathbf{F}_1 + \mathbf{F}_2 = \mathbf{F}$$

$$\text{Multiplication: } \lambda \mathbf{F}_1 = \mathbf{F}_2$$

现在就可以发现.第一,把共点力系的加法(力的叠加),现在就可以自然地推广到分布力系的加法来(数乘也同理).只需要逐点把每点处的两个力密度矢量加在一起就是合力系.例如,如果两个力系分别为 $\mathfrak{F}_1: \mathbf{f}_1(\mathbf{r})$ 和 $\mathfrak{F}_2: \mathbf{f}_2(\mathbf{r})$,那么:

$$\mathfrak{F}_1 + \mathfrak{F}_2 : \mathbf{f}_1(\mathbf{r}) + \mathbf{f}_2(\mathbf{r})$$

$$\lambda \mathfrak{F}_1 : \lambda \mathbf{f}_1(\mathbf{r})$$

这样就不需要分离散力系的并与共点力系的和来阐述物理上的“叠加”,它们本质上就是力密度分布函数的和.线性代数中一切应当符合的“泛性质”它也应当符合.其零元素就称作零力系(null force system):

$$\mathbf{o} : \mathbf{o}(\mathbf{r}) = \mathbf{0}$$

静力学自成一套的理论核心,是虽然质点平衡时一定意味着合力系为零力系.但是在一个刚体平衡的时候,它依然可以对应在不同点受到不同力的一个力系.事实上,即使不平衡,两个不同的力系其“效果”也可以是相同的.以往的力学告诉我们,效果其实是指产生的加速度或角加速度等运动学响应.但是,在公理化静力学的过程中,我们不需要关心“效果”一词的具体含义,而是抽象的定义一种新的关系:

这就是两个力系的等效(equivalence),如果两个力系 \mathfrak{F} 和 \mathfrak{G} 等效,就记作 $\mathfrak{F} \equiv \mathfrak{G}$.其实数学上有一对对应的概念为等价关系(equivalence relation),它需要满足以下三个条件,在物理观点下看是显然的:

- 反身性(reflexive property): $\mathfrak{F} \equiv \mathfrak{F}$
- 对称性(symmetric property): $\mathfrak{F} \equiv \mathfrak{G} \Leftrightarrow \mathfrak{G} \equiv \mathfrak{F}$
- 传递性(transitive property): $\mathfrak{F} \equiv \mathfrak{G}, \mathfrak{G} \equiv \mathfrak{H} \Rightarrow \mathfrak{F} \equiv \mathfrak{H}$

但是还有一点十分必要,这样一个等价关系必须还得是同余(congruence)的.即,它必须保持代数关系不变.这就构成了静力学第二公理:

$$\text{公理二: } \mathfrak{F}_1 \equiv \mathfrak{F}_2, \mathfrak{G}_1 \equiv \mathfrak{G}_2 \Rightarrow \mathfrak{F}_1 + \mathfrak{G}_1 \equiv \mathfrak{F}_2 + \mathfrak{G}_2, \lambda \mathfrak{F}_1 \equiv \lambda \mathfrak{F}_2$$

这两个公理赋予力系这一个代数系统以极其规则而且性质丰富的结构.例如,我们可以定义平衡力系(equilibrium force system)的概念:如果 $\mathfrak{K} \equiv \mathbf{o}$,即等效于零力系的力系就是平衡力系.那么显然任何一个力系上增减平衡力系的结果都会得到等效的力系:

$$\mathfrak{K} \equiv \mathbf{o} \Rightarrow \mathfrak{F} + \mathfrak{K} \equiv \mathfrak{F}$$

那么就可以把所有平衡力系全部都找到,构成一个平衡力系的集合,称作平衡核(kernal of equilibrium):

$$K = \{\mathfrak{K} | \mathfrak{K} \equiv \mathbf{o}\}$$

那么两个力系等效的充要条件就是其差在平衡核内,就是说如果把一个力系中的力完全做反向操作后两者就可以平衡:

$$\mathfrak{F} \equiv \mathfrak{G} \Leftrightarrow \mathfrak{F} - \mathfrak{G} \equiv \mathbf{o} \Leftrightarrow \mathfrak{F} - \mathfrak{G} \in K$$

最后就可以定义力系的等效类:所有与力系 \mathfrak{F} 等效的力系也构成了一个集合,记作 $[\mathfrak{F}]$.也就是说上面的平衡核其实都是 $[\mathfrak{o}]$.那么这个类的结构其实就是:

$$[\mathfrak{F}] = \{\mathfrak{F} + \mathfrak{K} | \mathfrak{K} \in K\}$$

这样就把所有可能的力系的集合分解为一个一个的等效类,每一个集合内部的每个力系彼此之间完全等效,而平衡核是它们之中所有平衡力系构成的“核心”等效类.不同的类之间就不存在等效关系.每一个等效类内部的差其实都是一些平衡力系.两个不同的等效类中的力系差依然是一个非平衡的力系.

所以这样就把最终问题直接等效为最后一个细节:那就是平衡核究竟有多大?怎样的力系是平衡力系,属于平衡核?

这个问题的答案其实简单:那便是找到最基础的平衡力系,它是平衡核的生成元(generating element).这就是静力学的第三公理:任意二力构成的,等大反向共线的力系全构成平衡力系,它是二力平衡的充分且必要条件:

$$\text{公理三: } \{(r_1, F_1), (r_2, F_2)\} \in K \Leftrightarrow F_1 + F_2 = \mathbf{0}, F_1//F_2//(r_2 - r_1)$$

这个定理又忽然把我们从分布力系的问题拉回了特殊的空间力系分析的问题,它们都实用且简单.但是从特殊到普遍,从二力平衡条件开始,一步一步去探究力系是否是平衡力系的判据,乃至找到所有可能的平衡力系的过程可是非常细致与漫长.后面我们会一个推论一个推论地来完成这个过程.在这里我们先尚且指出另外两个通常也被认为是静力学公理的性质,成为我们接下来一些推理的出发点.

当我们分析一个含约束的质点,刚体构成的系统时,它们不仅受到外力,还要受约束力(作为内力).那么以前的牛顿第三定律就可以完全被静力学化也成为公理:如果约束发生在两个物体的两点 1 和 2 之间,位置分别为 r_1 和 r_2 .那么 1 给 2 的力 F_{12} 和 2 给 1 的力 F_{21} 就要符合:

$$\text{公理四: } F_{12} + F_{21} = \mathbf{0}, F_{12}//F_{21//(r_2 - r_1)}$$

这样是否就能解决任何静力学平衡问题呢?对一个刚体而言这就够了.其力系是平衡力系自然就是平衡条件.然而对复杂体系的事实是:约束力作为被动力往往也要作为待求的量进入平衡方程.在约束力确定之前自然不能知道是否能平衡.那么如何评价含未知约束力时的平衡问题呢?我们规定:

公理五: 系统平衡充要条件是,存在一种约束力取值,考虑约束力后每一个部分单独平衡.

它还有一个常见推论:由于考虑每一个部分的力系的时候是平衡力系,那么把这些力系进行叠加也是平衡力系,再考虑到成对存在的内力符合公理四,再由公理三,公理四的条件恰好让这些内力构成平衡力系.故扣除掉以后,剩下的外力力系自然也必须是平衡力系,也就是说如果设想整个体系等效为一个大刚体,那么外力就是它受到的所有力,此时就能得出为了平衡,外力就必须构成平衡力系的条件,因此,这个推论得名**刚化原理**(principle of solidification).这是原来体系平衡的必要条件.

刚化原理: 系统平衡必要条件是外力 $\mathfrak{F} \equiv \mathbf{0}$

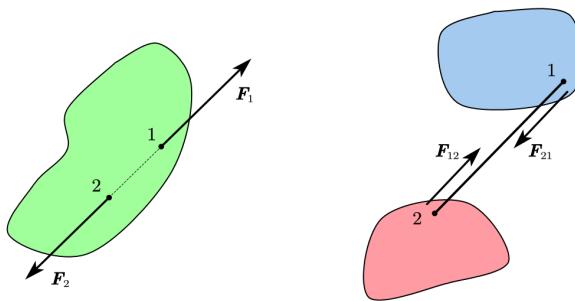


图 4.9: 第三公理和第四公理

有了这些公理的基础, 我们就可以获得十分强大的一些推论 (证明过程略), 它们都是我们进行力系化简的关键工具:

推论一: 力是滑移矢量: $(r_1, \mathbf{F}) \equiv (r_2, \mathbf{F}) \Leftrightarrow \mathbf{F} / \parallel (r_2 - r_1)$

从而, 描述一个刚体上力的效果时. 只有力的作用线是关键的. 将一个力沿作用线滑移不改变力的作用效果.

如果把两个力系进行“叠加”, 可以称作“合成”. 把共点力系做矢量和也可以称作“合成”但是如果出现一个完整的力系恰好等效于一个单力构成的力系的情况, 则称这个单力为力系的合力(join force).

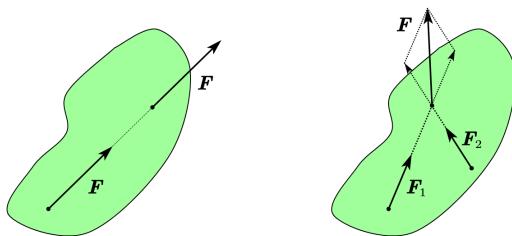
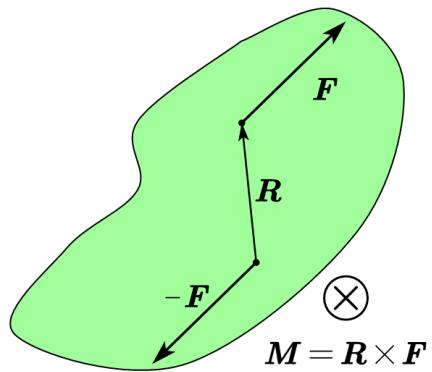


图 4.10: 滑移与合力

上一个推论的直接结果是: 可以在平面力系的情形下合成任意两个非平行力, 即非平行力求合力:

推论二: 非平行力求合力: 如果两个力作用线交于一点, 那么可以滑移到该点来做矢量合成求合力.



为了求解力的非滑移的平移如何等效的问题. 我们需要开始引入矩的概念. 如果一个单力力系保持大小方向不变但作用在另一个点上, 那么两个力系做差的同余类便是一个力偶(force couple). 最典型的情形下, 力偶由两个力构成:

$$\mathfrak{M} = \{(r_1, \mathbf{F}), (r_2, -\mathbf{F})\}$$

力偶的特征由力偶矩(moment of couple) 来描述. 它被定义为:

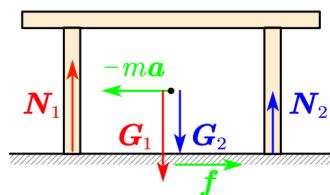
$$\mathbf{M} = (r_1 - r_2) \times \mathbf{F}$$

图 4.11: 力偶

可以分解为力偶的力系称作力偶系(system of couples). 例如如图, 在加速运动的车厢中静止的桌子的受力就构成功力偶系. 我们把重力分解为前后脚的支持力之和, 并分别与两个支持力构成力偶. 摩擦力则与惯性力构成功力偶.



这样一个力偶系的平衡条件是显而易见的. 事实上, 对力偶系的分析中, 力偶本身是一种非定位的元素. 只需要两个力构成的单力偶系的力偶矩一致, 它们就是等效的. 所以力偶系的描述方法为:



$$\mathfrak{M} = \{M_1, M_2 \dots M_n\}$$

每一个力偶都不需要具体指定是由哪两个力组成, 位置在哪儿. 也正因为如此, 力偶系描述时这些力矩都不需要支配任何作用点. 力偶系的化简更是简单: 直接对力偶做矢量和便是等效的.

图 4.12: 桌子受力分析

$$\{M_1, M_2 \dots M_n\} \equiv \{M = \sum_i M_i\}$$

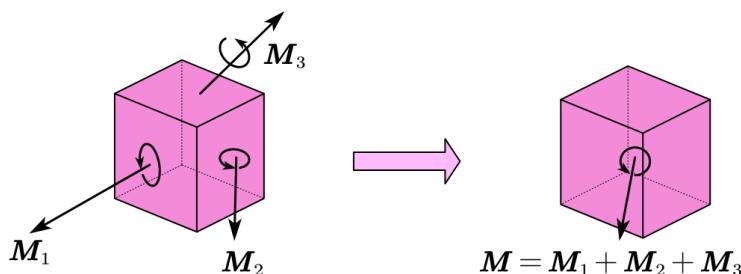


图 4.13: 力偶系合力

我们也把最后那个合力矩 M 称为力偶系的合力. 现在我们可以写出共点力系和力偶系的平衡充要条件:

推论三： 共点力系与力偶系平衡条件：元素的矢量和为零, 即合力 $\sum_i \mathbf{F}_i = \mathbf{0}$ 或 $\sum_i \mathbf{M}_i = \mathbf{0}$.

最后, 为了研究普遍的力系化简的方法, 利用力偶工具我们可以不限于滑移, 而是实现力的平移. 我们指出在之后的过程中至关重要的力的平移定理:

推论四： 力的平移定理 : $(\mathbf{r}, \mathbf{F}) \equiv \{(\mathbf{r} + \mathbf{R}, \mathbf{F}), \mathbf{M} = -\mathbf{R} \times \mathbf{F}\}$

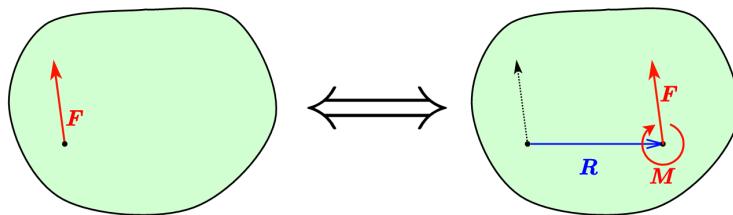


图 4.14: 力平移定理

即力平移以后的效果与原来的效果的区别可以被一个力偶等效.

如果力系简单, 那么其平衡条件也是简单的. 二力平衡条件就是等大, 反向, 共线. 而三力平衡条件也相对简单实用, 就是三力汇交, 力三角两个条件.

推论五： 三力平衡条件 : 三力作用线三线共点, 且三力首尾相连构成闭合三角形.

利用三力汇交可以把简单力系的平衡条件完全变成几何上的边角关系. 而“力三角”的矢量三角形在很多具体的例子中也恰好与位形空间中的三角形相似, 以提供一种完全几何化的解题思路.

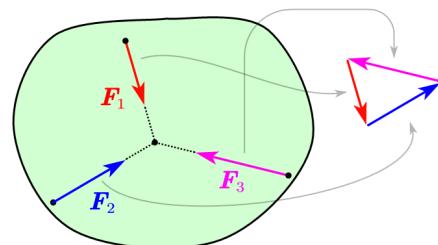


图 4.15: 三力汇交力三角

4.2.2 力系向一点简化

利用强大的力平移定理, 我们就可以实现力系向一点的简化. 对于任意一个空间力系, 在空间中取一点作为简化中心. 那么就以这个点作为原点来写出每个力的作用点, 力系被描述为:

$$\mathfrak{F} = \{\cdots (\mathbf{r}_i, \mathbf{F}_i) \cdots\}$$

但是又根据力平移定理:

$$(\mathbf{r}_i, \mathbf{F}_i) \equiv \{(\mathbf{0}, \mathbf{F}_i), \mathbf{M}_i\} , \quad \mathbf{M}_i = \mathbf{r}_i \times \mathbf{F}_i$$

故力系简化为:

$$\mathfrak{F} \equiv \{\cdots (\mathbf{0}, \mathbf{F}_i) \cdots \mathbf{M}_i \cdots\}$$

再由共点力系求合力与力偶系求合力的方法:

$$\{\cdots (\mathbf{0}, \mathbf{F}_i) \cdots\} \equiv (\mathbf{0}, \mathbf{F})$$

$$\{\cdots \mathbf{M}_i \cdots\} \equiv \mathbf{M}$$

这个和称作力系对简化中心的主矢(principal vector) 与主矩(principal moment):

$$\mathbf{F} = \sum_i \mathbf{F}_i ; \quad \mathbf{M} = \sum_i \mathbf{r}_i \times \mathbf{F}_i$$

其实,这样的结果可以推广到十分普遍的分布力系上来:

$$\mathbf{F} = \int \mathbf{f} dV ; \quad \mathbf{M} = \int \mathbf{r} \times \mathbf{f} dV$$

力系平衡的充要条件,是对任何一点主矢和主矩同时为零.也就是说有无数个彼此等价的条件.这些条件彼此之间有什么关系?这一点需要我们考察对不同点作简化的结果.事实上,如果力系对原点简化结果为:

$$\mathfrak{F} \equiv \{(\mathbf{0}, \mathbf{F}), \mathbf{M}\}$$

那么换成位矢为 \mathbf{R} 的一点,简化结果一定为:

$$\mathfrak{F} \equiv \{(\mathbf{R}, \mathbf{F}), \mathbf{M} - \mathbf{R} \times \mathbf{F}\}$$

也就是说,主矢是不会变的,只有主矩会变.从而主矢的不等于零是本质的,无法改变的.但是主矩的不等于零确是条件的.最后考虑若的确 $\mathbf{F} \neq \mathbf{0}$,但当 \mathbf{R} 任取时,主矩上多出来的项 $-\mathbf{R} \times \mathbf{F}$ 具有两个性质:一是必须与 \mathbf{F} 垂直,二是在与 \mathbf{F} 垂直的平面上可以取到任意值.我们还可以总结出平面力系与空间力系的最终简化结果:

平面力系的最终简化结果分类:

- 平衡力系: $\mathfrak{F} \equiv \mathbf{0}$
- 单力: $\mathfrak{F} \equiv (\mathbf{r}, \mathbf{F})$
- 单力偶: $\mathfrak{F} \equiv \mathbf{M}$

空间力系的最终简化结果分类:

- 平衡力系: $\mathfrak{F} \equiv \mathbf{0}$
- 单力: $\mathfrak{F} \equiv (\mathbf{r}, \mathbf{F})$
- 单力偶: $\mathfrak{F} \equiv \mathbf{M}$
- 力螺旋: $\mathfrak{F} \equiv \{(\mathbf{r}, \mathbf{F}), \mathbf{M}\}$ 且 $\mathbf{F} // \mathbf{M}$

4.3 平衡问题: 矢量力学

4.3.1 平衡问题的要素

在一个典型的平衡问题中, 作为已知量未知量的可能要素如下所示. 本节枚举中如果某要素只能为已知则标蓝, 如果只能为未知(即待求解的量) 则标红. 都可以则标紫.

- **体系的结构:** 所有平衡问题中, 体系的结构都是完全独立于其它量的已知未知提法的. 它在确定其它量的已知未知前就已经被完全确定是已知的. 它的内容包括: 约束个数, 约束方程形式, 约束力个数与写法, 自由度个数, 平衡方程个数, 平衡方程写法.
- **约束力:** 约束力从来不会在平衡问题中先行给出, 它永远是待求解的, 正因为如此我们称呼它们为被动力. 其它所有力都与它们对立, 称为主动力.
- **主动力:** 只要不是约束给出的力, 原则上都视为主动力. 但是不是所有主动力, 在一个典型的平衡问题中都会事先给出. 有些力, 比如重力, 总是已知. 但是, 比如一些问题问法就是求某外力等于多少时体系可以平衡; 有些题目中出现弹簧的弹力, 但是弹簧的伸长本身又是未知待求解的; 动摩擦力, 作为主动力, 它依赖于未知的支持力; 还有的题目干脆把约束解除, 用未知的主动力代替原来也未知的约束力. 这些情况下主动力也会变成未知力.
- **位形变量.** 很明确地, 一个平衡问题分为两大类. 一类问题 (由于自由度为零或负) 中每一个几何参量都在问题提出的事后都完全已知. 但是另一类问题 (自由度大于零) 中, 即使是平衡时的几何参量也是实现未知的, 需要通过平衡条件加以求解. 这里只指出这一种现象, 如何分类见之后小节: 负静定, 静定与超静定.

4.3.2 平衡条件与判据

平衡方程, 顾名思义, 就是平衡时各个已知量, 未知量之间要满足的具体数学方程. 平衡方程可以分为三类:

1. 平衡条件, 具体列法见下. 它可以同时包含位形变量, 主动力和约束力三类已知未知量.
2. 约束方程, 即以前讨论过的几何上的约束关系. 它只包含位形变量一类未知量.
3. 主动力的结构性方程, 有一些未知的主动力实际上是由一定的结构导致的, 弹簧上的力和动摩擦力就是典型的例子, 它们都有一些计算公式. 这样的方程也可以同时包含位形变量, 主动力和约束力三类已知未知量.

第一类方程, 也就是平衡条件(condition of equilibrium) 无疑是平衡问题中最丰富的一类方程. 具体问题中它却往往有很多种列法, 并造成或简单或复杂的求解过程. 这些条件中最基础的一类就是单个质点和刚体的平衡条件. 它们分别为:

A. 质点沿 l 方向的力平衡条件: 质点上的主动力与被动力构成共点力系 $\{\dots \mathbf{F}_i \dots\}$, 任意选取某方向 l 的单位矢量 e_l 以后有:

$$\sum_i e_l \cdot \mathbf{F}_i = 0$$

B. 刚体沿 l 方向的力平衡条件: 刚体受到主动力和被动力构成空间力系 $\{\dots (\mathbf{r}_i, \mathbf{F}_i) \dots\}$, 任意选取某方向 l 的单位矢量 e_l 作为投影方向, 平衡条件为:

$$\sum_i e_l \cdot \mathbf{F}_i = 0$$

C. 刚体绕 \mathbf{R} 点沿 l 方向的力矩平衡条件: 刚体受到主动力和被动力构成空间力系 $\{\cdots(\mathbf{r}_i, \mathbf{F}_i)\cdots\}$, 任意选取位矢为 \mathbf{R} 的点作为矩心, 任意选取某方向 l 的单位矢量 e_l 作为投影方向, 平衡条件为:

$$\sum_i e_l \cdot [(\mathbf{r}_i - \mathbf{R}) \times \mathbf{F}_i] = 0$$

在三类平衡条件中, 对一个具体的问题, 由于 l 方向与 \mathbf{R} 点选取的任意性, 平衡条件的列法都是无穷无尽的, 但这不一定意味着这些所有方程都是独立(independent)的, 也就是说往往只列出少数几个就能充分地包含每一种列法的正确性. 还有就是也不是所有列法在解决问题上都是同等简易的, 比如为了使得某个未知力不出现在方程中以减少未知数和方程的个数, 一方面利用整体法(integral method)来消灭内力, 一方面在列力平衡条件时选取垂直于力的方向作为投影方向, 一方面列力矩平衡条件时选取力作用线上的点作为矩心.

以上三类方程还都属于隔离法(isolation method)的平衡条件. 事实上根据刚化原理, 我们还可以把若干质点刚体看作整体, 把作为内力的约束力忽略, 列出整体法的平衡条件. 容易理解这些平衡条件全都是隔离法平衡条件的组合, 并不独立于隔离法的平衡条件.

于是一个重要的问题便浮现了出来, 如何找到所有可能的平衡条件中, 最少的, 相互独立的, 又能充分蕴含一切平衡条件的“核心平衡条件”? 这些固定的平衡条件构成的组被称作平衡判据(criterion of equilibrium). 对这个问题的研究给出以下结果:

I. 平面力系:

A. 质点的平衡判据

$$\begin{cases} \sum_i e_{l_1} \cdot \mathbf{F}_i = 0 \\ \sum_i e_{l_2} \cdot \mathbf{F}_i = 0 \end{cases} \quad (e_{l_1} \neq \pm e_{l_2}, \text{ 即两方向不平行})$$

B. 刚体的平衡判据

$$\begin{cases} \sum_i e_{l_1} \cdot \mathbf{F}_i = 0 \\ \sum_i e_{l_2} \cdot \mathbf{F}_i = 0 \end{cases} \quad (e_{l_1} \neq \pm e_{l_2}, \text{ 即两方向不平行})$$

$$\begin{cases} \sum_i e_l \cdot \mathbf{F}_i = 0 \\ \sum_i (\mathbf{r}_i - \mathbf{R}_1) \times \mathbf{F}_i = \mathbf{0} \quad (e_l \cdot (\mathbf{R}_1 - \mathbf{R}_2) \neq 0, \text{ 即矩心连线不能垂直于力投影方向}) \\ \sum_i (\mathbf{r}_i - \mathbf{R}_2) \times \mathbf{F}_i = \mathbf{0} \end{cases}$$

$$\begin{cases} \sum_i (\mathbf{r}_i - \mathbf{R}_1) \times \mathbf{F}_i = \mathbf{0} \\ \sum_i (\mathbf{r}_i - \mathbf{R}_2) \times \mathbf{F}_i = \mathbf{0} \quad (\text{三矩心不共线}) \\ \sum_i (\mathbf{r}_i - \mathbf{R}_3) \times \mathbf{F}_i = \mathbf{0} \end{cases}$$

II. 空间力系: A. 质点的平衡判据

$$\left\{ \begin{array}{l} \sum_i e_{l_1} \cdot \mathbf{F}_i = 0 \\ \sum_i e_{l_2} \cdot \mathbf{F}_i = 0 \quad (\text{三投影方向不共面}) \\ \sum_i e_{l_3} \cdot \mathbf{F}_i = 0 \end{array} \right.$$

B. 刚体⁹的平衡判据

$$\left\{ \begin{array}{l} \sum_i e_{l_1} \cdot \mathbf{F}_i = 0 \\ \sum_i e_{l_2} \cdot \mathbf{F}_i = 0 \\ \sum_i e_{l_3} \cdot \mathbf{F}_i = 0 \\ \sum_i e_{l_4} \cdot [(\mathbf{r}_i - \mathbf{R}_1) \times \mathbf{F}_i] = 0 \\ \sum_i e_{l_5} \cdot [(\mathbf{r}_i - \mathbf{R}_2) \times \mathbf{F}_i] = 0 \\ \sum_i e_{l_6} \cdot [(\mathbf{r}_i - \mathbf{R}_3) \times \mathbf{F}_i] = 0 \\ \dots \end{array} \right. \quad (\text{三力投影方向, 三力矩投影方向都不共面})$$

¹⁰

我们发现, 质点和刚体的平衡判据的方程个数都是严格等于无约束情况下各自的自由度数的: 平面问题质点是 2 而刚体是 3, 空间问题质点是 3 而刚体是 6. 线状刚体, 其力系也会具有共轴的特征, 读者可以证明, 此时空间问题中无论如何书写独立的方程数也将退化为 5.

平衡问题复杂, 多变. 牛顿矢量力学情况下有没有万全的求解套路呢? 对于一部分问题是有的, 这给出了:

4.3.3 亚静定, 静定与超静定

单独考虑约束方程与位形变量之间的关系是任何平衡问题的分类过程中首当其冲的. 参考之前对自由度与约束的讨论. 一个体系若包含 a 个质点, b 个线状刚体, c 个非线状刚体, 没有任何约束前自由度为 $s = 3a + 5b + 6c$. 至此我们已经知道了一点, 完全利用隔离法, 可以列的平衡方程就是所有物体的平衡判据组合在一起, 一共也是 $s = 3a + 5b + 6c$ 个方程.

这时候就开始分为三种典型的情形:

1. 体系具有 $r < s$ 个约束. 首先, 增添 r 个约束方程, 那么体系剩余自由度为 $f = s - r$, 这就是说凭约束方程不足以解出平衡时的位形变量. 称作**亚静定问题**(hypo-statically-determinate problem). 其次, 此时待解未知数至少有 f 个剩余独立位形变量, 和 r 个约束带来的约束力. 但如果所有主动力已知, 那么, 刚刚好就能够用 s 个平衡方程求解出所有未知数来. 如果 f 个独立的位形变量作为已知量由问题

⁹本条均指非线状刚体.

¹⁰还有 2 力 +4 力矩, 1 力 +5 力矩, 6 力矩三类平衡判据, 在此略去.

事先给定,那么方程多而未知数少,就有可能发生无解的情况,这实际上就代表着体系无法平衡,自由度开始演化,问题就从静力学问题转变为了动力学问题,也就是可能发生“不静”的情况.

2. 体系恰好具有 $r = s$ 个约束.那么,不同于上一种情况,平衡时的位形变量就可以被完全确定下来.就是说体系处于绝对的“静”的状态.称作**静定问题**(statically-determinate problem).其次, s 个约束带来的约束力,在已知所有主动力的前提下,恰好可以用 s 个平衡方程求解 s 个约束力.这就是“定”的来源.
3. 体系具有 $r > s$ 个约束.那么约束方程显然是随自治但冗余(不独立)的.相同于上一种情况的是,体系没有自由度了,处于绝对的“静”的状态.称作**超静定问题**(hyper-statically-determinate problem).这是因为,不同于上一种情况的是,体系将存在 r 个未知约束力,已经超过可列的所有平衡方程的个数 s 了.这会导致约束力不确定,存在多解的情况.此时体系平衡是一定可以实现的.毕竟体系平衡定义为任意一组解能导致各部分即可.但是由于多解性,它其实是“不定”.所以这一类问题又被称作**静不定问题**(statically-indeterminate problem).

静不定问题常见于带静摩擦的平衡问题中,此时支持力和静摩擦力都是约束力,带来的约束较多.摩擦力由于解不唯一往往有一个范围,那么范围中的最小静摩擦力小于接触点的临界静摩擦力就成为了必要的平衡条件.实际解题过程中往往还结合摩擦角,摩擦锥等几何化方法进行求解,在此不再赘述.

4.4 平衡问题: 虚功原理

在一个平衡问题中,如果想对它进行分析力学化,那么需要对体系进行较简单的抽象.

我们考虑一个包含 a 个质点, b 个线状刚体, c 个非线状刚体的体系,初始自由度为 $s = 3a + 5b + 6c$.加上 r 个约束.那么自由度和广义坐标的个数都是 $f = s - r$,不妨设广义坐标为 $q_1, q_2 \dots q_n$.再施加若干主动力.这些主动力要么作用于质点上,要么作用于刚体的某点上.而约束将能给出:任意一个点的坐标都可以用 f 个广义坐标来表述:

$$\mathbf{r}_k = \mathbf{r}_k(q_i, t) \quad i = 1, 2 \dots f$$

反过来理解,约束方程本身又是对那些点的坐标,进行限制,也就是说下面 r 个约束方程也能反映约束:

$$f_j(x_k, y_k, z_k, t) = 0 \quad j = 1, 2 \dots r$$

现在既然是静力学,我们就多提一点要求:以上两组方程都不能随时间变化:

$$\mathbf{r}_k = \mathbf{r}_k(q_i) \quad ; \quad f_j(x_k, y_k, z_k) = 0$$

这样我们就可以引入**虚位移**(virtual displacement)这一概念的初等理解方式.它就是指在平衡位置附近假想的位移(坐标改变),发生位移了以后不一定还能平衡,这很显然.但是这个位移决不能破坏约束.即:

$$\delta \mathbf{r}_k = \sum_i \frac{\partial \mathbf{r}_k}{\partial q_i} \delta q_i \quad ; \quad \sum_k \left(\frac{\partial f_j}{\partial x_k} \delta x_k + \frac{\partial f_j}{\partial y_k} \delta y_k + \frac{\partial f_j}{\partial z_k} \delta z_k \right) = 0$$

另一个要求,彻底决定了约束力的性质,让平衡问题成为分析力学中非常简单的一类问题.就是关于约束必须是理想约束的要求:

4.4.1 理想约束

定义方法一:如果约束力总是可以通过约束方程来如此计算,那么约束就是理想约束:

$$R_{kx} = \lambda \frac{\partial f_j}{\partial x_k}, \quad R_{ky} = \lambda \frac{\partial f_j}{\partial y_k}, \quad R_{kz} = \lambda \frac{\partial f_j}{\partial z_k}$$

定义方法二: 如果约束力永远不做虚功, 那么约束就是理想约束:

$$\forall \delta \mathbf{r}_k, \sum_k \mathbf{R}_k \cdot \delta \mathbf{r}_k = 0$$

结合对 $\delta \mathbf{r}_k$ 需要满足的约束方程来看, 以上两种定义方法的等价性是不言自明的.

值得注意的是, 根据之前引入的广义力的说法, 如果把作用在各个质点和刚体上的点上的约束力来求对各个广义坐标的广义力:

$$Q_{Ri} = \sum_k \mathbf{R}_k \cdot \frac{\partial \mathbf{r}_k}{\partial q_i}$$

又根据理想约束的定义, 其虚功为零, 但具有这样的算法:

$$\delta W_R = 0 = \sum_k \mathbf{R}_k \cdot \delta \mathbf{r}_k = \sum_k \mathbf{R}_k \cdot \sum_i \frac{\partial \mathbf{r}_k}{\partial q_i} \delta q_i = \sum_i \left(\sum_k \mathbf{R}_k \cdot \frac{\partial \mathbf{r}_k}{\partial q_i} \right) \delta q_i = \sum_i Q_{Ri} \delta q_i$$

由于对任意虚位移 $\forall \delta q_i$ 上式都成立, 这就得到了:

$$Q_{Ri} = 0$$

所以理想约束的约束力只会作为内力而存在, 其在每一个广义坐标上都没有任何广义力以影响这个自由度的运动. 广义坐标的选取本身就是一个隐藏内部约束效果的结果.

4.4.2 亚静定问题的虚功原理

上述讨论的情形实际上就是亚静定问题. 现在的平衡问题平衡条件就可以得到一个更简单的阐述. 既然无论质点还是刚体上, 考虑了主动力和约束力以后其合力都是零. 那么在发生一个虚位移时, 所有这些力做的总功都必然是零:

$$\delta W + \delta W_R = 0$$

既然现在是虚位移, 约束力的虚功一定是零, 那么主动力的虚功必然也有:

$$\delta W = 0$$

这其实就已经构成平衡条件了. 而它就是**虚功原理**(principle of virtual work): 一个理想约束系平衡时, 主动力的虚功和必然是零.

而如果找到主动力在每一个自由度上的广义力, 上式还可以被表述地更加具体:

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

与矢量力学做比较, 对于亚静定问题, 矢量力学用 s 个平衡条件求解出 r 个约束你和 f 个位形变量. 但是分析力学, 直接用以上 f 个广义力为零的方程求解出 f 个位形变量, 即广义坐标的平衡位置的值来. 对问题的简化可见一斑.

4.4.3 静定问题的求解

对于静定问题以上方法也完全具有求解各个约束力的过程. 如果要求解某一个约束上的约束力, 那么就解除这一个约束, 新的体系就成为一个亚静定的体系了. 此时把原来根据约束方程确定的约束力(未知大小)换成主动力, 如果一共解除了 f 个约束, 就会出现 f 个未知的主动力, 新的体系的广义坐标其实反倒是继承了原来静定问题的广义坐标值, 是已知的, 无需求解的. 从而依然是未知的 f 个主动力在 f 个自由度上的广义力为零:

$$Q_i = 0$$

就能够通过 f 个方程求解出原来的 f 个约束力来.

4.5 分析力学初步 *

经过上一节对虚功原理的探讨, 我们察觉到了分析力学理论建立的最基本条件: 约束必须是理想的. 但是与矢量力学方便处理的情况不同的是, 理想的约束不一定是稳定的, 此时分析力学照样也能够处理.

这就是说, 与上一节相同, 考虑一个包含 a 个质点, b 个线状刚体, c 个非线状刚体的体系, 初始自由度为 $s = 3a + 5b + 6c$. 加上 r 个约束. 那么自由度和广义坐标的个数都是 $f = s - r$, 不妨设广义坐标为 $q_1, q_2 \dots q_n$. 再施加若干主动力. 这些主动力要么作用于质点上, 要么作用于刚体的某点上. 而约束将能给出两组不同表述的方程:

$$\mathbf{r}_k = \mathbf{r}_k(q_i, t) \quad i = 1, 2 \dots f$$

$$f_j(x_k, y_k, z_k, t) = 0 \quad j = 1, 2 \dots r$$

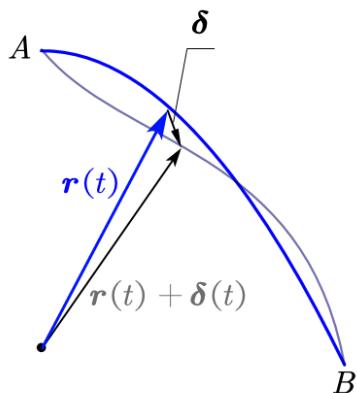


图 4.16: 实位移与虚位移

不同点在于, 现在根本不需要以上方程不含 t , 下面的过程都是照推无误的. 现在对于这样的非稳定约束, 我们依然希望这样的约束不做虚功, 不过此时的虚位移该如何表述?

在任何一个实际发生的动力学过程中, 确定每个广义坐标的值就确定了在某一时刻的体系的位形, 确定每个广义坐标随着时间的演化实际上就确定了一个具体的过程. 而每一个质点在这样的具体的过程发生实际位移就是**实位移**(real displacement). 它们应该满足:

$$d\mathbf{r}_k = \sum_i \frac{\partial \mathbf{r}_k}{\partial q_i} dq_i + \frac{\partial \mathbf{r}_k}{\partial t} dt$$

$$\sum_k \left(\frac{\partial f_j}{\partial x_k} dx_k + \frac{\partial f_j}{\partial y_k} dy_k + \frac{\partial f_j}{\partial z_k} dz_k \right) + \frac{\partial f_j}{\partial t} dt = 0$$

那虚位移又以何种形式存在? 回忆光学中的费马原理¹¹我们就可以找到答案: 比如一个粒子做平抛运动从 A 到达 B . 那么蓝色的轨迹就是粒子的实位移 $\mathbf{r}(t)$, 但是如果我们设想一条并不符合动力学规律, 仅仅符合运动学约束的灰色轨迹, 其运动方程为 $\mathbf{r}'(t) = \mathbf{r}(t) + \boldsymbol{\delta}(t)$, 那么这一条新的运动相对原来的运动的偏移就是虚位移 $\boldsymbol{\delta}(t)$. 对应到之前的情形, 就是说在同一时间, 约束的方程对原来正确的符合动力学规律的运动和虚位移以后的运动是同样成立的:

$$\mathbf{r}_k = \mathbf{r}_k(q_i, t), \quad f_j(x_k, y_k, z_k, t) = 0$$

$$\mathbf{r}'_k = \mathbf{r}_k(q'_i, t), \quad f_j(x'_k, y'_k, z'_k, t) = 0$$

用下式减上式, 就得到了:

$$\delta\mathbf{r}_k = \sum_i \frac{\partial \mathbf{r}_k}{\partial q_i} \delta q_i, \quad \sum_k \left(\frac{\partial f_j}{\partial x_k} \delta x_k + \frac{\partial f_j}{\partial y_k} \delta y_k + \frac{\partial f_j}{\partial z_k} \delta z_k \right) = 0$$

这和 t 作为常数的稳定约束对运动的限制才是一致了. 下一步, 自然时要求约束力不能做虚功, 注意, 理想的却不确定的约束情形中, 约束力是可以做实功的, 从而体系能量一般都不会守恒. 总之, 所谓理想约束, 依然指的是:

$$R_{kx} = \lambda \frac{\partial f_j}{\partial x_k}, \quad R_{ky} = \lambda \frac{\partial f_j}{\partial y_k}, \quad R_{kz} = \lambda \frac{\partial f_j}{\partial z_k}$$

$$\forall \delta\mathbf{r}_k, \delta W_R = \sum_k \mathbf{R}_k \cdot \delta\mathbf{r}_k = 0$$

最后只需解决主动力如何描述, 最终重要的其实是主动力造成的广义力. 也就是说我们计算以下量并以此描述所有主动力造成的“效果”:

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

¹¹它与分析力学的理论是同源的.

4.5.1 用广义坐标表示能量

一个体系最终会具有的能量,主要包括动能和势能. 我们分别来研究两者:

动能

一个体系如果还包含刚体,那么其实总动能应当体现为:

$$T = \sum_k \frac{1}{2} m_k \mathbf{v}_k^2 + \sum_l \frac{1}{2} \boldsymbol{\omega}_l \cdot \mathbb{I}_l \cdot \boldsymbol{\omega}_l$$

这里不可避免地需要使用到以后引入的刚体的惯量张量 \mathbb{I}_l 的概念. 为了简化问题,我们可以考虑把刚体也抽象为带约束的质点系,那么只需要写出:

$$T = \sum_k \frac{1}{2} m_k \mathbf{v}_k^2$$

就注意表达任意复杂体系的动能函数. 现在我们来进行广义坐标化. 利用上一小节给出的:

$$d\mathbf{r}_k = \sum_i \frac{\partial \mathbf{r}_k}{\partial q_i} dq_i + \frac{\partial \mathbf{r}_k}{\partial t} dt$$

等号两侧同时除 dt , 得到质点速度的广义坐标表达法:

$$\mathbf{v}_k = \sum_i \frac{\partial \mathbf{r}_k}{\partial q_i} \dot{q}_i + \frac{\partial \mathbf{r}_k}{\partial t}$$

带回动能表达式, 我们得到:

$$T = \sum_i \sum_j \sum_k \frac{1}{2} m_k \frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial q_j} \dot{q}_i \dot{q}_j + \sum_i \sum_k m_k \frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial t} \dot{q}_i + \sum_k \frac{1}{2} m_k \left(\frac{\partial \mathbf{r}_k}{\partial t} \right)^2$$

这就是说, 动能可以分解广义速度 \dot{q}_i 的二次齐次项 T_2 , 一次齐次项 T_1 , 和随时间变的常数项 T_0 的和:

$$T = T_2 + T_1 + T_0$$

$$T_2 = \sum_i \sum_j A_{ij} \dot{q}_i \dot{q}_j, \quad A_{ij} = \frac{1}{2} \sum_k m_k \frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial q_j}$$

$$T_1 = \sum_i B_i \dot{q}_i, \quad B_i = \sum_k m_k \frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial t}$$

$$T_0 = C, \quad C = \frac{1}{2} \sum_k m_k \left(\frac{\partial \mathbf{r}_k}{\partial t} \right)^2$$

如果体系包含的约束全都是稳定约束,那么一切关于 t 的偏导数消失, 动能就只剩下二次齐次项了, 即 $T = T_2$.

我们再定义两个量,一个是与广义坐标 q_i 相伴相生的那些广义速度, 广义加速度, 广义力同系列的最后一个量: **广义动量**(general momentum). 它的定义始终为:

$$p_i = \frac{\partial T}{\partial \dot{q}_i}$$

也就是把实际的动能看作 \dot{q}_i 的至多二次多项式,而把下面的偏导数定义为与第 i 广义坐标 q_i 共轭的广义动量:

$$p_i = \frac{\partial}{\partial \dot{q}_i} \left(\sum_i \sum_j A_{ij} \dot{q}_i \dot{q}_j + \sum_i B_i \dot{q}_i + C \right) = 2 \sum_j A_{ij} \dot{q}_j + B_i$$

第二个量,在数学上称作勒让德变换(Legendre transform),我们定义一个动能的变式:

$$U = \sum_i p_i \dot{q}_i - T$$

很容易验证,在以上三项分解法下,这个变式的动能其实为:

$$U = T_2 - T_0$$

广义动量和变式动能的意义何在?至少从定义上考虑, $p_i dq_i$ 将会是动能的实际改变中的一项,完整的写法为:

$$dT = \sum_i \frac{\partial T}{\partial q_i} dq_i + \sum_i \frac{\partial T}{\partial \dot{q}_i} d\dot{q}_i + \frac{\partial T}{\partial t} dt$$

三项中第二项就是 $p_i dq_i$. 而外力做功中的对应项为 $Q_i dq_i$, 虽然两者不存在直接的等量关系,但是它们至少是同量纲的. 现在即使让两者相等也得不到任何有意思的结果. 这是因为这个动能函数还需要修正. 如果考虑变式动能 U ,根据定义,有:

$$dU = - \sum_i \frac{\partial T}{\partial q_i} dq_i + \sum_i \dot{q}_i dp_i - \frac{\partial T}{\partial t} dt$$

乍一看这样的做法找不到经典力学中的对应,其实不然:经典力学中根本就不会有非稳定约束,从而 $T = U = T_2$. 所以在更普遍的情况下,用 U 在某些场合代替原来的 T 也是可以理解的. 最关键是,如果设想让外力在 q_i 上做功转化为上式第二个求和项中对应项(注意下式是错误的结论,仅仅是一种在特殊情况下成立的设想):

$$Q_i dq_i = \dot{q}_i dp_i = \frac{dq_i dp_i}{dt} = \dot{p}_i dq_i \Rightarrow Q_i = \dot{p}_i$$

这就是说,广义力的效果其实就是在改变对应自由度上的广义动量.这个说法后面却可以说明是对的,虽然上式不对,它漏项了.

势能

之前,我们简单地把体系中质点和刚体的受力二分为约束力(即被动力)和主动力.现在我们要重新观察众多种类的主动力中,保守力的可能性.有时候,体系内部的部分之间具有势能,带来非约束力的内力.有时候体系受到的外力也可以由势能来描述.多个势能函数直接求和,自然可以同时产生多个对应的保守力.从而我们定义,如果体系的所有主动力都有对应的势能,即都是保守力,那么体系称作**保守**(conserved)的.也就是说,主动力全都由势能函数 $V(\mathbf{r}_k)$ 描述:

$$\mathbf{F}_k = - \frac{\partial V}{\partial x_k} \mathbf{e}_x - \frac{\partial V}{\partial y_k} \mathbf{e}_y - \frac{\partial V}{\partial z_k} \mathbf{e}_z = - \nabla_k V$$

但是旋即便发现,如果用广义坐标表示 \mathbf{r}_k 并最终表示势能,在约束为非稳定约束情形下,势能将变成显含时间的形式 $V(q_i, t)$. 所以我们从一开时就默认最初势能也可以含有 t ,即 $V(\mathbf{r}_k, t)$. 对应的例子比如,在质量变化的太阳 $M(t)$ 外距离为 r 的地球 m 的势能,就是:

$$V(\mathbf{r}, t) = - \frac{GM(t)m}{r}$$

势能,以后都要理解为势能函数,作为一个数它不一定与动能的和守恒,但是上面这个势能至少可以告诉我们,地球的受力就是这个函数的负梯度:

$$\mathbf{F}(\mathbf{r}, t) = - \nabla V = - \frac{GM(t)m}{r^2} \mathbf{e}_r$$

故现在,我们直接总结,保守体系的主动力部分,由势能函数 $V(q_i, t)$ 来描述.十分容易验证的一点是,此时主动动力在广义坐标上的广义力,也恰好可以通过负偏导数来计算:

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

但是一定要注意,由于势能本身就在随时间变化,故主动力做的实功并不会等于势能的降低,而是有:

$$-\mathrm{d}V = \sum_i Q_i \mathrm{d}q_i - \frac{\partial V}{\partial t} \mathrm{d}t$$

反倒是在虚位移情形下,有

$$-\delta V = \sum_i Q_i \delta q_i$$

4.5.2 拉格朗日方程

至此,一个由广义坐标描述的,反映广义力是如何改变广义动量的严谨正确的最终方程就呼之欲出,只差临门一脚了. 我们先不做任何体系上的主动力是否保守的假设,在一个十分普遍的问题中,每一个自由度 q_i 上的广义力为 Q_i . 最初我们只能知道每一个质点都符合牛顿定律:

$$\mathbf{R}_k + \mathbf{F}_k = m_k \ddot{\mathbf{r}}_k$$

如何从这个定律抽象到广义坐标呢? 核心的问题是要解决 $\ddot{\mathbf{r}}_k$ 的广义坐标化. 为了将求导的阶数降低, 我们构造真实路径 $\mathbf{r}_k(t)$ 周围的虚位移以后的路径 $\mathbf{r}_k(t) + \delta \mathbf{r}_k(t)$. 一如既往地, 虚位移要符合约束条件. 而在牛顿定律两侧点乘虚位移, 再来求和计算虚功和相伴产生的项:

$$\sum_k \mathbf{R}_k \cdot \delta \mathbf{r}_k + \sum_k \mathbf{F}_k \cdot \delta \mathbf{r}_k = \sum_k m_k \ddot{\mathbf{r}}_k \cdot \delta \mathbf{r}_k$$

第一项是约束力的虚功, 由于是理想约束所以等于零. 第二项才是真正的主动力虚功, 用广义力和广义位移表示为 $Q_i \delta q_i$ 的求和. 第三项就是尚未算过的, 这样处理也恰好提供了去除二阶导数的便利, 从而方便广义坐标化:

$$\begin{aligned} \ddot{\mathbf{r}}_k \cdot \delta \mathbf{r}_k &= \sum_i \ddot{\mathbf{r}}_k \cdot \frac{\partial \mathbf{r}_k}{\partial q_i} \delta q_i \\ \ddot{\mathbf{r}}_k \cdot \frac{\partial \mathbf{r}_k}{\partial q_i} &= \frac{\mathrm{d}}{\mathrm{d}t} \left(\dot{\mathbf{r}}_k \cdot \frac{\partial \mathbf{r}_k}{\partial q_i} \right) - \dot{\mathbf{r}}_k \cdot \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathbf{r}_k}{\partial q_i} \\ \sum_k m_k \ddot{\mathbf{r}}_k \cdot \delta \mathbf{r}_k &= \sum_i \left[\frac{\mathrm{d}}{\mathrm{d}t} \left(\sum_k m_k \dot{\mathbf{r}}_k \cdot \frac{\partial \mathbf{r}_k}{\partial q_i} \right) - \sum_k m_k \dot{\mathbf{r}}_k \cdot \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathbf{r}_k}{\partial q_i} \right] \delta q_i \end{aligned}$$

现在我们暂时把动能写作:

$$T = \sum_k \frac{1}{2} m_k \dot{\mathbf{r}}_k^2$$

而其中 $\dot{\mathbf{r}}_k$ 视作 q_i, \dot{q}_i 函数, 有下列结论成立:

$$\begin{aligned} \frac{\partial T}{\partial q_i} &= \sum_k m_k \dot{\mathbf{r}}_k \cdot \frac{\partial \dot{\mathbf{r}}_k}{\partial q_i} \\ \frac{\partial T}{\partial \dot{q}_i} &= \sum_k m_k \dot{\mathbf{r}}_k \cdot \frac{\partial \dot{\mathbf{r}}_k}{\partial \dot{q}_i} \\ \dot{\mathbf{r}}_k &= \sum_j \frac{\partial \mathbf{r}_k}{\partial q_j} \dot{q}_j + \frac{\partial \mathbf{r}_k}{\partial t} \quad \Rightarrow \quad \frac{\partial \dot{\mathbf{r}}_k}{\partial \dot{q}_i} = \frac{\partial \mathbf{r}_k}{\partial q_i} \\ \frac{\partial \dot{\mathbf{r}}_k}{\partial q_i} &= \sum_j \frac{\partial^2 \mathbf{r}_k}{\partial q_i \partial q_j} \dot{q}_j + \frac{\partial^2 \mathbf{r}_k}{\partial q_i \partial t} \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathbf{r}_k}{\partial q_i} = \frac{\partial \dot{\mathbf{r}}_k}{\partial q_i} \end{aligned}$$

后两式带入前两式并对比之前待化简的项, 发现:

$$\sum_k m_k \ddot{\mathbf{r}}_k \cdot \delta \mathbf{r}_k = \sum_i \left[\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial T}{\partial \dot{q}_i} \right) - \frac{\partial T}{\partial q_i} \right] \delta q_i$$

所以原来的牛顿定律, 最后就变为:

$$Q_i = \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_i} \right) - \frac{\partial T}{\partial q_i}$$

这就是著名的**拉格朗日方程**(Lagrange equation) 的原型. 也可以写成:

$$Q_i + \frac{\partial T}{\partial q_i} = \dot{p}_i$$

所以产生第 i 自由度上广义动量变化的原因除了广义力, 还有一个动能本身对该自由度变化偏导数的因子. 可以称做“几何力”, 它的原因在于广义坐标与广义速度之间的耦合.

一个典型的例子比如在平面极坐标中研究粒子的运动, 以矢径 r 和极角 θ 为两自由度. 矢量力学给出粒子受力 F_r, F_θ . 但是粒子广义力为:

$$Q_r = F_r, \quad Q_\theta = rF_\theta$$

写出动能:

$$T = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2)$$

从而两个广义动量:

$$p_r = m\dot{r}, \quad p_\theta = mr^2\dot{\theta}$$

后者即角动量. 注意到在半径方向还有几何力:

$$\frac{\partial T}{\partial r} = m\dot{\theta}^2r$$

故拉格朗日方程, 就变为:

$$F_r + m\dot{\theta}^2r = \frac{d}{dt}(m\dot{r}), \quad rF_\theta = \frac{d}{dt}(mr^2\dot{\theta})$$

这与矢量力学给出了同样的结果.

在体系为保守体系时, 拉格朗日方程还可以获得一种更普遍的形式. 只需要注意到势能函数, 由于不能含有广义速度, 与广义力之间符合以下关系:

$$Q_i = -\frac{\partial V}{\partial q_i} + \frac{d}{dt} \left(\frac{\partial V}{\partial \dot{q}_i} \right)$$

这样就得到了:

$$\frac{d}{dt} \frac{\partial(T - V)}{\partial \dot{q}_i} - \frac{\partial(T - V)}{\partial q_i} = 0$$

一件令人震惊的事情就发生了: 动能势能不是做和, 而是做差, 以函数的形式组合为了一个矢量力学中从未见过的新函数 $L = T - V$. 它就是著名的**拉格朗日量**(Lagrangian). 于是我们就遇到了标准的拉格朗日方程:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0$$

对于 f 自由度系统拉格朗日方程也恰恰是 f 个, 恰好能把问题解决.

在以上过程中, 我们对势能的性质放松到了可以含有时间. 这一步可以进一步放松. 在有的场合下, 拉格朗日量也是由两项构成, 一项是常规意义上的动能, 但是另一项势能也是含有速度的. 这种势能函数叫做**广义势能**(general potential energy). 一种典型的情形是带点粒子在电磁场中运动的情形, 此时动能势能为:

$$T = \frac{1}{2}m\dot{r}^2$$

$$V = q(\varphi + \dot{r} \cdot \mathbf{A})$$

其中 φ 是往往与电场对应的标势, \mathbf{A} 则是往往与磁场对应的矢势. 在这样的场合下, 广义力的计算也要以下式为准:

$$Q_i = -\frac{\partial V}{\partial q_i} + \frac{d}{dt} \left(\frac{\partial V}{\partial \dot{q}_i} \right)$$

容易得到, 最终粒子的拉格朗日方程, 化简以后就是:

$$m\ddot{\mathbf{r}} = q \left[-\nabla \varphi + \frac{\partial \mathbf{A}}{\partial t} + \dot{\mathbf{r}} \times (\nabla \times \mathbf{A}) \right]$$

故只要把电场, 磁场视作:

$$\mathbf{E} = -\nabla \varphi + \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}$$

就能给出正确的洛伦兹力公式和牛顿定律. 以往在经典力学中与势能毫无关联的磁场力, 分析力学下也是广义势能的结果.

最后我们来探讨守恒量的问题. 用动能减势能这个操作固然不会得到什么守恒量, 经典力学告诉我们保守系统守恒量应当是两者加起来, 但是它导致了拉格朗日方程, 以它为基础, 我们能先后得到两个可能的守恒量.

一是在可能含有广义势的普遍情况下, 广义动量被重新定义为:

$$p_i = \frac{\partial L}{\partial \dot{q}_i}$$

直接从拉格朗日方程可以看出, 如果有

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

也就是拉格朗日量中不显含某个坐标 q_i , 这样的坐标称作**循环坐标**(cyclic coordinate), 那么这个自由度上的广义动量就不会随着时间改变. 也就是“坐标平移不变导致动量守恒”.

二是, 我们还记得之前引入过变式动能 U , 那么其实 $U + V$ 才是对应经典力学中的总能量. 或者, 更普遍的, 对任意拉格朗日量, 定义**哈密顿量**(Hamiltonian):

$$H = \sum_i p_i \dot{q}_i - L$$

那么直接计算其对时间导数就可以得到:

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t}$$

从而如果拉格朗日量中不显含时间 t^{12} . 那么哈密顿量就会守恒. 也就是“时间平移不变导致能量守恒”.

4.5.3 再论冲击问题

利用拉格朗日方程也可以求解碰撞问题. 其实就是考虑拉格朗日方程的积分形式:

$$p'_i - p_i = \int \left(\frac{\partial T}{\partial q_i} + Q_i \right) dt$$

其中 p_i, p'_i 就是碰前碰后的广义动量, 考虑到实际碰撞过程中几何力其实都是一些仅仅取决于广义坐标和广义速度的函数, 过程中最多突变, 但不会无限大, 所以在极短时间内积分是趋于零的, 而广义力积分给出了广义冲量 I_i , 它也正是各个主动力冲量 (一般也就一个) \mathbf{I}_k 计算的结果:

$$I_i = \sum_k \mathbf{I}_k \cdot \frac{\partial \mathbf{r}_k}{\partial q_i}$$

¹²即使约束不稳定这也有可能发生.

4.6 平衡态稳定性

4.6.1 一自由度体系

4.6.2 多自由度体系

第五章 简谐振动

5.1 方程与谐振

5.1.1 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 $Q_0 = 0$:

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

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

5.1.2 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:

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

5.1.3 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 $v^2 \neq \Re(\tilde{v}^2)$

5.1.4 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.

5.2 阻尼振动与受迫振动

5.2.1 阻尼振动

阻尼振动,狭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:

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

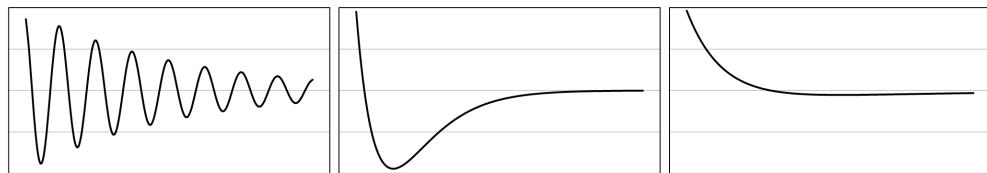


图 5.1: 欠阻尼, 临界阻尼与过阻尼

- **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:

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

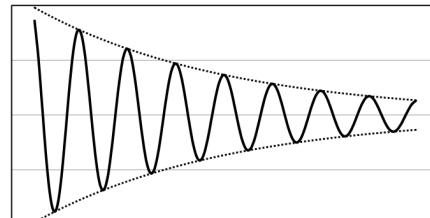


图 5.2: 小阻尼下的振幅衰减

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.

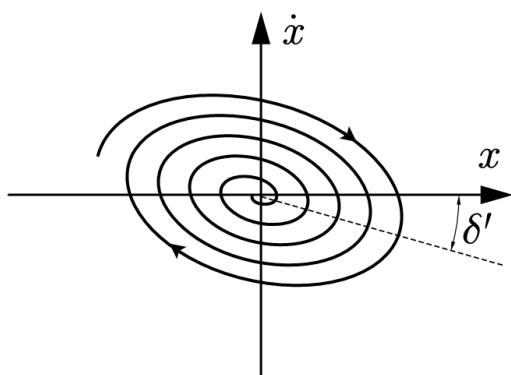


图 5.3: 阻尼振动的相图

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:

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

在这个解下,其速度为:

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

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:

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

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.

5.2.2 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5.4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:

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

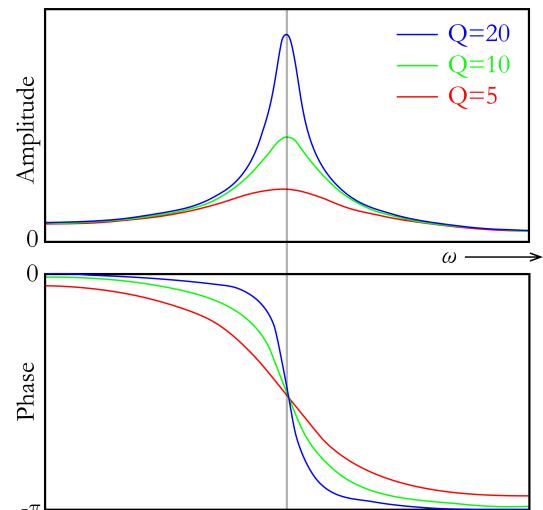


图 5.4: 幅频, 相频特性

$$\varphi \approx -\pi$$

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5.4.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:

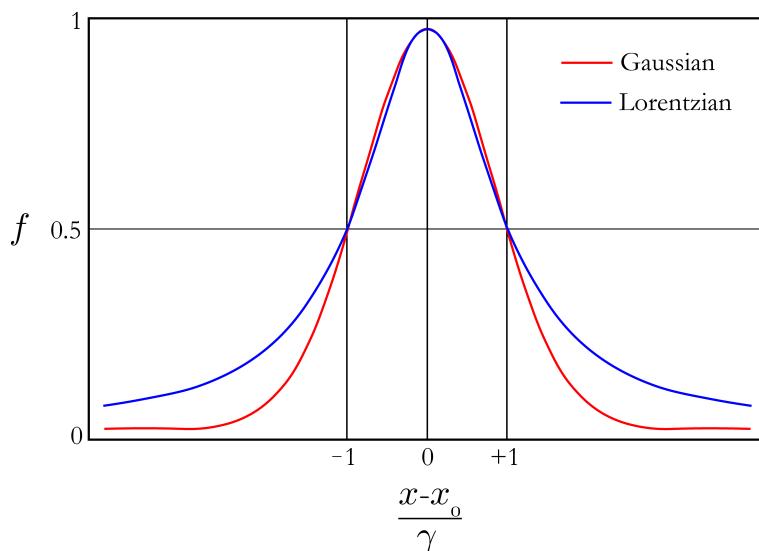
$$Q = \frac{\omega_0 \bar{E}}{W}$$

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 5.5: 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:

$$I(x) = I_0 \left/ \left[(1 - x^2)^2 + \frac{x^2}{Q^2} \right] \right. \Rightarrow I_{\max} \approx I_0$$

$$I = \frac{I_{\max}}{2} \Rightarrow 1 - x^2 \approx \pm \frac{x}{Q}$$

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:

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

5.3 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5.6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5.7中, 两个摆直接悬挂在起. 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:

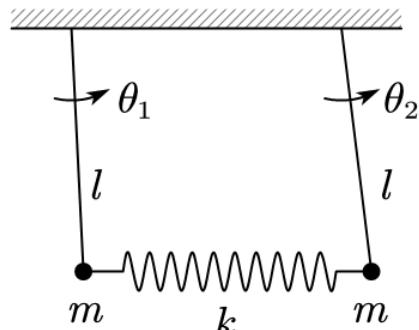


图 5.6: 耦合摆

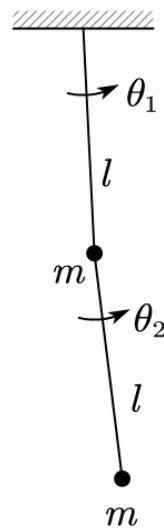


图 5.7: 双摆

5.3.1 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.

5.3.2 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:

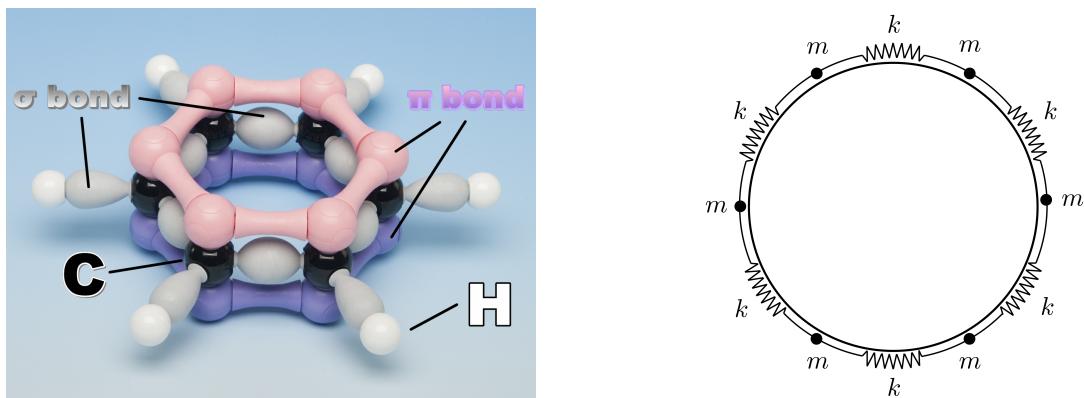


图 5.8: 苯: 6 原子环问题

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:

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

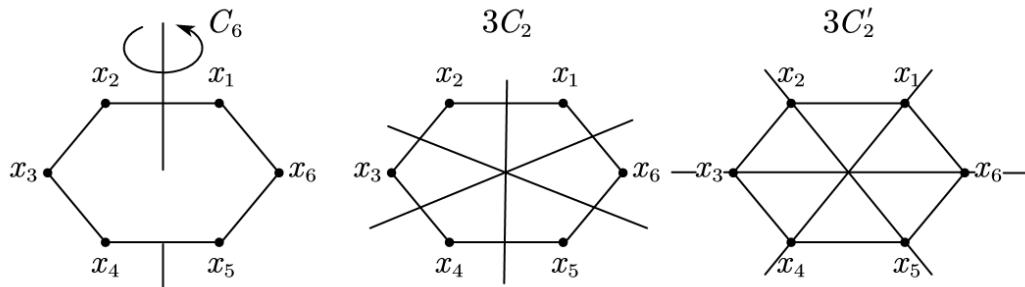


图 5.9: 6 原子环对称性

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:

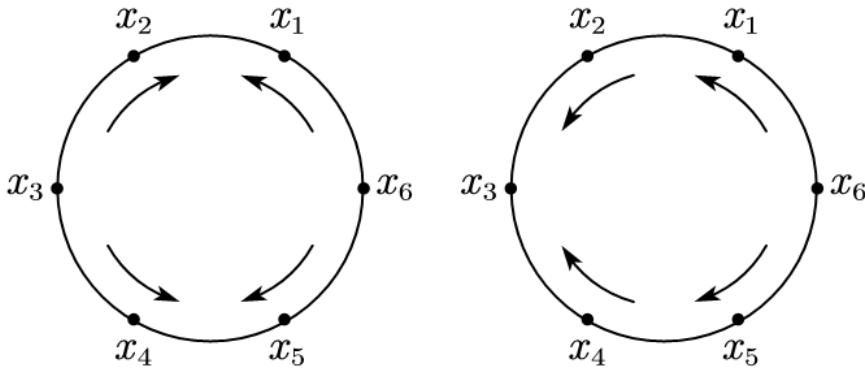


图 5.10: 二重简并的振动模式

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:

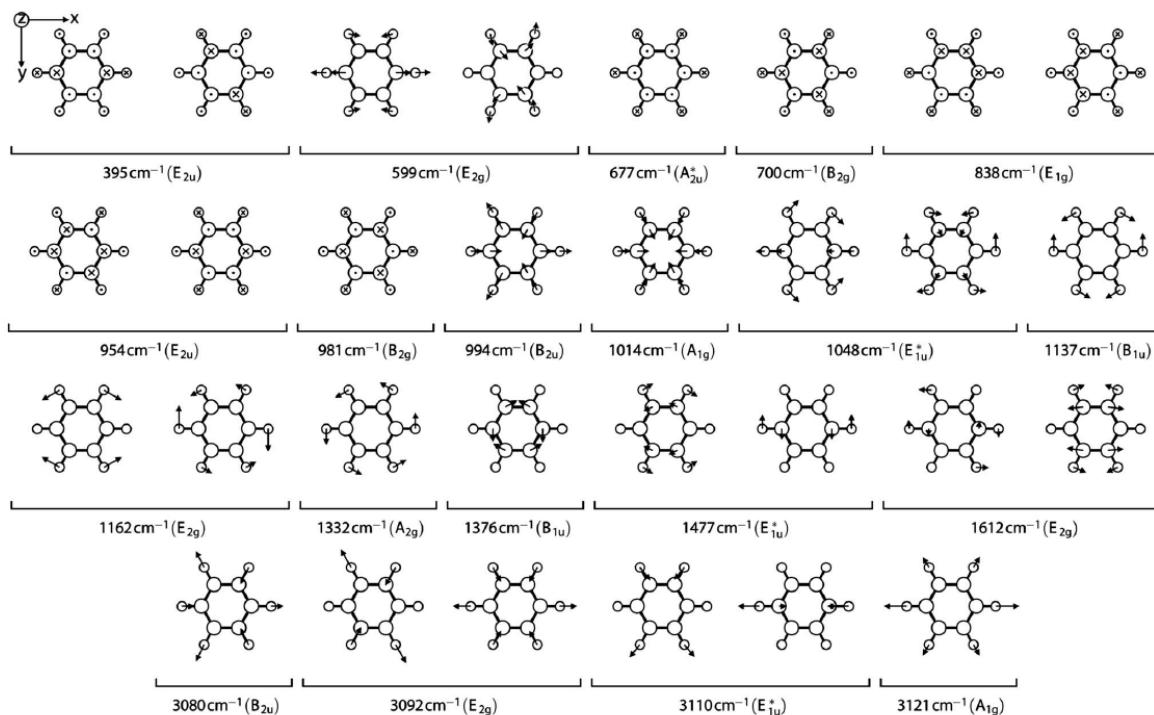


图 5.11: 苯分子的完整振动模式

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:

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

5.4 非线性摄动 *

5.5 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.

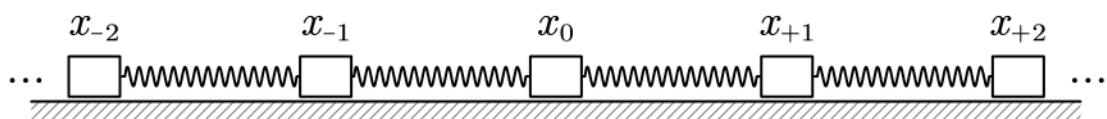


图 5.12: 一维原子链

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 5.13.

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.

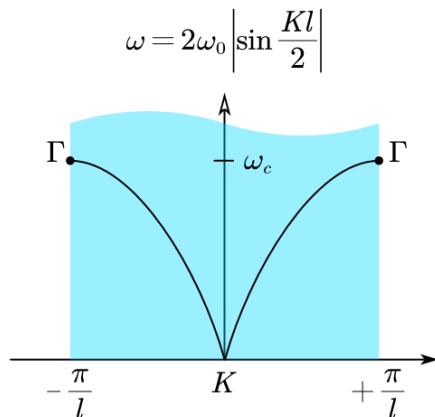


图 5.13: 色散关系

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 :

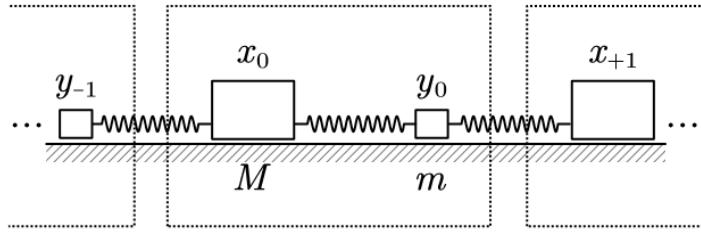


图 5.14: 双原子链

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:

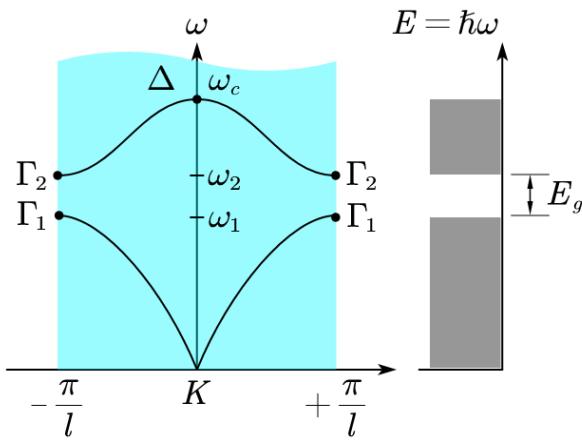
$$\omega \approx \frac{2\omega_0}{\sqrt{\beta}} \cdot \sqrt{1 + \beta \cos^2 \frac{Kl}{2}}$$

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5.15. 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.

第六章 万有引力

6.1 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:

6.1.1 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:

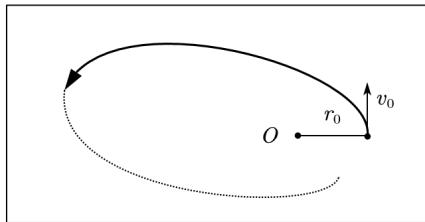
$$H = \frac{1}{2}m(\dot{r}^2 + \dot{\theta}^2 r^2) + V(r) = \text{Const.}$$

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:

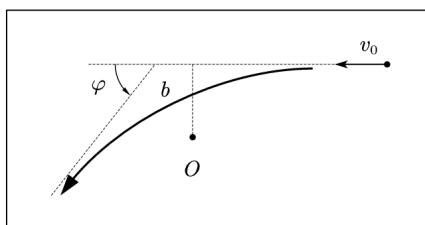
$$\dot{r}^2 + \dot{\theta}^2 r^2 = \frac{2[E - V(r)]}{m}$$

6.1.2 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 6.1: 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:

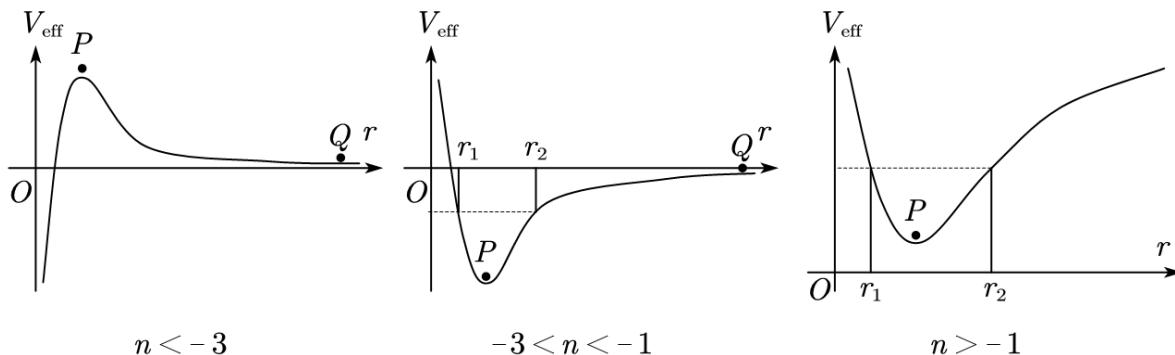


图 6.2: 三类幂律力

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6.2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6.2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.

6.2 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:

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

6.3 二体与潮汐

第七章 刚体

7.1 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:

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

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

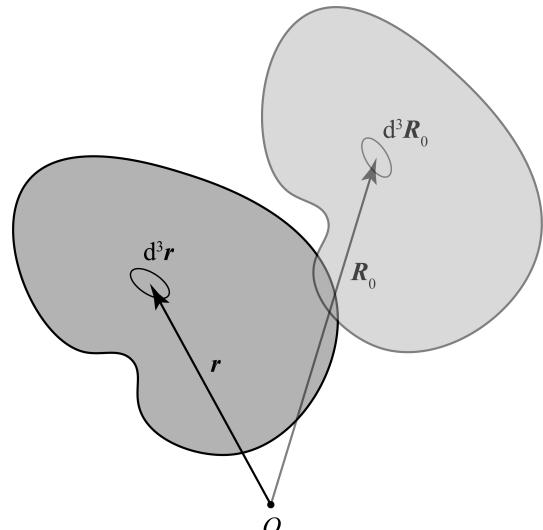


图 7.1: 刚体的描述

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:

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

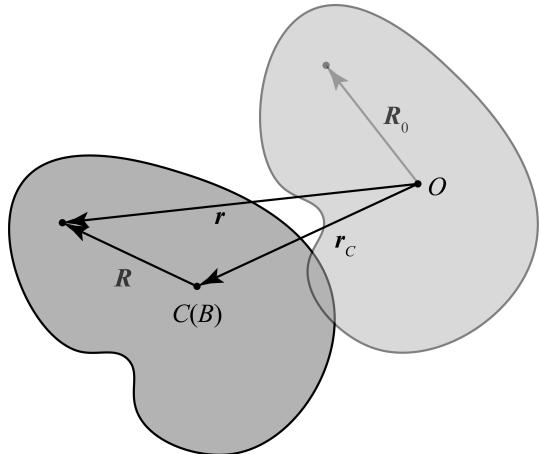


图 7.2: 基点法

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:

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

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7.3阐明.现在我们写出相对质心系能量角动量的定义:

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

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

7.2 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):

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

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

$$\sum_i M_i = I_C\beta$$

7.3 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:

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

那么这个映射一定是线性的.这是因为如果角速度为 $\boldsymbol{\omega} = \lambda_1\boldsymbol{\omega}_1 + \lambda_2\boldsymbol{\omega}_2$. 那么根据第一节给出的角动量计算方法:

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

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

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:

$$\mathbf{I} = I_x \mathbf{e}_x \mathbf{e}_x + I_y \mathbf{e}_y \mathbf{e}_y + I_z \mathbf{e}_z \mathbf{e}_z + I_{xy} \mathbf{e}_x \mathbf{e}_y + I_{xz} \mathbf{e}_x \mathbf{e}_z + I_{yx} \mathbf{e}_y \mathbf{e}_x + I_{yz} \mathbf{e}_y \mathbf{e}_z + I_{zx} \mathbf{e}_z \mathbf{e}_x = I_{zy} \mathbf{e}_z \mathbf{e}_y$$

其中 $\mathbf{e}_i \mathbf{e}_j$ 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:

$$\mathbf{e}_i \mathbf{e}_j \neq \mathbf{e}_j \mathbf{e}_i$$

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:

$$(\mathbf{e}_i \mathbf{e}_j) \cdot \mathbf{e}_k = \mathbf{e}_i (\mathbf{e}_j \cdot \mathbf{e}_k)$$

即如果 $j = k$ 则就是 \mathbf{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:

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

利用之前定义的二级矩:

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

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:

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

第八章 弹性体

8.1 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:

$$\begin{aligned} \boldsymbol{\delta}(d\mathbf{R}, t) &= \boldsymbol{\delta}(\mathbf{0}, t) + d\phi \times d\mathbf{R} \\ \nabla \boldsymbol{\delta} : & \begin{bmatrix} 0 & d\phi_z & -d\phi_y \\ -d\phi_z & 0 & d\phi_x \\ d\phi_y & -d\phi_x & 0 \end{bmatrix} \end{aligned}$$

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:

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

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

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8.1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 $\overset{\leftrightarrow}{T}_{ij}$ 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:

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

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

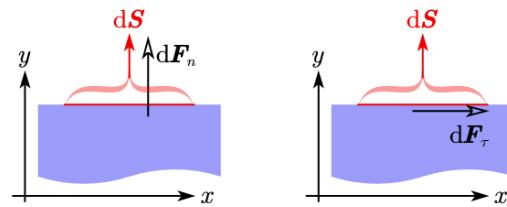


图 8.2: 正应力与剪应力

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

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8.2, 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 $\nu = \lambda/2(\lambda + \mu)$:

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

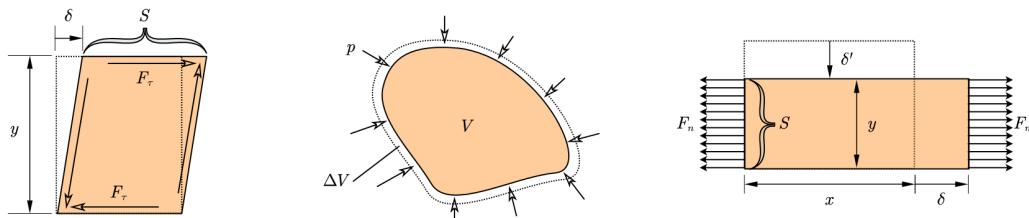


图 8.3: 剪变模量, 体弹性模量与杨氏模量

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.

8.2 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.

8.2.1 弹性棒

考虑一根弹性棒^{8.6},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:

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

8.2.2 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.

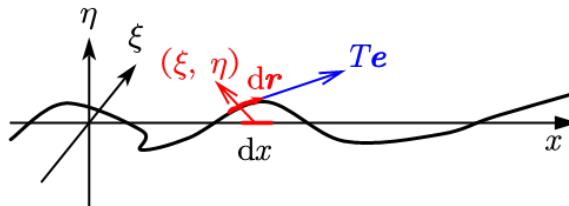


图 8.5: 弹性绳的建模

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.

8.2.3 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:

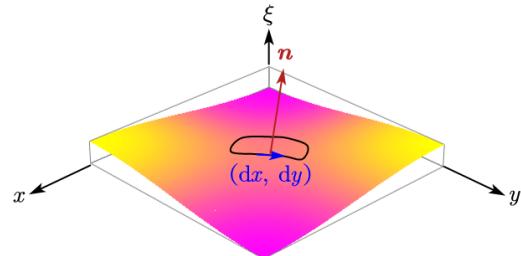


图 8.6: 弹性膜的建模

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

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:

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

8.2.4 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:

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

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:

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

8.3 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:

8.3.1 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.

8.3.2 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.

8.3.3 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。

8.3.4 再论格波

⁶少量 S 波有可能转化为 P 波继续向地核传播

第九章 流体

9.1 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

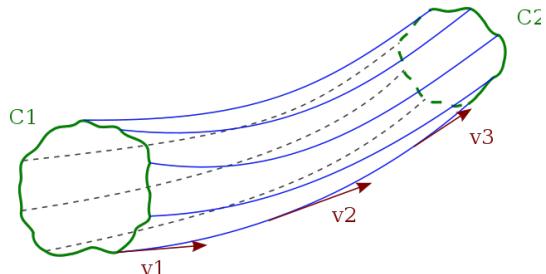


图 9.1: 流线与流管

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).

9.2 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

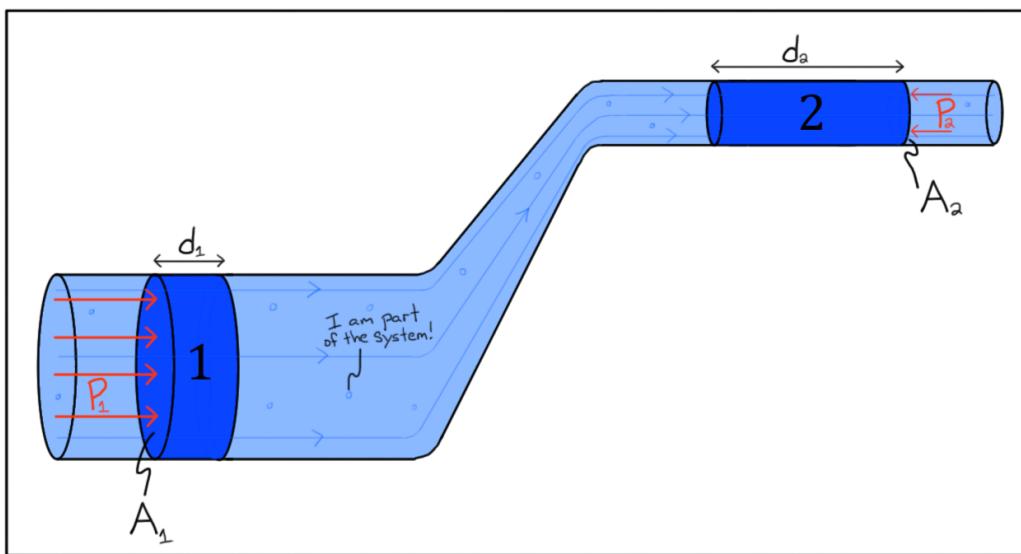


图 9.2: 贝努力方程

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:

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

9.3 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9.3. 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.

9.4 * 流体中的波

9.5 波的色散

第十章 热力学第一定律

10.1 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运输过程(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.

10.2 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.

10.2.1 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.

10.2.2 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.

10.2.3 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

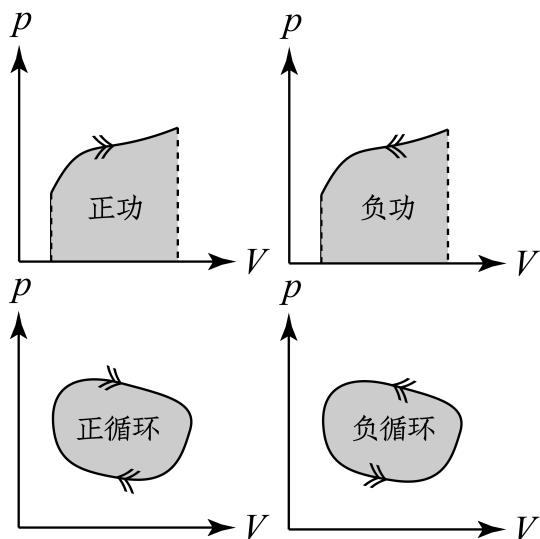


图 10.1: 准静态过程的做功

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.

10.2.4 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.

10.2.5 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 (Prigogine) 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.

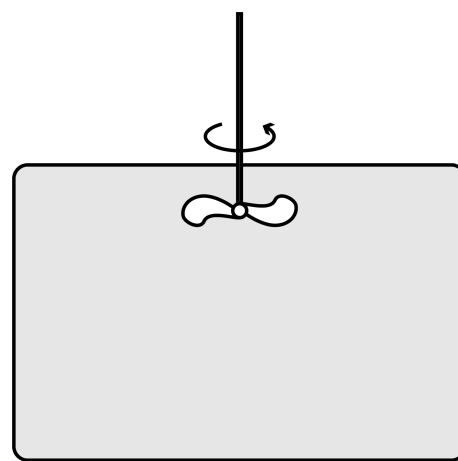


图 10.2: 搅动空气发热



图 10.3: 自组织现象: Bénard 对流与斑马身上的条纹

10.3 理想气体

10.3.1 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:

$$\bar{v} = \int_0^\infty v f(v)dv ; \quad \overline{v^2} = \int_0^\infty v^2 f(v)dv$$

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:

$$V(t) = \pi D^2 \cdot \sqrt{2}\bar{v}t$$

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:

$$U = \frac{f}{2}\nu RT$$

10.3.2 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.

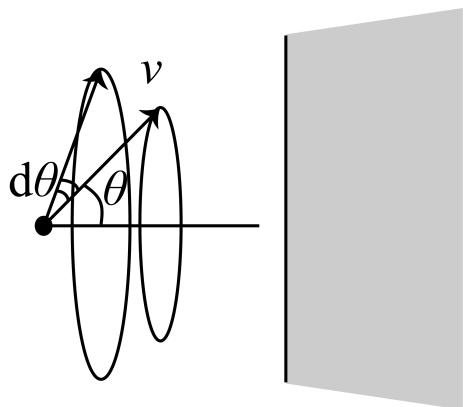


图 10.5: 粒子速度方向分布

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:

$$\begin{aligned} p &= \int_0^\infty \int_0^{\frac{\pi}{2}} n dS v \cos \theta dt \cdot f(v) dv \cdot \frac{\sin \theta d\theta}{2} \cdot 2mv \cos \theta / dS dt \\ &= nm \cdot \int_0^\infty v^2 f(v) dv \cdot \int_0^{\frac{\pi}{2}} \sin \theta \cos^2 \theta d\theta \\ &= \frac{1}{3} nm v^2 \\ &= \frac{2}{3} n \bar{v}_k \\ &= nkT \end{aligned}$$

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 $C_{mV} = \frac{f}{2} R$.

10.3.3 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_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:

$$\nu_i : \nu_j = \frac{n_i V}{N_A} : \frac{n_j V}{N_A} = n_i : n_j ; \quad \mu_i = N_A m_i$$

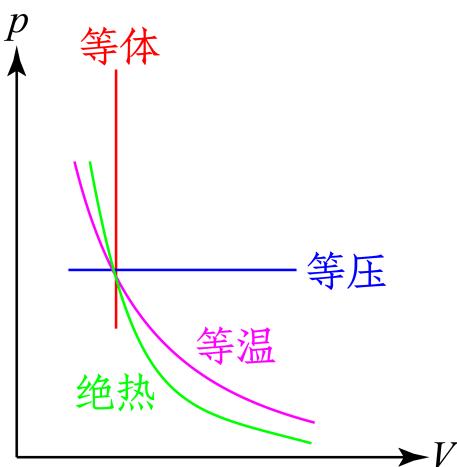
能量自由度数影响的是摩尔内能 u :

$$u = \nu \overline{C_{mV}} T = \sum u_i = \sum \nu_i C_{mVi} T \Rightarrow \overline{C_{mV}} = \sum x_i C_{mVi} = \sum \frac{x_i f_i}{2} R$$

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.

10.3.4 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10.2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 10.6 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)

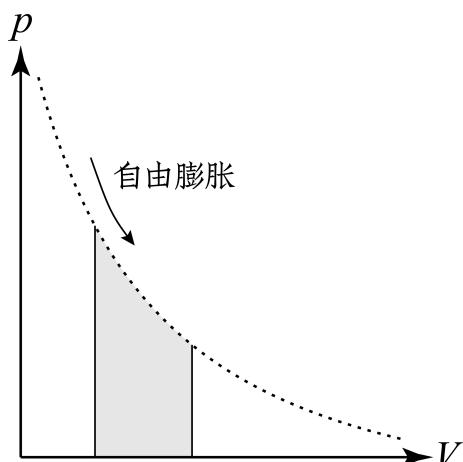
多方过程可以由过程方程 $pV^n = \text{Const.}$ 描述, 多方指数 $1 < n < \gamma$. 实际过程若发生缓慢, 可以近似处理为某种形式的多段多方过程. 多方过程的热容为:

$$C_{mn} = \left(\frac{1}{\gamma-1} - \frac{1}{n-1} \right) R$$

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 10.7: $p - V$ 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: 理想气体的自有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有膨胀过程中的状态, 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. 如果在某个时刻命自有膨胀结束, 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

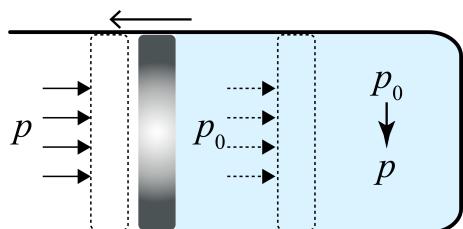


图 10.8: 向固定压强膨胀

考虑原平衡态压强为 p_0 的气体, 突然命外压变小为 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).

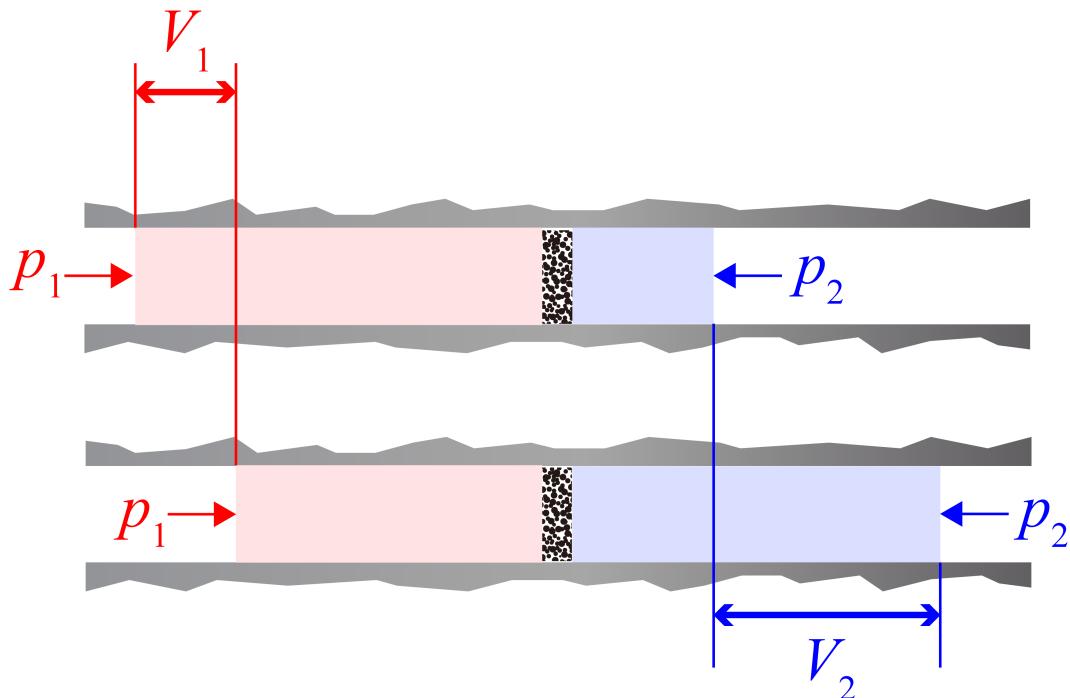


图 10.9: 多孔塞制节流阀

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.

10.4 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.

10.4.1 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:

$$p(z) = p(0) \cdot [\frac{T'(z)}{T(0)}]^{\frac{\gamma}{\gamma-1}}$$

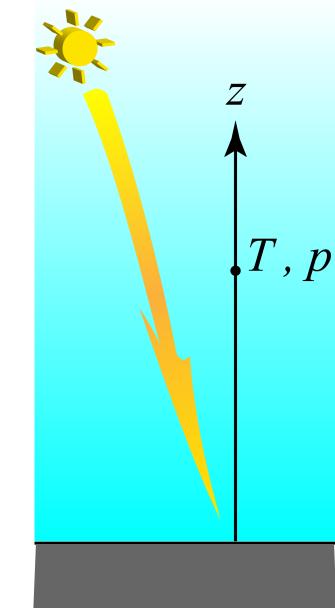


图 10.10: 地球大气

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.

10.4.2 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:

$$p + \frac{1}{2}\rho v^2 = \text{const.}$$

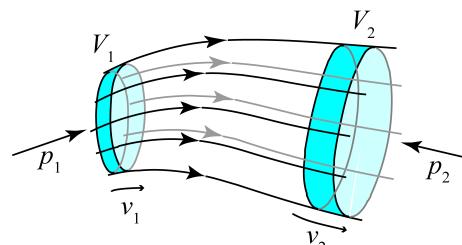


图 10.11: 气体沿一条流管流动

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:

$$p' = p + \frac{1}{2}\rho v^2 ; \quad h' = \frac{p'}{\rho}$$

10.4.3 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:

$$\rho = \frac{\mu p}{RT}$$

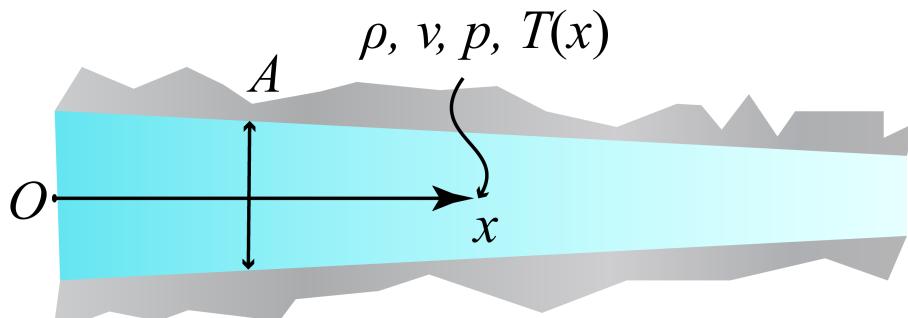


图 10.12: 截面积减小的管道内的气流

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:

$$\frac{\Delta A}{A} \ll 1$$

第十一章 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.

11.1 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.

11.1.1 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).

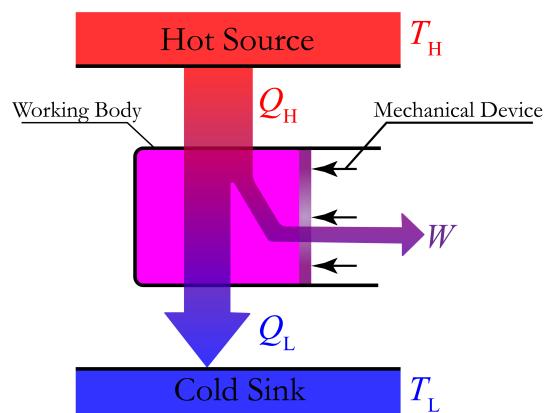


图 11.1: 热机模型

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 11.2: 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.

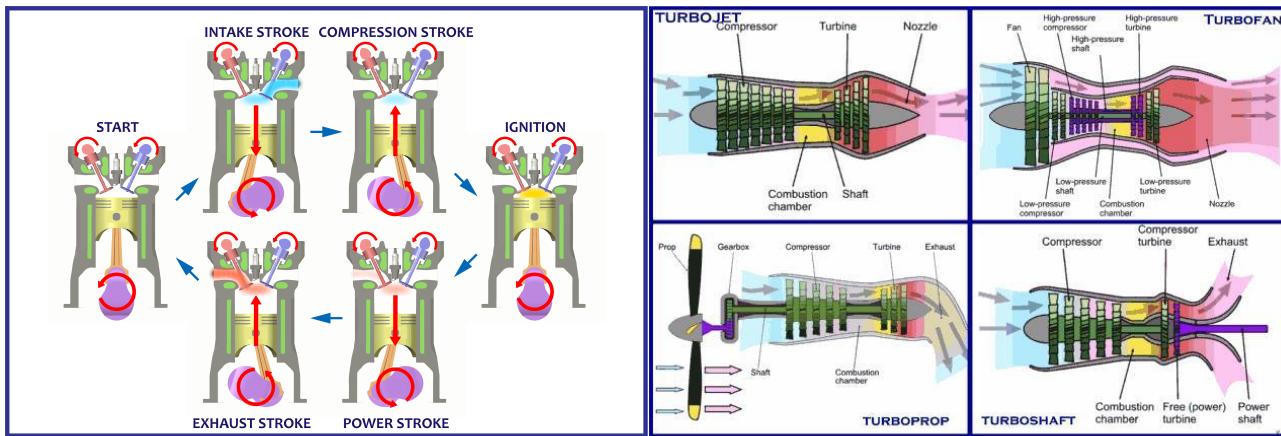


图 11.3: 左: 四冲程活塞内燃机; 右: 四类涡轮机

我们发现, 将实际的热机抽象为前图11.1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.

²火箭除外, 它需要自备燃料与氧化剂.

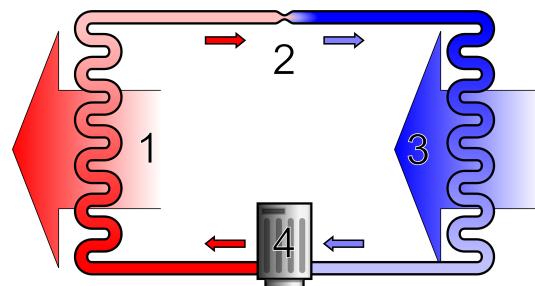


图 11.4: 热泵的普遍构成

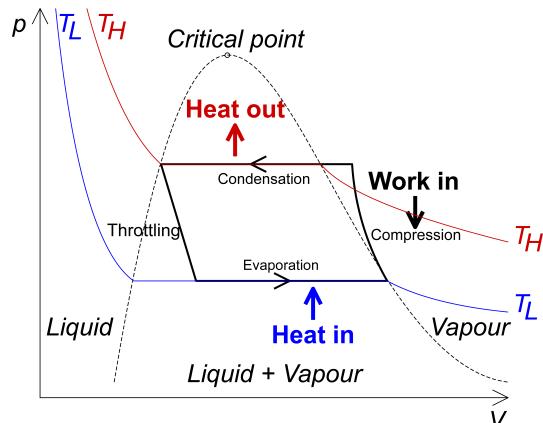


图 11.5: 热泵的工作循环

11.1.2 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)

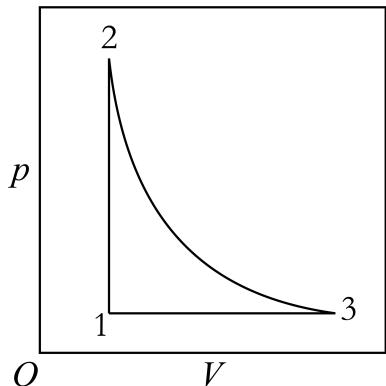


图 11.6: Lenoir cycle

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)

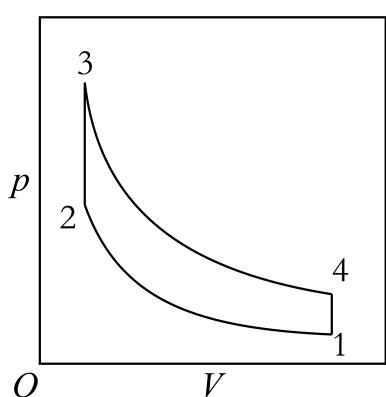


图 11.7: Otto cycle

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:

$$\eta = 1 - \frac{1}{\gamma} \cdot \frac{T_4 - T_1}{T_3 - T_2} = 1 - \frac{\alpha^\gamma - 1}{\gamma r^{\gamma-1} (\alpha - 1)}$$

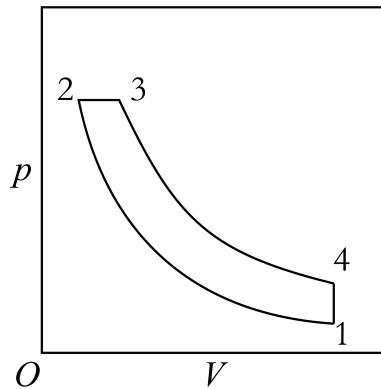


图 11.8: Diesel cycle

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:

$$\eta = 1 - \frac{T_1}{T_2} = 1 - r_p^{-\frac{\gamma-1}{\gamma}}$$

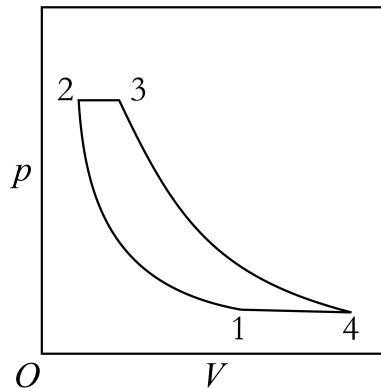


图 11.9: Brayton cycle

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中的实际热机。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。

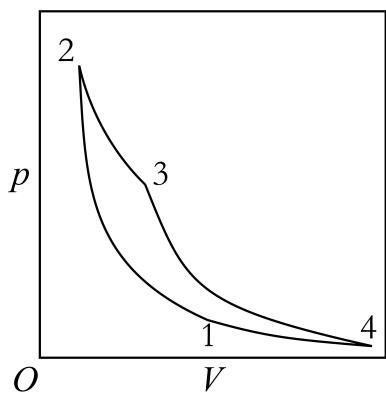


图 11.10: Carnot cycle

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：

$$\eta = 1 - \frac{T_L}{T_H}$$

11.2 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。

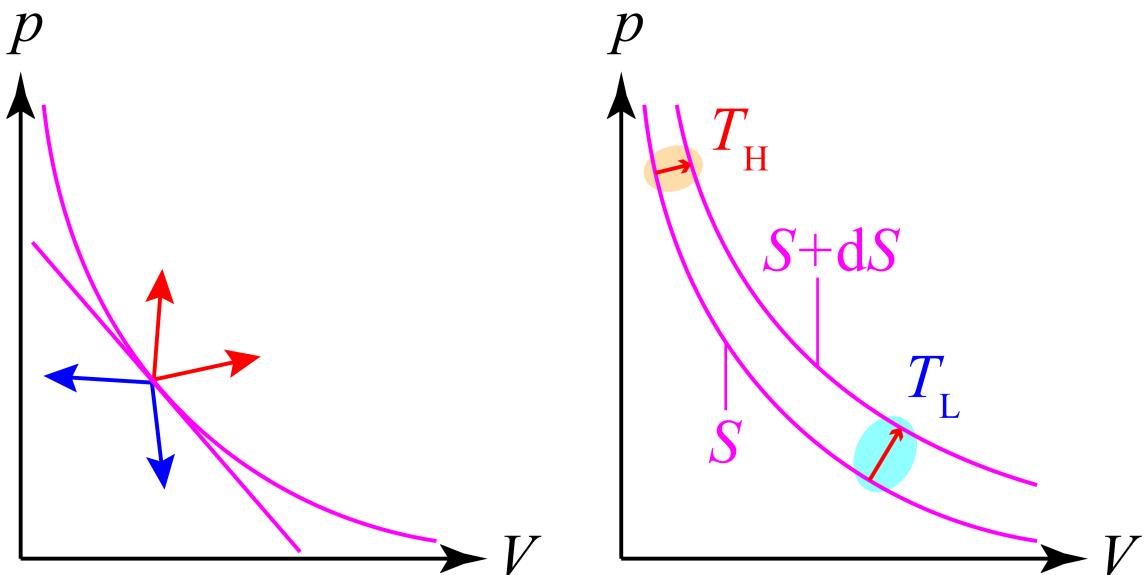


图 11.11: 左: 吸放热的元过程方向; 右: 两根等熵线间的“热量差”

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11.11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:

$$Q_H = T_H dS \quad ; \quad Q_L = T_L dS$$

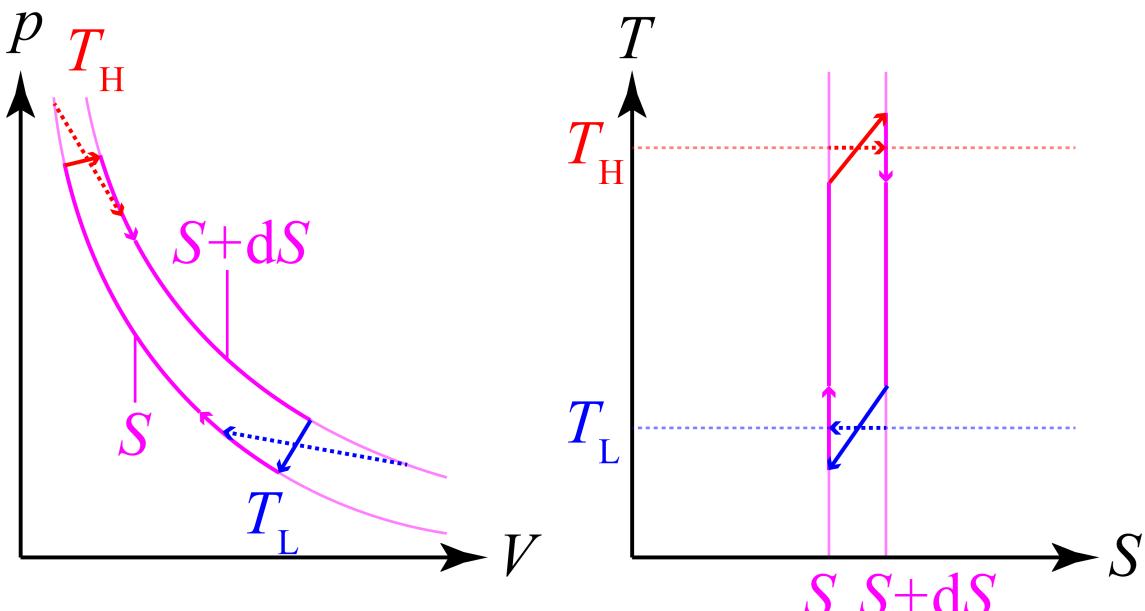


图 11.12: 示功图与示热图

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.

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

11.3 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:

$$\eta = 1 - \frac{Q_L}{Q_H}$$

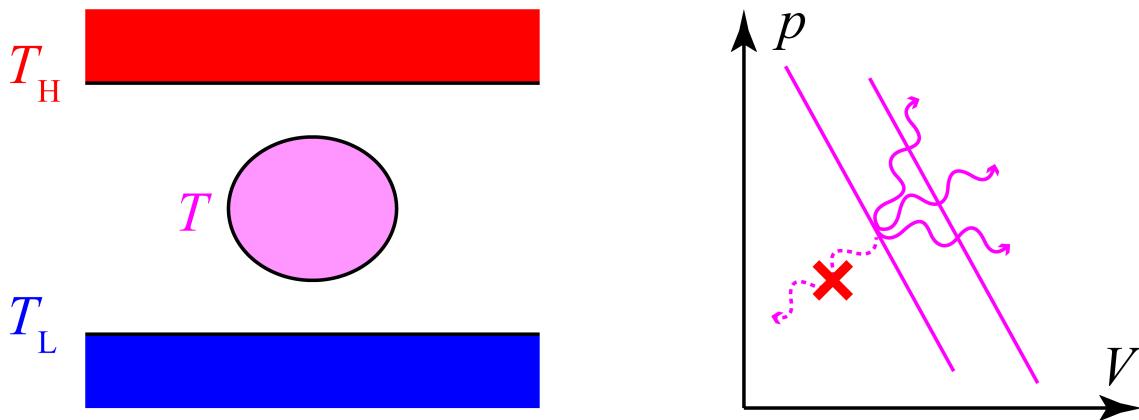


图 11.13: 熵增原理

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.

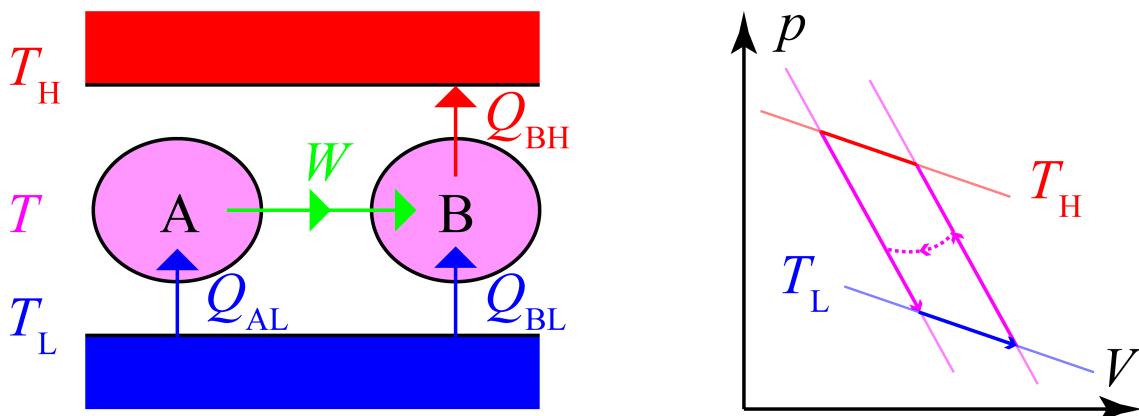


图 11.14: 三种不可逆过程的联系

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.

11.4 熵的计算

11.4.1 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.

11.4.2 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:

$$S = C \ln \frac{T}{T_0}$$

11.4.3 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:

$$\Delta S = C \ln \frac{T + \Delta T}{T} \approx \frac{C\Delta T}{T} = \frac{\Delta U}{T} \approx \Delta \frac{U}{T}$$

即此时等价的定义不会造成区别:

$$S = \frac{U}{T}$$

11.4.4 传热熵

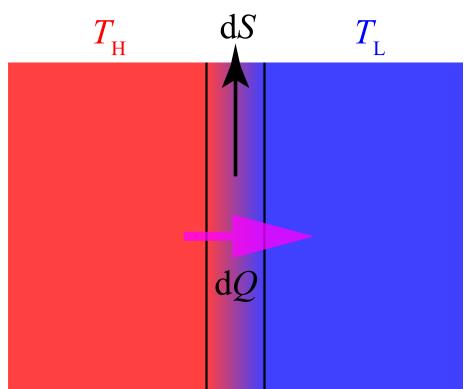


图 11.15: 传热熵产生

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.

11.4.5 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.

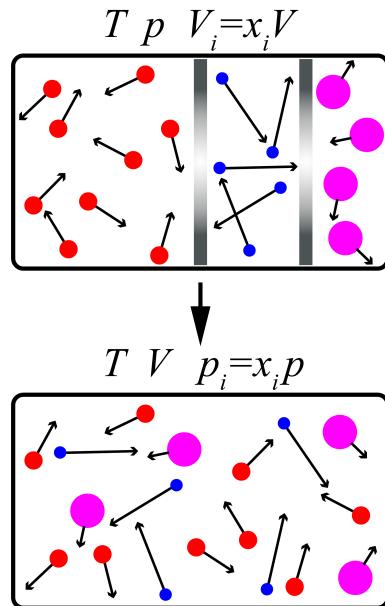


图 11.16: 混合熵

⁵然而低温下影响巨大, 这与量子效应有关

11.5 热力学函数与其特性

11.5.1 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:

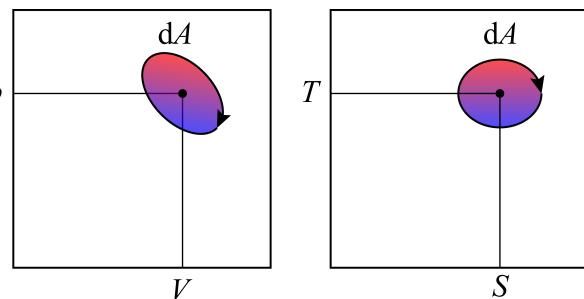


图 11.17: 循环过程的做功与吸热相等

$$\frac{\partial(T, S)}{\partial(p, V)} = \begin{vmatrix} \frac{\partial T}{\partial p} & \frac{\partial S}{\partial p} \\ \frac{\partial T}{\partial V} & \frac{\partial S}{\partial V} \end{vmatrix} = \left(\frac{\partial T}{\partial p}\right)_V \left(\frac{\partial S}{\partial V}\right)_p - \left(\frac{\partial S}{\partial p}\right)_V \left(\frac{\partial T}{\partial V}\right)_p = 1$$

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.

11.5.2 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:

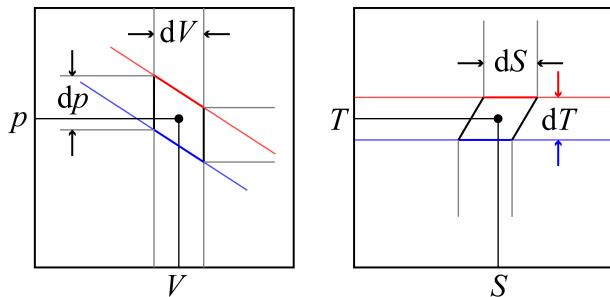


图 11.18: 构造循环过程证明能态定理

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 $n \approx \frac{11}{9}$, 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. 所以即使增加 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.

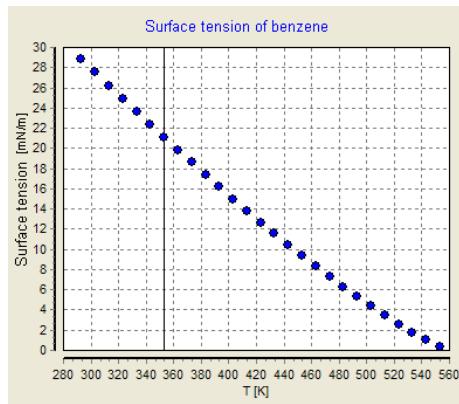


图 11.19: 苯的 $\sigma \sim T$ 关系

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:

$$\left(\frac{\partial}{\partial T} \right)_V \frac{F}{T} = -\frac{U}{T^2}$$

$$\left(\frac{\partial}{\partial V} \right)_T \frac{F}{V} = -\frac{G}{V^2}$$

$$\left(\frac{\partial}{\partial p} \right)_T \frac{G}{p} = -\frac{F}{p^2}$$

11.5.3 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:

$$W'' \leq G_{\text{initial}} - G_{\text{final}}$$

11.5.4 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.

11.6 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.

11.6.1 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11.20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.

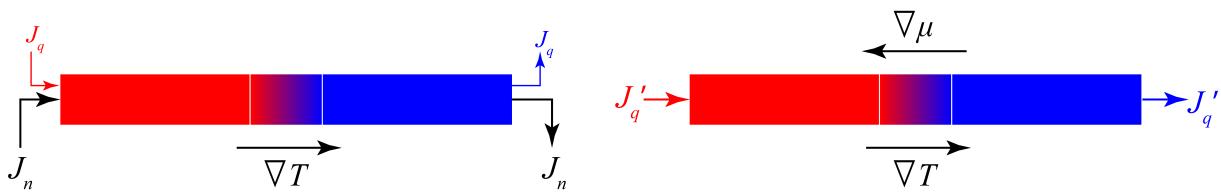


图 11.20: Soret 效应与 Dufuer 效应

杜福尔 (Dufuer) 效应

如图11.20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 11.21 左演示了其半导体型热电偶的工作原理.

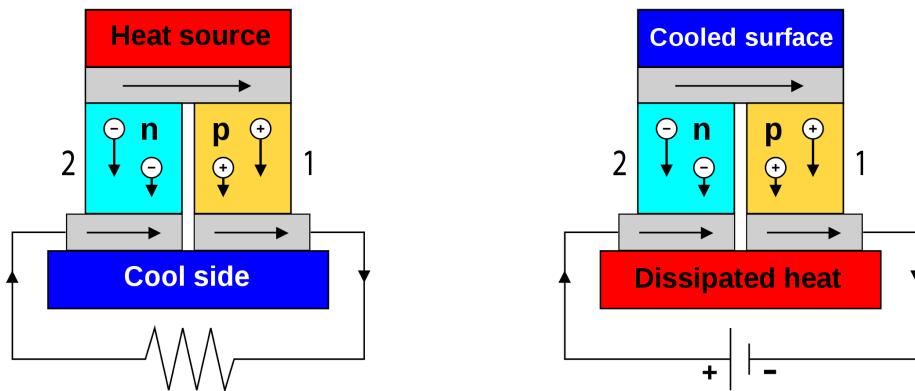


图 11.21: Seebeck 效应与 Peltier 效应

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.

11.6.2 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：

$$\frac{\partial s}{\partial t} = -\nabla \cdot \mathbf{J}_s + \theta$$

11.6.3 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

11.6.4 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.

第十二章 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.

12.1 固体晶格论

12.1.1 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12.1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:

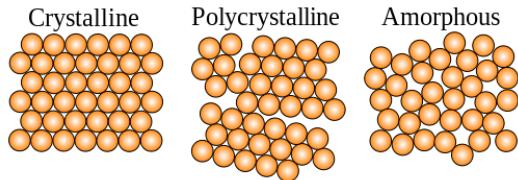


图 12.1: 单晶体, 多晶体与无定形体

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:

$$H = \sum_g T(\mathbf{v}_{i,g}) + V(\mathbf{r}_{i,g}) \quad , \quad i = 1, 2 \dots n ; g \in G$$

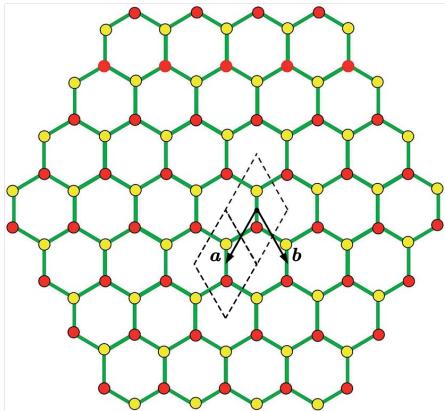


图 12.2: 石墨烯的结构

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.

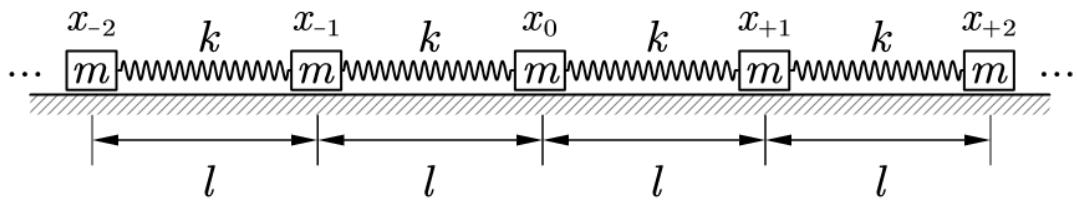


图 12.3: 弹性

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.

12.1.2 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：

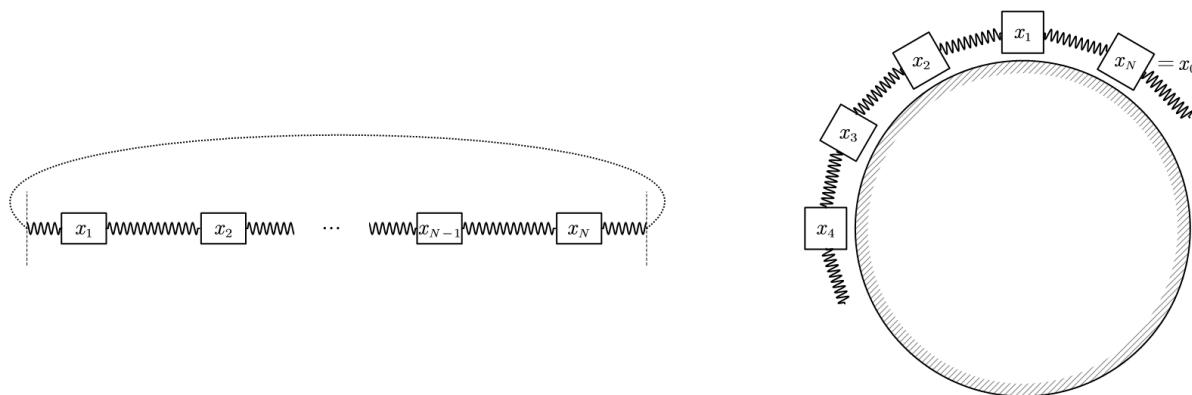


图 12.4: 周期性边界条件

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：

$$N\kappa l = j \cdot 2\pi \quad j = 1, 2, 3 \dots N$$

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:

$$H = \sum_{j=1}^N \frac{1}{2} N m (|\dot{q}_j|^2 + \omega_j^2 |q_j|^2) \quad , \quad \omega_j^2 = \frac{4k}{m} \sin^2 \frac{\pi j}{N}, \quad \kappa_j = \frac{2\pi j}{Nl}$$

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:

$$x_n(0) = \int_{-\pi/l}^{\pi/l} [A(\kappa) \cos n\kappa l + B(\kappa) \sin n\kappa l] d\kappa$$

$$A(\kappa) = \frac{l}{2\pi} \sum_{n=-\infty}^{+\infty} x_n(0) \cos n\kappa l \quad ; \quad B(\kappa) = \frac{l}{2\pi} \sum_{n=-\infty}^{+\infty} x_n(0) \sin n\kappa l$$

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：

$$v_g = \frac{d\omega}{d\boldsymbol{\kappa}}$$

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 $\omega, \boldsymbol{\kappa}$ 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。

12.2 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。

但是就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。

12.2.1 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:

$$p \cdot n \propto e^{-\frac{V_g}{kT}} ; \quad \sigma = p e \mu_p + n e \mu_n$$

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).

12.3 液体的彻体性质

12.3.1 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12.5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.

图 12.5: 在有“弹性”的液体上行走
的流动性的材料,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.

液体内部一般无序.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12.6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:

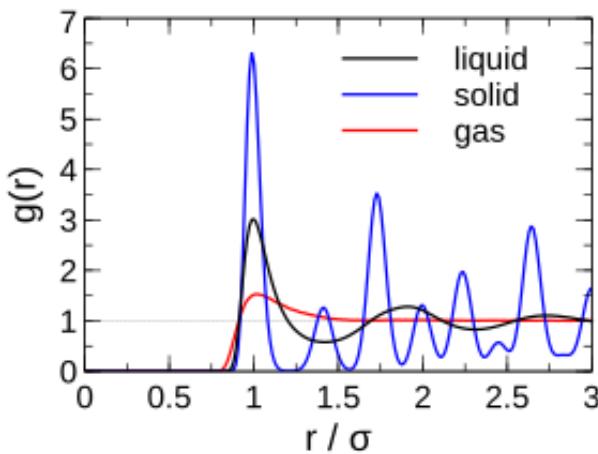


图 12.6: 径向分布函数

- 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.

12.3.2 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”.

12.4 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.

12.5 极端条件下的其他物态

- 超冷 ($\sim \mu K$): **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:

$$u = \frac{74\sigma T^4}{c}$$

第十三章 相与相变摘要

13.1 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 $0.5\text{cal/g} \cdot \text{K}$

13.2 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):

$$\frac{dp}{dT} = \frac{s_g - s_l}{v_g - v_l} = \frac{l}{T(v_g - v_l)}$$

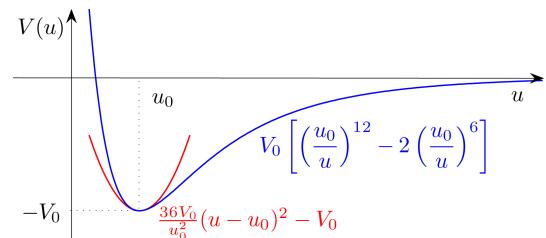


图 13.1: 范氏等温线

- **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.

第十四章 统计物理基础

14.1 数学基础

14.1.1 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:

在古典模型(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 $A = \{2, 4, 6\}$.

但在几何模型(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:

$$p(x, y) = p_1(x)p_2(y)$$

14.1.2 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:

$$p_y(y) = \frac{dp}{dy} = \frac{p(x)dx}{f'(x)} = \frac{1}{f'(x)}p(x)$$

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

$$p_{e+} = \frac{p(v)}{v} = \frac{1}{2\sqrt{\pi e}}e^{-\frac{v^2}{2}}(v > 0) ; p_{e-} = \frac{p(v)}{-v} = \frac{1}{2\sqrt{\pi e}}e^{-\frac{v^2}{2}}(v < 0)$$

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):

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

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.

14.1.3 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.

14.2 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.

14.3 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:

$$\Gamma = \frac{1}{4} nc ; \quad d\Gamma = \frac{1}{4} dnc$$

$$p = \frac{1}{3} n \bar{\varepsilon} ; \quad dp = \frac{1}{3} dn \varepsilon$$

14.4 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...

14.5 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:

$$dq = T ds$$

第十五章 光波与光线

15.1 界面上的反射与折射

15.1.1 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大至 1mm 左右的红外光(infrared light, IR) 与短至 10nm 的紫外光(ultraviolet light, UV).

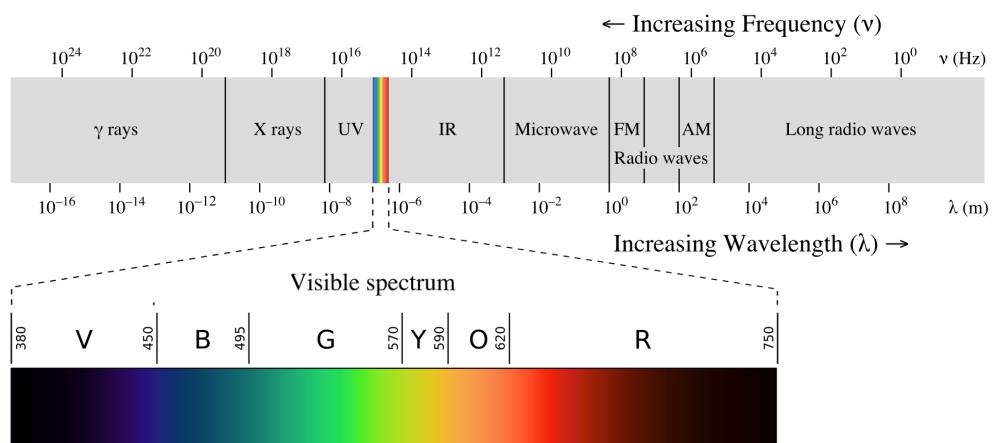


图 15.1: 光学研究范围

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.

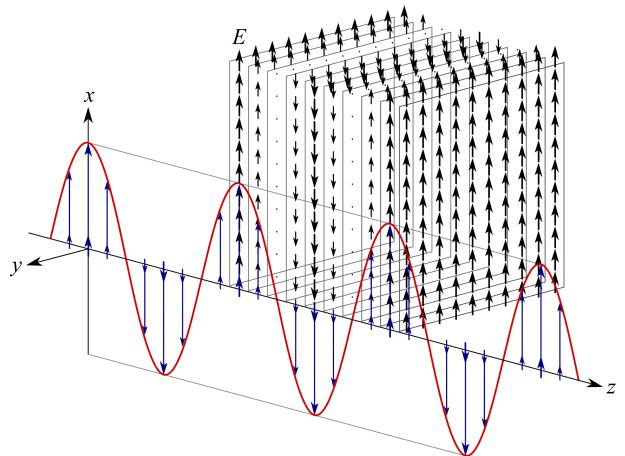


图 15.2: 平面偏振电磁波

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:

$$\left[\nabla^2 - \frac{1}{(c/n)^2} \frac{\partial^2}{\partial t^2} \right] A e^{-i\omega t} = 0 \Rightarrow (\nabla^2 + k^2) A = 0$$

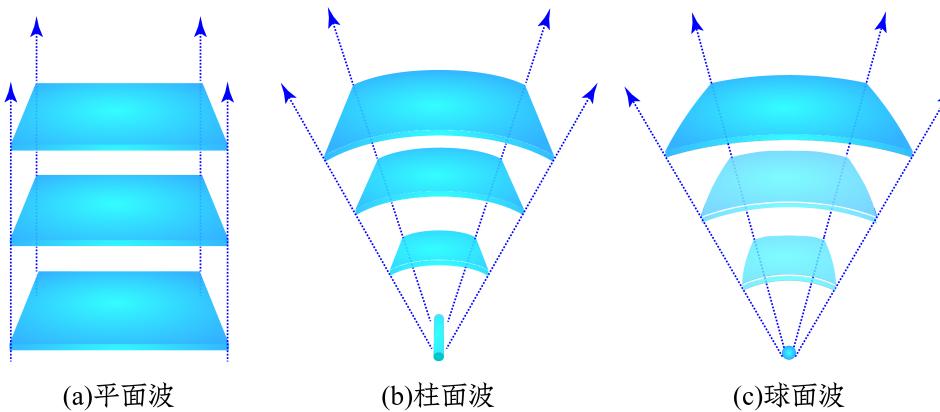


图 15.3: 三种常用波形式

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.

15.1.2 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.

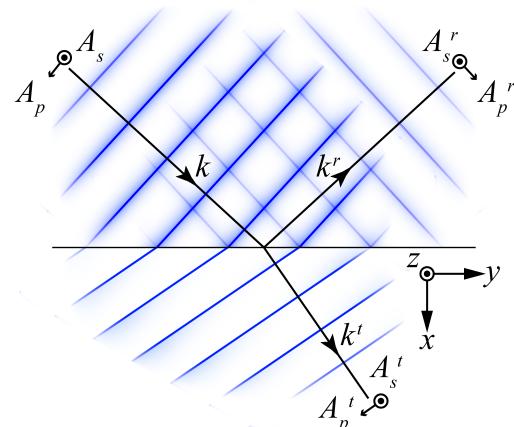


图 15.4: 界面与两种模式

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.

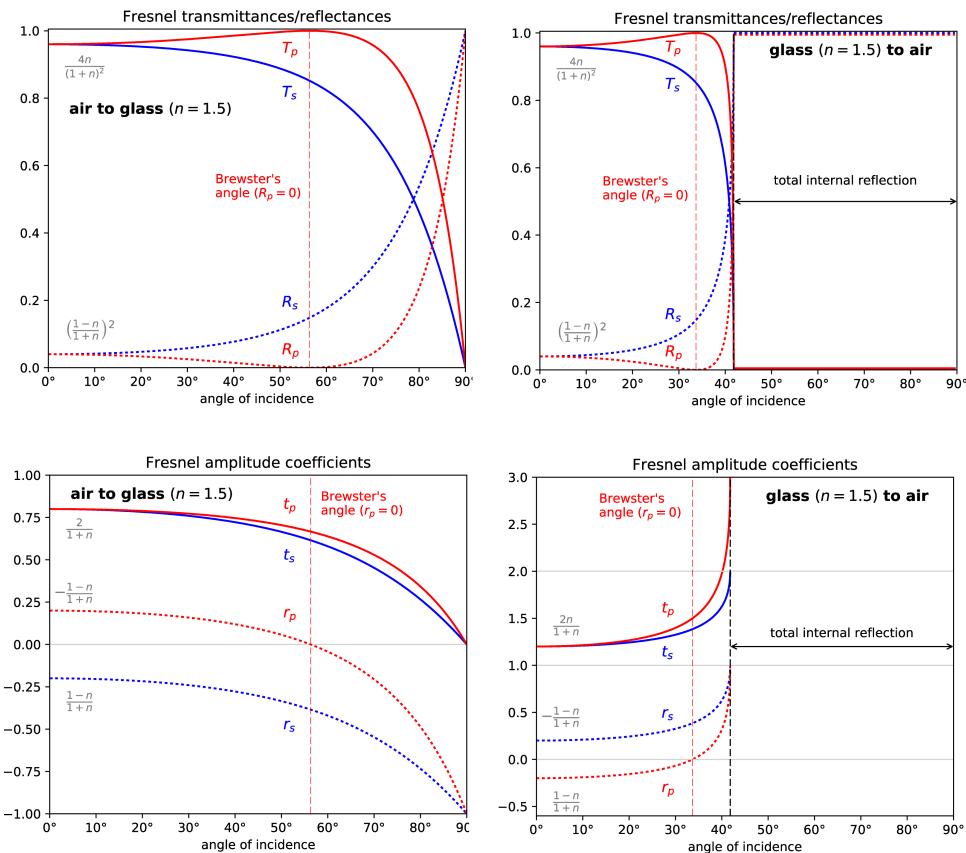


图 15.5: 界面反射与折射

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

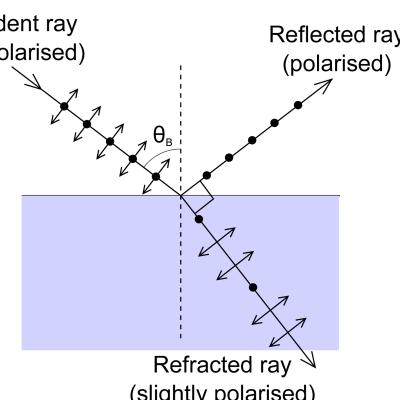


图 15.6: 布儒斯特角

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 -光没有反射的直接出射, 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:

$$r_s = e^{-2i\delta_s}, \quad t_s = \frac{2n \cos \theta}{\sqrt{n^2 - n'^2}} e^{-i\delta_s}, \quad \delta_s = \arctan \frac{\sqrt{n^2 \sin^2 \theta - n'^2}}{n \cos \theta}$$

$$r_p = e^{-2i\delta_p}, \quad t_p = \frac{2nn'}{\sqrt{(n^2 - n'^2)(n^2 \tan^2 \theta - n'^2)}} e^{-i\delta_p}, \quad \delta_p = \arctan \frac{n \sqrt{n^2 \sin^2 \theta - n'^2}}{n'^2 \cos \theta}$$

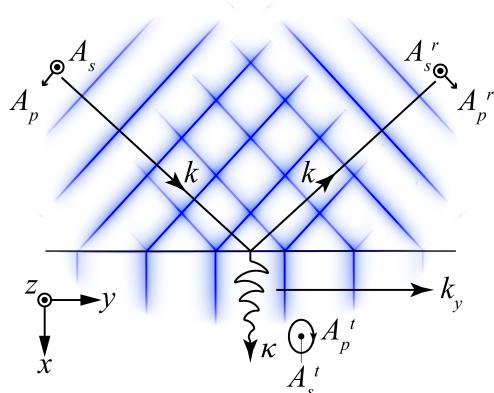


图 15.7: 全内反射与隐失波

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.

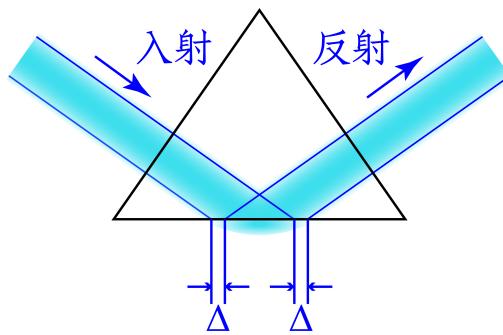


图 15.8: Goos-Hänchen Shift

15.2 光线方程

15.2.1 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:

$$d(n \sin \varphi) = 0 \Rightarrow d\varphi = -\frac{\sin \varphi dn}{n \cos \varphi} = -\frac{|\nabla n|}{n} ds \sin \varphi$$

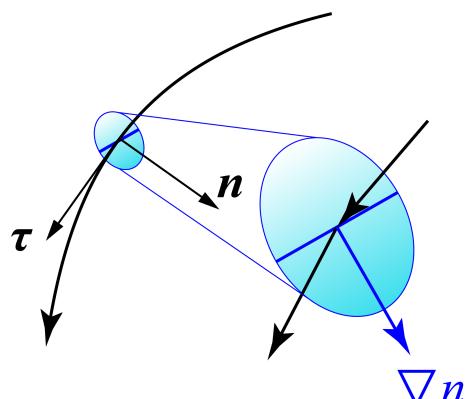


图 15.9: 光线方程推导

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.

15.2.2 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:

$$nr \sin \theta = L = \text{Const.}$$

15.3 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:

$$\Gamma : \quad \mathbf{r} = \mathbf{R}(t), \quad \mathbf{R}(t_A) = \mathbf{r}_A, \quad \mathbf{R}(t_B) = \mathbf{r}_B$$

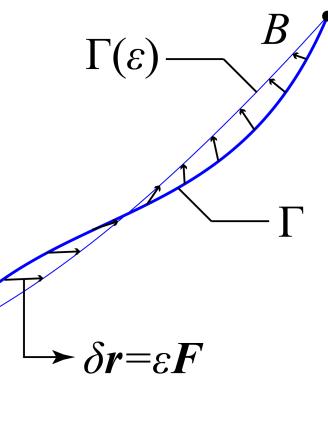


图 15.10: 光程变分

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.

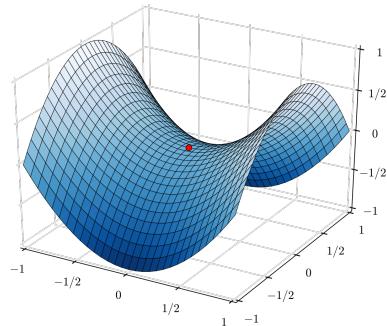


图 15.11: 二维鞍点处势取平稳值

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)

第十六章 光学成像

16.1 傍轴光成像

16.1.1 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

$$\frac{z}{l} \ll 1$$

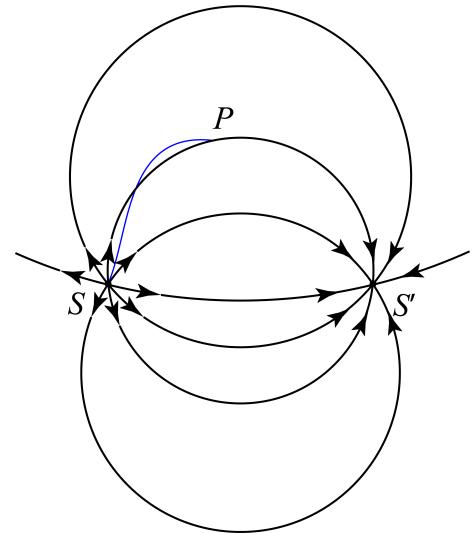


图 16.1: 物与像

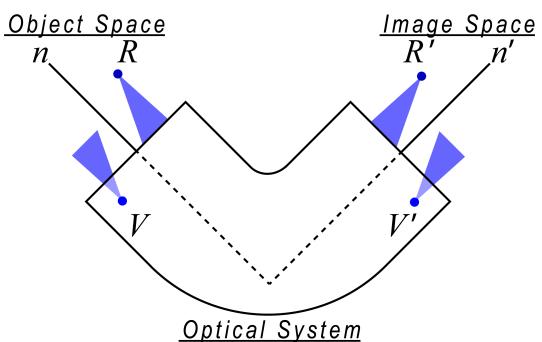


图 16.2: 实与虚

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).

16.1.2 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:

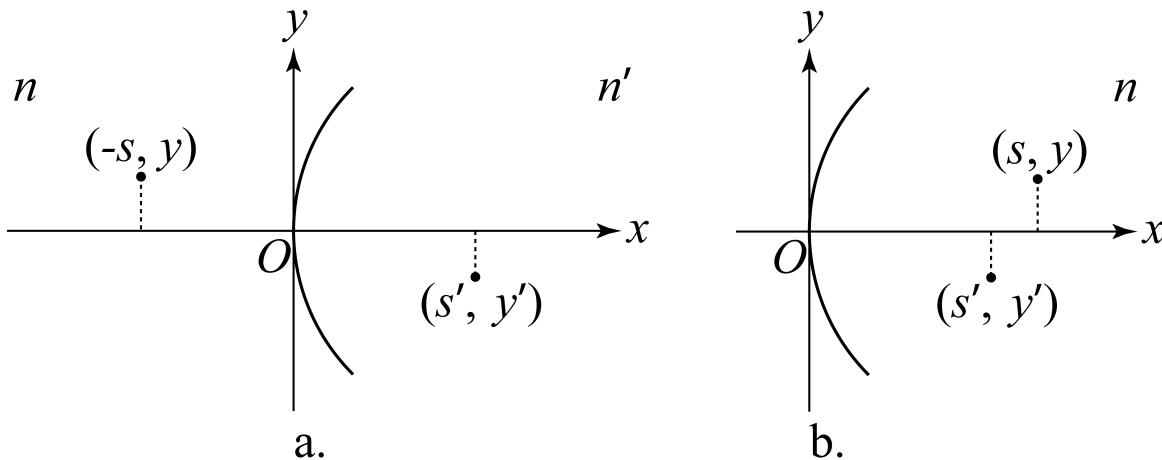


图 16.3: 球面折射与球面反射

在球面折射情况16.3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:

$$\beta = \frac{y'}{y} = -\frac{ns'}{n's}$$

$$\frac{n}{s} + \frac{n'}{s'} = \frac{n'-n}{R}$$

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”. $1D = 1\text{m}^{-1}$.

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:

$$f = \frac{n}{\Phi} = \frac{nR}{n' - n} ; f' = \frac{n'}{\Phi} = \frac{n'R}{n' - n}$$

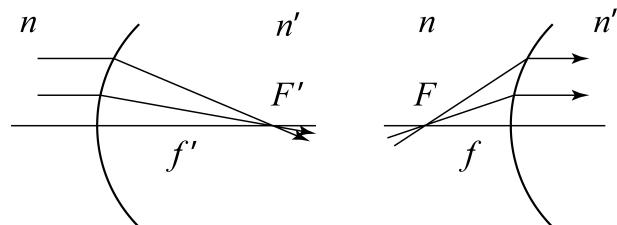


图 16.4: 焦点

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16.3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.

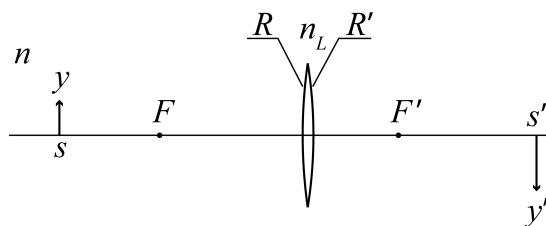


图 16.5: 薄透镜

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.

16.1.3 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.

16.2 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.

16.2.1 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.

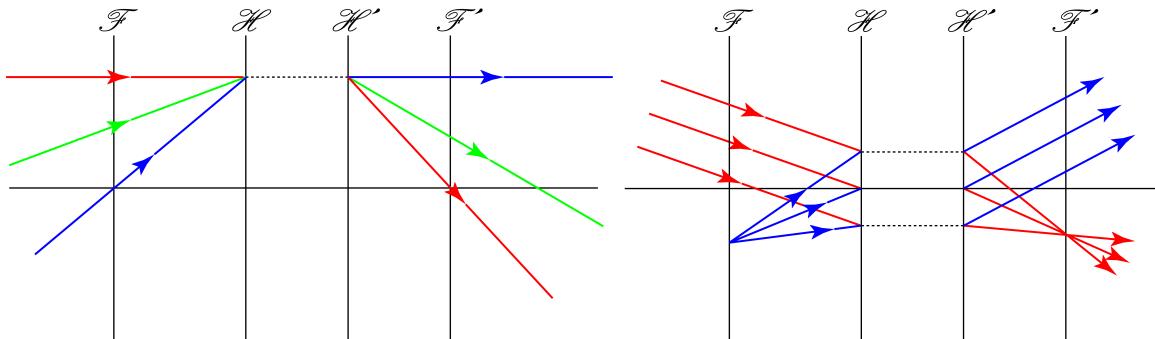


图 16.7: 特殊光线

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.

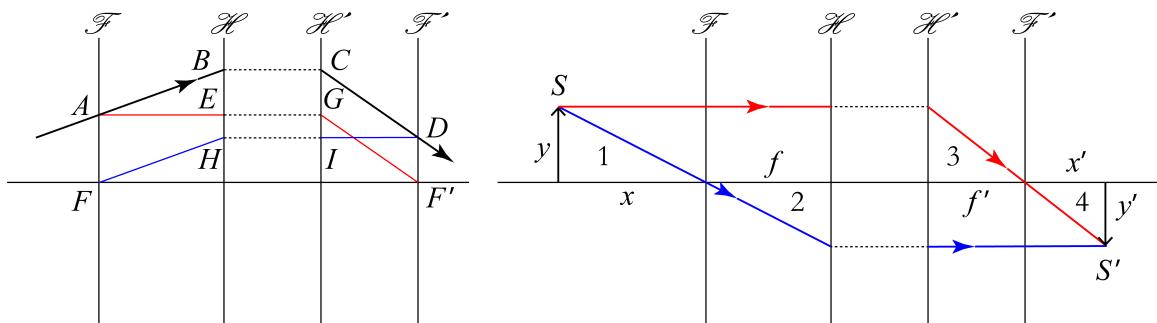


图 16.8: 共轭线与共轭点

16.2.2 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

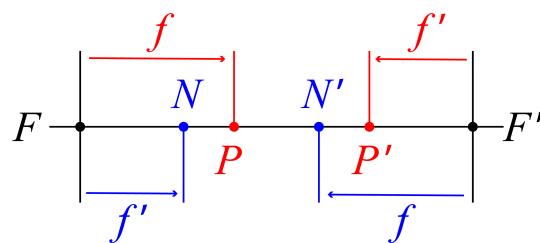


图 16.10: 三对关键基点

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 .

16.2.3 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.

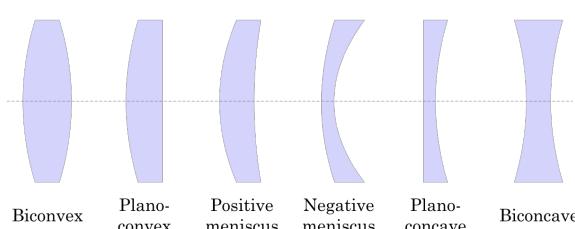


图 16.11: 透镜种类

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16.12, 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.

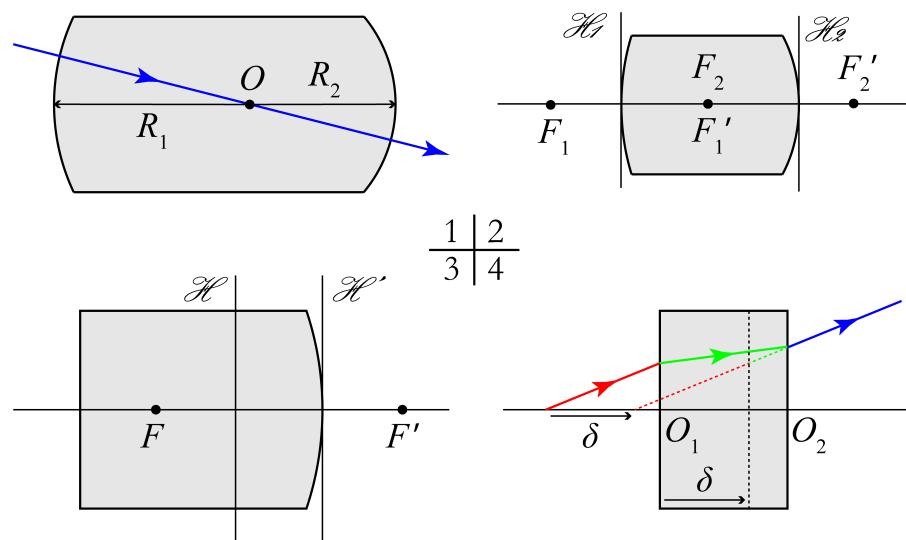


图 16.12: 四种特殊厚透镜

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16.12,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.

16.3 更多讨论 *

16.3.1 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).

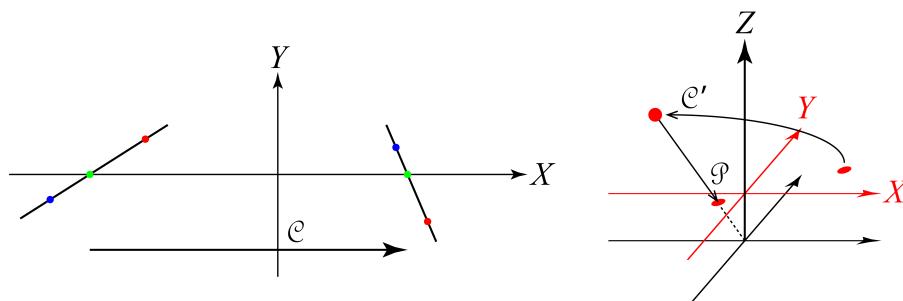


图 16.13: 射影映射

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.

第十七章 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.

17.1 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:

17.1.1 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^{17.1}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

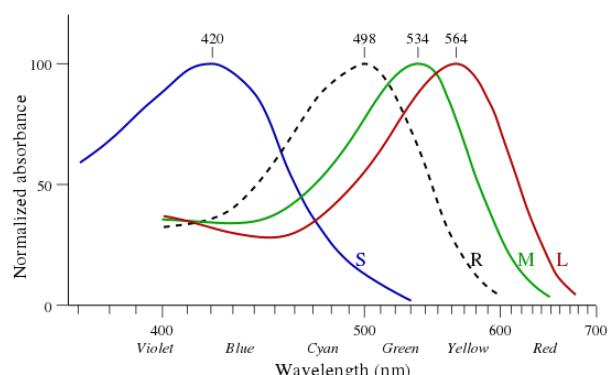


图 17.1: 四种感光细胞约化吸收率

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 17.2: 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:

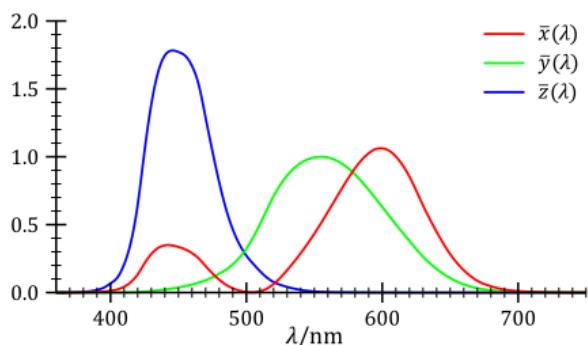


图 17.3: 色匹配函数

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。

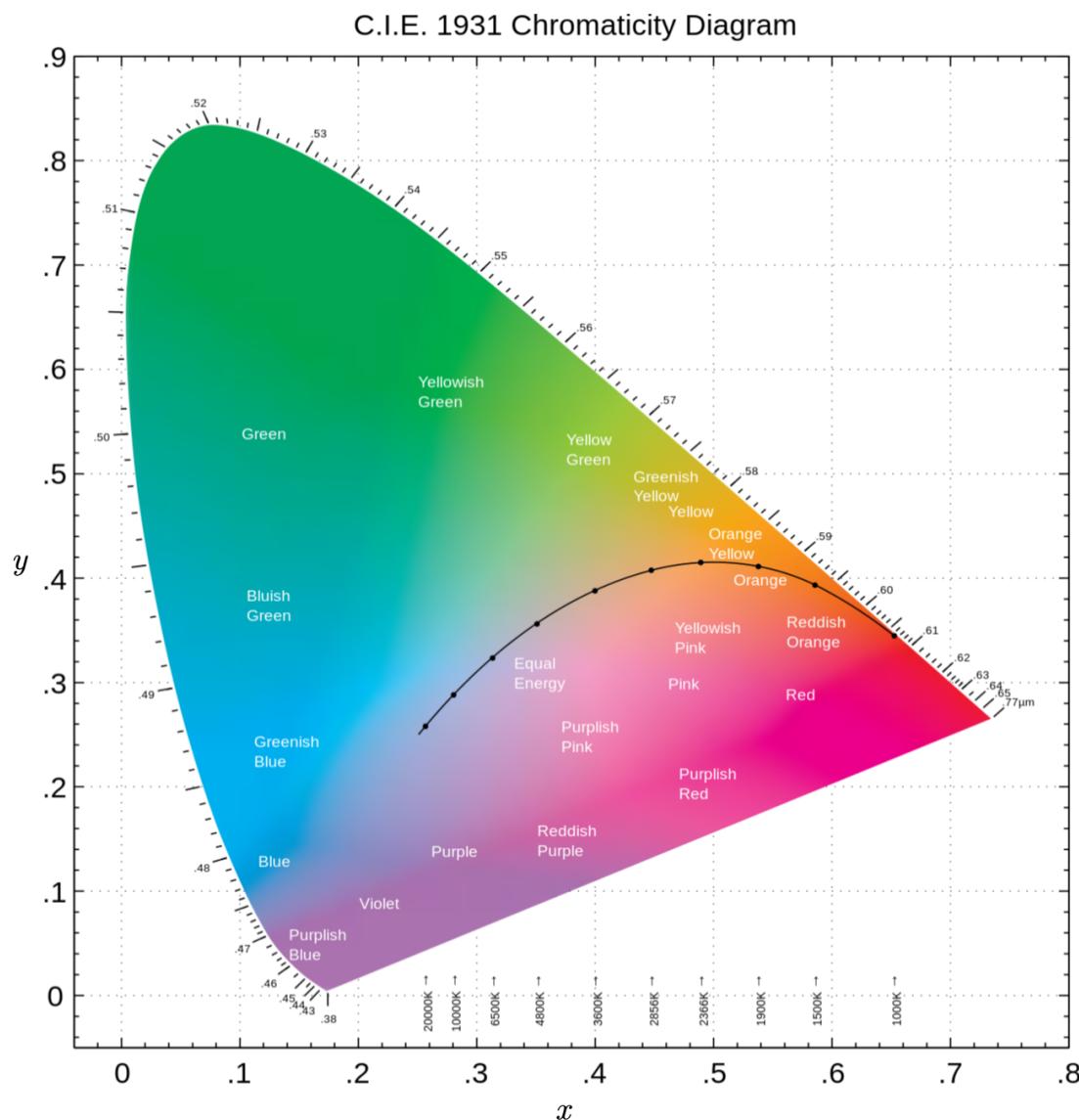


图 17.4: CIE1931 色度图

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 $(1, 0, 0), (0, 1, 0), (0, 0, 1)$.

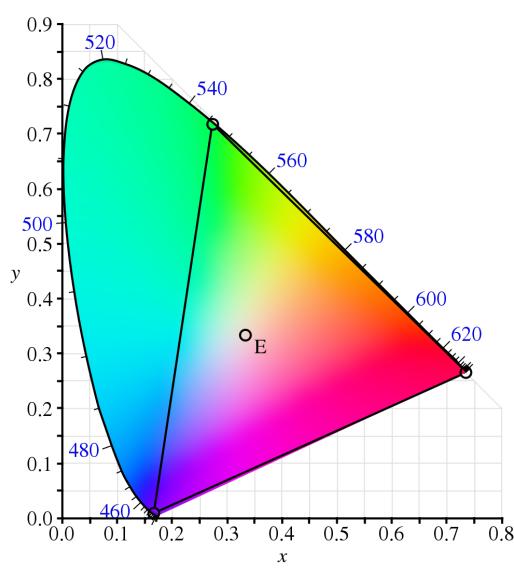


图 17.5: CIE1931-RGB 色域

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.

17.1.2 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 17.2⁴。

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：

$$I = 683.002 \text{cd} \cdot \text{sr}/\text{W} \cdot \int V(\lambda) i(\lambda) d\lambda$$

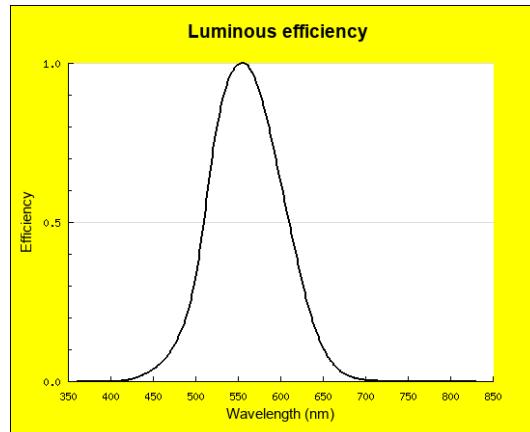


图 17.6: 光度函数图像

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 17.1: 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 17.2: 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

17.1.3 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。

17.1.4 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:

$$B = B'$$

17.2 光阑与光瞳

17.2.1 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.

17.2.2 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.

17.3 常见仪器简介

17.3.1 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.

17.3.2 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:

$$M = \frac{s\Delta}{f_O f_E}$$

17.3.3 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:

$$M = \frac{f_O}{f_E}$$

第十八章 静电学

18.1 电荷与电场

18.1.1 电磁相互作用与电荷

物质若携带电荷(charge), 则可以参与电磁相互作用(electromagnetic interaction). 在经典情形下理解为电荷受到了一个力. 我们先研究静电学(electrostatics), 它要求受力物体与施力物体¹都处于静止状态, 或者认为一切电荷都处于静止状态.

电荷分正负, 阴阳激荡.

作为协变电磁场理论的基本要求与实验规律, 电荷是一个参考系不变的标量, 微观地看电荷具有量子化(quantized) 的奇妙特性². 质子的质量与电子的质量没有简单的整数比关系, 但电荷量确实严格相反. 如果设想两者电量有着微小差异, 那么构成自然界的原子也就携带了一个能够传递远程相互作用的库仑力, 从而导致自然界的不稳定³. 而质子(proton) 的电量被定义为基本电荷(elementary charge)⁴:

$$e = 1.602176634 \times 10^{-19} \text{ C}$$

实际上今天我们知道, 电荷量的量子化单位实际上更小, 至少是 $e/3$, 夸克的电量就是这个电量的一倍或两倍, 可正可负. 而电子(electron) 带一个单位的负电荷. 当然, 特殊场合还会有带一个单位正电荷的正电子(positron) 和一个单位负电荷的反质子(anti-proton) 甚至它们组成的反物质. 中子(neutron) 是不带电的.

电荷如同静质量, 是基本粒子的基本属性. 它只能随着基本粒子运动, 没有相互作用时不能被创造与消灭. 在存在电磁相互作用时电荷可以被创造与消灭, 典型的过程如电子与正电子湮灭(annihilation) 为两个光子, 或高能光子在原子核附近产生(production) 电子正电子对. 这些过程电荷的代数和守恒, 其实现要求反物质与相对论性运动的存在. 在经典情形下有严格的电荷守恒定律(charge conservation), 电荷可以流动形成电流, 但流动的过程中电荷的总量不能发生改变. 我们总是计算一定体积内的电荷代数和:

$$dQ = \sum_i n_i q_i dV$$

其中 n_i 是第 i 种基本构成粒子的数密度, q_i 是这种基本粒子的带电量. dV 则为微元体积. 而上式也被写为:

$$dQ = \rho dV ; \quad \rho = \sum_i n_i q_i = \sum_i \rho_i$$

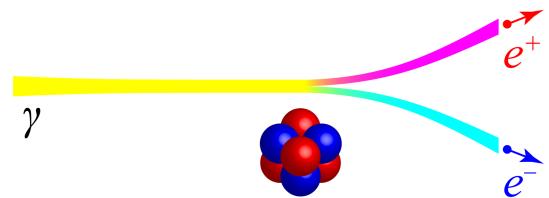
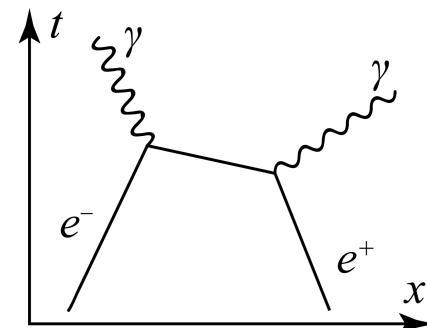


图 18.1: 电子对湮灭与产生

¹实际上物理上不允许超距的施力物体, 施力物体就是局域的电场.

²与磁单极子的存在与量子化有关.

³是的, 这种思考方式就是人择原理(anthropic principle).

⁴这个值是 2018 新国际标准单位制提出的规定值.

18.1.2 库仑定律

库仑定律(Coulomb's law) 用于描述真空中的两个孤立静止点电荷之间的相互作用力:

$$\mathbf{F}_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \mathbf{e}_{12}$$

其中系数 $k_e = \frac{1}{4\pi\epsilon_0} \approx 8.99 \times 10^9 \text{ N} \cdot \text{m}^{-2} \cdot \text{C}^{-2}$ 为库仑常数.

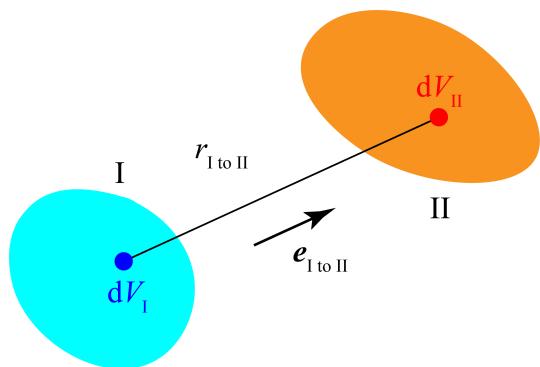


图 18.2: 电荷体系间的相互作用力
一个复杂电荷系统 (包括构成介质的电荷) 可以用连续的电荷密度描述:

$$dQ = \rho dV$$

那么两部分电荷之间的相互作用力为:

$$\mathbf{F}_{I \text{ to } II} = \frac{1}{4\pi\epsilon_0} \int_{V_{II}} \rho(\mathbf{r}_{II}) \int_{V_I} \frac{\rho(\mathbf{r}_I)}{r_{I \text{ to } II}^2} \mathbf{e}_{I \text{ to } II} dV_I dV_{II}$$

相互作用力在定义上就是成对存在的, 所以两部分电荷体系间的相互作用力满足牛顿第三定律: 大小相等, 方向相反, 对整个体系产生的力矩的贡献为零.

18.1.3 电场

现代物理认为, 狭义相对论下的时空观是被事实验证了的. 一个重要的推论就是**局域性原理**(principle of locality): 物体仅仅被此时此刻周围的物质所影响. 所有相互作用必须被理解为**近距作用**(action upon contact) 而不能是**超距作用**(action at a distance). 这是为了能与更本质的**因果律**(principle of causality) 相适应. 如果物体这一时刻的状态及其改变取决于同一时刻的空间上相隔一定距离的另外一点 (类空间隔). 那么就可以另一个点的事件称作因, 而物体的状态与状态改变称作果. 但类空间隔在洛伦兹变换下具有可逆性. 也就是原则上可以在果的时空点也可以制造事件去对因的点产生影响. 从而造成逻辑上的紊乱. 所以我们必须规定这两个时空点间不存在因果关系, 互相独立没有影响.

局域性原理的直接结论便是, 在引力与电磁相互作用时, 作为物质的**场**(field) 的概念的引入. 场本是数学概念. 在一个三维的欧几里得空间 \mathbb{R}^3 中, 如果每一个点都定义了一个数 (标量), 代表某种物理量, 物理现象可以在不同参考系, 惯性系中观察, 所对应的时空点的位置坐标可能发生改变, 但特定的时空点的概念, 和与之相结合的物理现象与描述它的这个数, 不应该发生改变, 称为**标量**(scalar):

$$f: \mathbf{r} \mapsto f(\mathbf{r})$$

⁵而在经典电磁场论的观点下库仑定律为其矢量场论形式的推论.

如,压强场,温度场,波动的相位场,都是典型的标量场的例子.数学上定义了标量场的梯度(gradient)来描述在一点极其附近的标量场的行为,它是一个矢量,方向是使得标量上升最快的欧几里得空间中的方向,大小是沿这个方向的标量变化的方向导数.数学上可以证明:

$$\nabla f = \frac{\partial f}{\partial x} e_x + \frac{\partial f}{\partial y} e_y + \frac{\partial f}{\partial z} e_z$$

物理量在某点 r 的局部,偏离一个 dr 后值变为:

$$f(r + dr) = f(r) + \nabla f \cdot dr$$

这就是微元形式的牛顿-莱布尼茨定理(Newton-Leibniz theorem).而标量场的梯度往往也具有重要的含义.压强场的梯度决定浮力的方向与大小(反向),温度场的梯度决定热传导的方向与强弱(也是反向),波动的相位场的梯度则决定波的传播方向与波长(波矢).

而在三维的欧几里得空间中在每一个空间点也可以引入一个矢量(vector),它不占三维空间的有限体积⁶而是定义在一个点上,可以认为是这个点上扩展出了一个新的与原三维空间坐标平行的 $e_x - e_y - e_z$ 坐标系.而这一点定义的矢量就位于这个“浓缩在一点”的坐标系中⁷:

$$\mathfrak{F}: r \mapsto \mathbf{F}(r)$$

显然,标量场的梯度恰是一个矢量场.矢量不能简单理解为三个标量场的合成.因为在空间旋转(或时空的惯性系变化)下,标量应该是不变的,而矢量将发生与坐标系相似的旋转,从而三个分量之间相互转化,只有矢量的模长是不变的.另一种典型的矢量场是流体的流速场 v ,或者更普遍的,某种物理属性的流密度场 j ,前者定义为单位时间通过单位面积的体积,后者则推广为单位时间通过单位面积的某物理量 Q :

$$dQ = j \cdot dA dt$$

在一点附近的矢量场的行为则比标量场复杂很多:

$$\mathbf{F}(r + dr) = \mathbf{F}(r) + dr \cdot \nabla \mathbf{F}$$

此处 $\nabla = e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y} + e_z \frac{\partial}{\partial z}$ 可以形式地视为矢量微分算符,它写在上一式的后一项中实际上是在表示:

$$\begin{aligned} d\mathbf{F} &= dF_x e_x + dF_y e_y + dF_z e_z \\ &= \left(\frac{\partial F_x}{\partial x} dx + \frac{\partial F_x}{\partial y} dy + \frac{\partial F_x}{\partial z} dz \right) e_x \\ &\quad + \left(\frac{\partial F_y}{\partial x} dx + \frac{\partial F_y}{\partial y} dy + \frac{\partial F_y}{\partial z} dz \right) e_y \\ &\quad + \left(\frac{\partial F_z}{\partial x} dx + \frac{\partial F_z}{\partial y} dy + \frac{\partial F_z}{\partial z} dz \right) e_z \\ &= \left(dx \frac{\partial}{\partial x} + dy \frac{\partial}{\partial y} + dz \frac{\partial}{\partial z} \right) (F_x e_x + F_y e_y + F_z e_z) \\ &= dr \cdot \nabla \mathbf{F} \end{aligned}$$

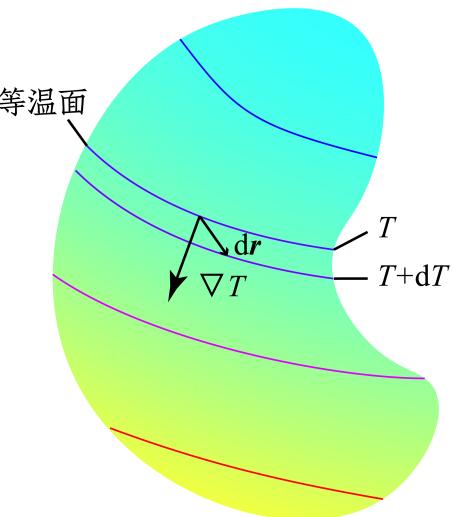


图 18.3: 温度场的梯度

⁶即,不同于位移矢量那样必须与三维空间中的不同点同时建立联系

⁷数学术语把空间叫做微分流形(differentiable manifold),而局部的矢量位于的空间叫做纤维(fiber)或切空间(tangent space),每一点都有一个这样的空间,合称纤维丛(fiber bundle).

而也可以把 ∇F 认为能完整描述这个矢量在该点处的一阶变化行为,也称为矢量的梯度,但它实际上由九个分量组成,分别描述三个分量在三个方向的变化率,一般写在一张数表中,也就是矩阵⁸:

$$\nabla \mathbf{F} = \begin{bmatrix} \frac{\partial F_x}{\partial x} & \frac{\partial F_y}{\partial x} & \frac{\partial F_z}{\partial x} \\ \frac{\partial F_x}{\partial y} & \frac{\partial F_y}{\partial y} & \frac{\partial F_z}{\partial y} \\ \frac{\partial F_x}{\partial z} & \frac{\partial F_y}{\partial z} & \frac{\partial F_z}{\partial z} \end{bmatrix}$$

显然用这样的一个矩阵去描述矢量场在一点处的行为是代数的,抽象地,不几何直观的.所以我们引入两个物理上会更实用的概念.

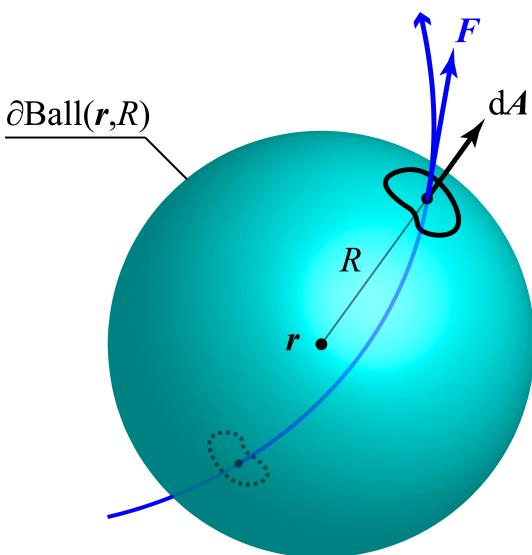


图 18.4: 矢量场的散度

一是矢量场的散度(divergence),它是一个标量,定义为矢量场的源强度:

$$\nabla \cdot \mathbf{F} = \lim_{R \rightarrow 0} \left(\oint_{\partial \text{Ball}(r, R)} \mathbf{F} \cdot d\mathbf{A} \right) / \left(\frac{4}{3}\pi R^3 \right)$$

其中 $\text{Ball}(r, R)$ 表示以待研究的点为球心的,半径为 R 的球形体,而 ∂Ball 表示它的表面.积分号 \oint 用来代表此场合下用来积分的区域是一个闭合的区域⁹.积分的量称为通量(flux),即矢量场与面积的电场,速度场情况下即单位时间通过面 $d\mathbf{A}$ 的体积.现在对闭合的面进行积分,再除以面包含的体积,代表的是该点单位体积内产生这个矢量场的场线的源的多少,即源强度.数学上可以证明,它与体积以怎样的取法趋于零无关,如果取成与坐标系平行的长方体的形状,能够在坐标系下计算出它的值为上面那个矩阵的迹:

$$\nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

第二个概念便是旋度(curl)了.它实际上代表矢量场的变化矩阵中反对称的那一部分¹⁰.它可以被写为一个矢量.与在这一点处的无限小的环积分有关,数学上不难证明:

$$\mathbf{n} \cdot \nabla \times \mathbf{F} = \lim_{R \rightarrow 0} \left(\oint_{\partial \text{Circ}(r, n, R)} \mathbf{F} \cdot dl \right) / (\pi R^2)$$

由于半径为 R 的圆的选取还依赖于其朝向.故在上式中我们对积分的方向和垂直于积分圆面的法向单位矢量 n 采取了右手定则式的对应.而最后积分出来得到的数恰好是面法向单位矢量与从 \mathbf{F} 的各个导数中提取的旋度矢量的点乘.环积分通常被称为环量(circulation),旋度则反应矢量场局域的涡旋程度.数学上可以证明旋度可以由行列式表示为:

$$\nabla \times \mathbf{F} = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \\ e_x & e_y & e_z \end{vmatrix}$$

⁸实际上,形成了一个张量场(tensor filed).张量将不在本部分教材的讨论范围内.

⁹数学上称为又闭又开(clopen)的集合,表示它没有边界.

¹⁰任何一个矩阵 A 都可以被唯一地分解为一个对称矩阵 $(A + {}^t A)/2$ 与反对称矩阵 $(A - {}^t A)/2$ 之和,前者有六个独立分量,后者仅有三个独立分量.

作为与牛顿-莱布尼茨定理对应的散度与旋度版本,根据两个量的定义可知前者发生在体积内部的积分与其闭合表面上,后者发生在曲面上面积分和它的闭合曲线边界上:

$$\int_V \nabla \cdot \mathbf{F} dV = \oint_{\partial V} \mathbf{F} \cdot d\mathbf{A}$$

$$\int_A \nabla \times \mathbf{F} \cdot d\mathbf{A} = \oint_{\partial A} \mathbf{F} \cdot dl$$

前者为[奥斯特诺格拉德斯基-高斯定理](#)(Ostrogradsky-Gauss theorem),后者则为[开尔文-斯托克斯定理](#)(Kelvin-Stokes theorem)¹¹.

现在让我们回到库仑定律,现在就必须把它理解为两个公式:

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \mathbf{e}_r ; \quad \mathbf{F} = q\mathbf{E}$$

前一半式子说明,即使没有引入任何试探电荷 q ,在空间中距离场源电荷 Q 为 r 处依然存在某种独立于电荷而存在的“场”.否则也就不可能对引入的电荷瞬间产生力的作用.这个场由电荷激发,以后将知道它有自己的能量与动量,而且都是可以被局域地确定¹².经典物理中似乎只能由电荷产生场,但高能物理下场也能反过来产生电荷粒子,实际上此时会将电磁场理解为“光子”这样的物质.电磁场的特点是能对带电荷的物质产生作用.

而后一半的式子则给出了电磁场中一点处电场强度矢量 \mathbf{E} 的定义.带电粒子如果在电磁场中静止,那么将受到一个与电荷量成正比的力.把力与电荷量的比值定义为电场强度矢量.

显然,带电粒子的引入也会产生电场.这个电场与原来场源电荷产生的电场叠加,形成了空间的真实电场分布.但在计算这个“试探电荷”处的电场时,不应该考虑自己在自己这一点处产生的电场.因为自己对自己的力总是零.而总是计算“外场”在试探电荷这一点处的值.不过,试探电荷的引入与否可能会对外界的场源电荷部分产生改变.从而改变外场.这又需要另说了,我们下一章的导体静电感应就是在不断讨论这个问题.

电场满足[叠加原理](#)(superposition principle).我们总是一对一对地考虑电荷之间的二体作用,把相互作用力做非相干的矢量叠加.这也就是说,给定一个场源电荷分布,可以叠加计算出空间中任意一点的电场强度:

$$\mathbf{E}(\mathbf{r}) = \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{\rho e_R}{R^2} dV' , \quad \mathbf{R} = \mathbf{r} - \mathbf{r}'$$

$$\mathbf{E}(\mathbf{r}) := \frac{\mathbf{F}_{\text{static}}}{q}$$

也是出于这个原因,所有由电场强度派生出来的量:电势,电通量等等都符合线性的叠加原理.马上就会介绍,这是由于描述电场与电荷关系的基本方程是一个线性的微分方程的缘故.

18.2 两个定律与电势

库仑定律是静电学最基本的公理,但我们研究电场作为场的性质,它与电荷的关系.为以后深入了解电磁场整体打好基础.

¹¹微分流形上的微分形式有一个[嘉当-斯托克斯定理](#)(Cartan-Stokes theorem) 统一了以上所有形式:

$$\oint_{\partial\Omega} \omega = \int_{\Omega} d\omega$$

¹²即,可以确定每一点的场能量动量密度.

18.2.1 电场的高斯定律

首先让我们研究电通量(electric flux), 通过一个曲面(不一定闭合)的电通量定义为曲面上电场强度的积分:

$$\Phi = \int_A \mathbf{E} \cdot d\mathbf{A}$$

我们在这儿必须引入有向曲线, 曲面, 体积的概念. 固然, 曲线, 曲面, 体积都没有方向与正负, 它们仅仅是特定空间中的几何形状. 但定义各种积分时, 曲线积分的线元矢量 $d\mathbf{l}$, 曲面积分的面元矢量 $d\mathbf{A}$ 都有两个可能的方向. 而体积元 dV 也不一定恒正, 在某些情况下将取负. 那么我们通常会有一个符号约定:

- 对于可定向(orientable)曲面(不可定向的比如莫比乌斯环), 总是定义一个方向为正, 即为每一个面元指定一个面元矢量, 它取垂直于面的两个方向之一, 而相邻的面元具有相同的取向. 而如果涉及到面边界上闭合曲线的环积分, 则线元矢量取的方向与面元构成右手定则(right-hand rule).
- 对于连通.connected)体积区域, 如果取体积微元为正, 那么涉及到区域边界闭合曲面上的面积分时, 面元取从体积内到体积外的方向. 如果取体积微元为负, 则面上面元取体积外到体积内的方向.

而我们之前写出的奥-高定理和开-斯定理, 都是在这个意义下正确的.

从而以上电通量也就因为我们选择的面的两种可能的正向而造成两个不同的结果, 它们相差一个相反数.

电场的高斯定律(Gauss's law)指出:

$$\oint_{\partial V} \mathbf{E} \cdot d\mathbf{A} = \frac{1}{\epsilon_0} \int_V \rho dV$$

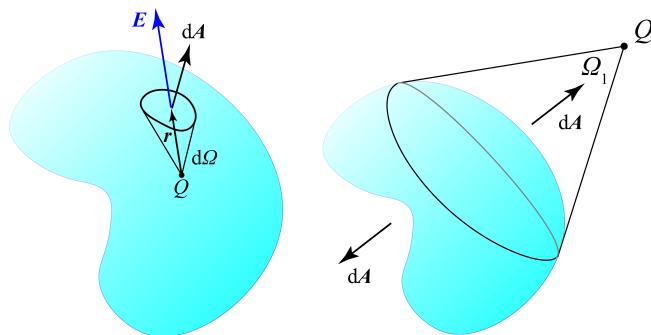


图 18.6: 证明高斯定律

我们来用库仑定律证明高斯定律. 首先考虑一个在闭合曲面内部的点电荷, 我们发现面元上的通量:

$$d\Phi = \mathbf{E} \cdot d\mathbf{A} = \frac{1}{4\pi\epsilon_0} \frac{Q e_r \cdot d\mathbf{A}}{r^2}$$

而这一个小面元对点电荷张的立体角恰好是:

$$d\Omega = \frac{e_r \cdot d\mathbf{A}}{r^2}$$

故对面积分总立体角给出 4π 的结果:

$$d\Phi = \frac{Q}{\epsilon_0} \cdot \frac{d\Omega}{4\pi} \Rightarrow \Phi = \frac{Q}{\epsilon_0} \cdot \oint_{\partial V} \frac{d\Omega}{4\pi} = \frac{Q}{\epsilon_0}$$

我们再考虑在面外的点电荷 Q 对面产生的电通量, 此时面的正向与点电荷产生电场方向的点乘就有了正负号, 积分以后得到的结果是通过点电荷去看曲面, 把曲面对点电荷张的最大视角处的曲线找到, 将曲面

分为两个部分,立体角都是 Ω_1 ,总积分为:

$$\Phi = \frac{Q}{\varepsilon_0} \cdot \oint_{\partial V} \frac{d\Omega}{4\pi} = \frac{Q}{\varepsilon_0} (\Omega_1 - \Omega_1) = 0$$

从而我们证明了,电荷若在闭合曲面内,则电通量贡献为 $\frac{Q}{\varepsilon_0}$.而若在闭合曲面外,则电通量贡献为 0.再由叠加原理,得任意情况下闭合曲面的电通量等于内部总电荷量除真空介电常数 ε_0 .考虑电荷连续分布或准连续分布的情况下,有:

$$\oint_{\partial V} \mathbf{E} \cdot d\mathbf{A} = \frac{1}{\varepsilon_0} \int_V \rho dV$$

此即高斯定律,由数学上的奥-高定理,我们可以把上式写为局域的微分形式:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}$$

高斯定律旨在说明:电荷产生电场(电荷是电场的源强度).

18.2.2 电势与电场的环路定律

对点电荷产生的电场 $\mathbf{E} = \frac{Q}{4\pi\varepsilon_0 r^2} \mathbf{e}_r$,再引入以下标量场是方便的:

$$\varphi = \frac{1}{4\pi\varepsilon_0} \frac{Q}{r}$$

这是因为如果计算球坐标下的梯度:

$$\nabla = \mathbf{e}_r \frac{\partial}{\partial r} + \frac{\mathbf{e}_\theta}{r} \frac{\partial}{\partial \theta} + \frac{\mathbf{e}_\varphi}{r \sin \theta} \frac{\partial}{\partial \varphi}$$

便得到以下关系式:

$$\mathbf{E} = -\nabla \varphi$$

即电场可以表示为我们引入的另一个标量场的负梯度的形式.这个标量场就叫做**电势**(electric potential).而以上关系,结合梯度的含义理解,便是沿电场线方向电势降低.可以通过牛-莱定理写为积分形式(下式与路径无关):

$$\int_A^B \mathbf{E} \cdot d\mathbf{l} = \varphi_A - \varphi_B$$

以上关系式由叠加原理,在任意电荷分部的情况下都成立,此时对应的电势标量场为:

$$\varphi(\mathbf{r}) = \int \frac{1}{4\pi\varepsilon_0} \frac{\rho}{|\mathbf{R}|} dV' , \quad \mathbf{R} = \mathbf{r} - \mathbf{r}'$$

对于电势的引入可能存在一点疑惑:我们都应该知道电场的定义是可以独立于场源电荷的,它为在一点处引入静止的试探电荷时单位电荷的受力.那么电势是否可以独立于场源电荷而定义呢?观点一比较基础,它认为在静电场情况下既然全空间电场已经被定义了,它就应该反映了该点处场的全部实际信息,那么电势作为与电场符合一个梯度关系的梯度场,也就能被相应的确定下来,两个可能的电势还可以差一个全空间都一致的常数.也就是这种观点把电场视为更加基础的场的特征.下面便可以知道电势的含义是引入试探电荷后单位试探电荷在该点处的电势能.而观点二则更加本质,我们今天知道了势比场强实际上包含更多的信息.由于微观粒子的量子特性,力的概念已然失效,从而场强反而不如用势来描述一个场更为直接而实用.电磁场论实际上也就是一个矢量场论,它由矢势和标势组成一个四维矢量来完整地描述,但取法不唯一,多出来的冗余自由度被称为**规范**(gauge),它恰好就在表示电荷与场的相互作用时,与电荷的量子相位协同变化,从而造成实际可以观测的影响.为了实际问题的方便,通常约定势有特定的取法.在静电学的常见情况下(电荷分布区域有限时),常常取无穷远处电势为零,注意上电势积分式也恰恰符合这一点.

我们进一步考察电势的物理含义,如果把以上电场视为外电场,引入试探电荷 q ,并在以上积分形式的场强与势的关系中乘以 q ,便发现 $q\mathbf{E}$ 就是电荷的受力,左边就是表示电场力的做功:

$$W_{A \rightarrow B} = q\varphi_A - q\varphi_B$$

很自然地我们发现电场力是保守力,而对应的电势能,基于它应该与电荷量成正比的观点,我们总是直接取为:

$$E_p = q\varphi$$

从而电势的效果,是为试探电荷带来一个电势能.

最后我们再考察电场的一个特性,由于任何路径上的电场积分可以根据牛-莱定理写为端点的电势差的形式,考虑闭合的曲线积分,它可以看成一个可定向曲面的边界,由于两端点重合:

$$\oint_{\partial A} \mathbf{E} \cdot d\mathbf{l} = 0$$

这就是电场的环路定律(circuital law),再根据数学上的开-斯定理,我们写出环路定律的微分形式:

$$\nabla \times \mathbf{E} = \mathbf{0}$$

也就是静电场无旋.值得指出这不过是静电场可以表示为标量电势场的负梯度的直接结果,因为任意一个标量场的梯度都是无旋的¹³:

$$\oint_{\partial A} \nabla f \cdot d\mathbf{l} = 0 \Rightarrow \nabla \times \nabla f = \mathbf{0}$$

18.2.3 总结

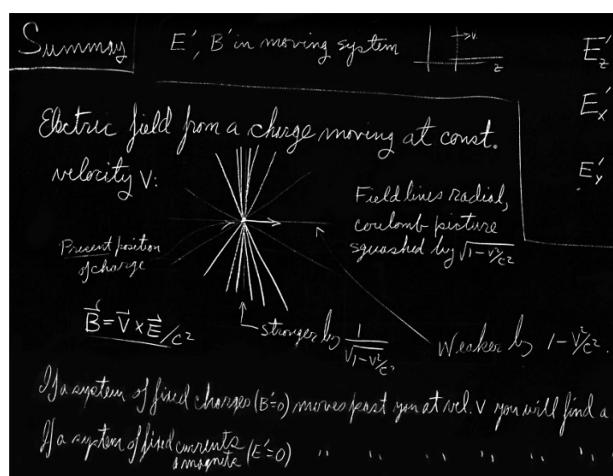


图 18.7: Feynman 讲座一瞥

我们现在有三个不同的层次去看待电荷与电荷之间的相互作用问题¹⁴:

1. 不引入场:

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0} \frac{Qq}{r^2} \mathbf{e}$$

2. 引入场,用场强描述:

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} ; \quad \nabla \times \mathbf{E} = \mathbf{0} \\ \mathbf{F} &= q\mathbf{E} \end{aligned}$$

3. 引入场,用势描述:

$$\nabla^2 \varphi = -\frac{\rho}{\epsilon_0}$$

$$E_p = q\varphi$$

¹³这个结论向流形推广就是著名的庞加莱引理(Poincaré lemma)了,它指出任意微分形式 ω 的二次全微分为零:

$$d\omega = 0$$

一个以后会用得上的推论可以通过考察矢量场的旋度在闭合曲面上的面积分,而把闭合曲面看成是边界曲线缩小到零的有边界曲面而得:

$$\oint_{\partial V} \nabla \times \mathbf{F} \cdot d\mathbf{A} = 0 \Rightarrow \nabla \cdot \nabla \times \mathbf{F} = 0$$

¹⁴拉普拉斯算子 $\nabla^2 = \nabla \cdot \nabla$,如作用在标量上即表示先求梯度后求散度:

$$\nabla^2 f = \nabla \cdot (\nabla f)$$

以上描述一个比一个适用范围更广,第一个仅仅适用于静电场或低速情况,第二个则是对高速运动的电荷与电磁场适用,比如匀速运动的电荷由简单相对论讨论可得:

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \frac{1 - \beta^2}{(1 - \beta^2 \sin^2 \theta)^{3/2}} \mathbf{e}_r$$

$$\mathbf{B} = \frac{\mathbf{v} \times \mathbf{E}}{c^2}$$

它要符合第二个式子(旋度需要做相对论修正),而第三个式子更是适用于非经典场论情况.

18.3 静电能

我们将进一步深入研究电场相互作用中的能量概念.

18.3.1 静电势能

之前已经引入了电势能的概念,我们把外场中引入的试探电荷(点电荷)所具有的势能的大小计做电势能,并得出了电势能的表达式:

$$E_p = q\varphi$$

还指出了,它的作用是用来表示不变的外电场对电荷的做功的多少,基于这样的观察,我们把这个能量称为静电势能(electrostatic potential energy),静电的含义,是要求外场不能随着试探电荷的引入与移动而改变.而更常见的一类情况是,随着试探电荷的移动,场源电荷也在随着改变,如之后将会经常讨论的静电感应的情况.此时仍然可以计算外场的场强与电势,并引入试探电荷以后依然可以形式地写出电势能(electro potential energy)的值 $E_p = q\varphi$,但它不再代表实际移动试探电荷电场对电荷做的功了,它代表什么?电场对电荷做功普遍地应该怎么表示?外力做功又如何和它发生联系,这是我们之后的讨论关心的问题.

首先对于这个问题的第一步考察是,虽然试探电荷的特定位置带来特定的外场,但一旦确定了外场的场源的电荷分布,便可以独立于试探电荷地产生一个分布于全空间的电势与电场,而可以设想固定这个电荷分布不变,而把试探电荷放在空间中的任意一点,讨论其受力与能量.这些改变试探电荷位置而不改变外场的位移称为虚位移(virtual displacement).它是一种广泛用来建立理论模型的方法.它的强大之处在于虚位移的任意性,在每一个实际时刻试探电荷可以在任意位置,外场在缓慢随时间发生改变,而试探电荷发生的所有虚位移中的一个特例.在这个意义上,电势能的含义便在于,试探电荷发生虚位移时,即场源电荷分布不变时,电场力对试探电荷做的虚功即为电势能的减小,此即静电场下的虚功原理:

$$\delta E_p = -q\mathbf{E} \cdot \delta\mathbf{r}$$

我们可以把这个概念推广,如果在该外场中引入不止一个点试探电荷,我们写出一个电荷体系在外场中的电势能来:

$$E_p = \sum_i q_i \varphi_i$$

这个写法存在一些问题,首先是如果没有引入任何试探电荷,体系是否固有一个能量?用场的角度来理解就会发现不对,场是一种物质,物质必定伴随着能量.或者考虑场源电荷之间的相互作用,它也一定会伴随着能量.再者,引入两个以上的点电荷时,它们之间的相互作用能没有被包含在以上项中.第三,即使是单个点电荷单独存在,也必定会带来其自能.我们需要从更基础的角度来讨论电荷体系的能量.

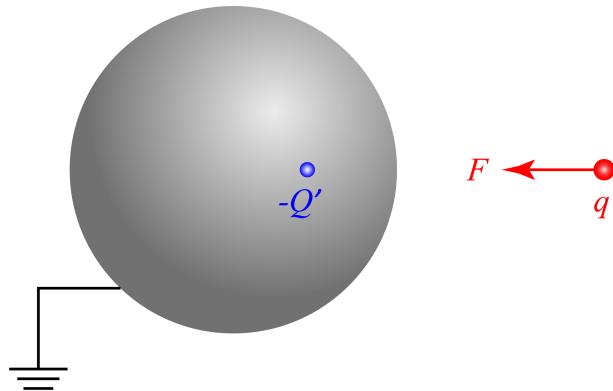


图 18.8: 情形: 场源电荷取决于试探电荷

18.3.2 自能与相互作用能

考虑真空中的两个点电荷之间的相互作用. 电荷 q_1 在电荷 q_2 处产生的电场与电势为 \mathbf{E}_{12} 与 φ_{12} , 电荷 q_2 在电荷 q_1 处产生的电场与电势为 \mathbf{E}_{21} 与 φ_{21} 无疑, 相互作用力是相互的, 满足牛顿第三定律 $q_1\mathbf{E}_{21} + q_2\mathbf{E}_{12} = \mathbf{0}$. 那么电势能是相互的吗? 注意到:

$$q_1 U_{21} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r} = q_2 U_{12}$$

从而两个电势能实际上是相等的. 根据电势能的定义, 它等于 1 不动 2 移到无穷远电场力所做的功, 也会等于 2 不动 1 移到无穷远电场力所做的功. 这两种情况平移对称, 从而我们有理由判断: 无论如何移动两个电荷, 将 1 与 2 分开到无限远时电场力做功都会是以上值, 这一点也不难在数学上证明.

从而我们把这个能量称为 12 间相互作用力对应的相互作用能(interaction energy). 它可以用任何一个电荷乘以另一个电荷在这个电荷处产生的势相乘得到, 或者找到这一对相互作用, 直接写出:

$$I_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}}$$

按照这思路, 由叠加原理, 我们可以很轻易地写出点电荷体系与连续分布电荷体系的总相互作用能:

1. 点电荷体系:

$$I = \sum_{\{i,j\}!} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$

2. 连续分布电荷体系:

$$I = \iint_{\{\mathrm{d}V, \mathrm{d}V'\}!} \frac{1}{4\pi\epsilon_0} \frac{\rho \rho'}{r} \mathrm{d}V \mathrm{d}V'$$

关于以上求和我们做出以下解释:

首先, 点电荷体系的相互作用能表示把点电荷从计算这个能量的状态移动到点电荷彼此之间都相距无穷远时电场力所做的功. 但一定要注意每个点电荷移动到无穷远以后还会因为自身的电磁相互作用而产生一个待定的能量, 叫做自能, 我们紧接着会讨论到这个问题. 而连续分布电荷体系则表示把每一份电荷微元移动到无穷远以后电场力做的功, 移动到无穷远以后电荷无限分散(宏观意义上), 所以每一份微元电荷对应的电磁相互作用能真实地趋于零, 故我们计算的这个能量就是这个体系的总电磁相互作用对应的能量.

第二, 是这个求和表示的含义. 我们用 (i, j) 来表示特定集合中两个元素 i, j 构成的有序对(ordered pair), 而 $\{i, j\}$ 则表示交换 i, j 后表示同一代数元素的¹⁵对称无序对(symmetrized unordered pair), 即 $\{i, j\} = \{j, i\}$. 求和下加这个对符号表示对所有一对一对的情况进行求和. 每一对只需要要求一次即可. 这之后还跟着一个感叹号, 它表示 i, j 必须取不同值(而 $\mathrm{d}V$ 与 $\mathrm{d}V'$ 可以取相同值, 这些项会在体积很小时求和也很小而可以忽略). 那么由简单的代数变形, 我们发现上两式即:

$$I = \frac{1}{2} \sum_i \sum_{j \neq i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} = \frac{1}{2} \sum_i q_i \phi_i$$

$$I = \frac{1}{2} \int \int_{\mathrm{d}V \mathrm{d}V'} \frac{1}{4\pi\epsilon_0} \frac{\rho \rho'}{r} \mathrm{d}V \mathrm{d}V' = \frac{1}{2} \int \rho \varphi \mathrm{d}V$$

注意到第一式中 ϕ_i 表示的是别的电荷在 q_i 处产生的电势, 不包括自己在自己上产生的, 我们把符号稍微变动一下以示区别. 而在后一个表达式中则不需要考虑这种区别, 这个式子将作为一个普遍的求总电磁相互作用能的表达式而使用.

¹⁵以后还会用到 n 个元素的对称无序类(symmetrized unordered setoid) 与反对称无序类(antisymmetrized unordered setoid):

$$\{i_1 \cdots i_p \cdots i_q \cdots i_n\} = \{i_1 \cdots i_q \cdots i_p \cdots i_n\}$$

$$[i_1 \cdots i_p \cdots i_q \cdots i_n] = -[i_1 \cdots i_q \cdots i_p \cdots i_n]$$

第三,是这个求和的使用方法.实际上,如果电荷连续分布系统从一个状态变化到另一个状态,电荷分布与它产生的电势分布都发生改变,那么前后的能量增加:

$$\Delta I = \frac{1}{2} \int \rho' \varphi' dV - \frac{1}{2} \int \rho \varphi dV$$

自然表示电场对电荷体系所做的负功,或者表示由电荷流向电场的能量,它使得场的能量,也就是 I ,发生了增长¹⁶.而反过来考察电荷体系,如果外力对电荷体系做功为 W ,那么根据动能定理应该有:

$$W - \Delta I = \Delta E_k$$

在十分缓慢地移动电荷的情况下,亦或是导体在载流子为电子,其动能十分的小可以忽略的情况下,动能的增长可以视为零,从而有:

$$W = \Delta I$$

即,外力做功,通过电荷为媒介,转化为了场的能量的增长.这个观点也是我们经常用到的.

我们最后回到电势能与相互作用能的关系,考虑整个体系由两部分电荷体系构成,那么由叠加原理,整个空间中每一点的电势可以写成两部分贡献的和 $\varphi_1 + \varphi_2$.我们只需要对全空间的电荷积分,就得到了整个体系的相互作用能:

$$\begin{aligned} I &= \frac{1}{2} \int_{I+II} \rho \varphi dV \\ &= \frac{1}{2} \int_I \rho_1 (\varphi_1 + \varphi_2) dV_1 + \frac{1}{2} \int_{II} \rho_2 (\varphi_1 + \varphi_2) dV_2 \\ &= \frac{1}{2} \int_I \rho_1 \varphi_1 dV_1 + \frac{1}{2} \int_{II} \rho_2 \varphi_2 dV_2 + \frac{1}{2} \left(\int_I \rho_1 \varphi_2 dV_1 + \int_{II} \rho_2 \varphi_1 dV_2 \right) \\ &= I_1 + I_2 + I_{12} \end{aligned}$$

我们于是找到了三个能量, I_1, I_2 是两个电荷分布体系单独存在时各自的自相互作用能, 我们把它简称作**自能**(self-energy), 它一定是大于零的, 这一点下一节就可以发现. 而余下的交叉求和项就是表示两个体系之间的相互作用能. 仔细一看不难发现一项其实就是电荷体系 II 产生的外场下电荷体系 I 的电势能, 而另一项就是电荷体系 I 产生的外场下电荷体系 II 的电势能, 两者仔细研究下发现其实相等:

$$I_{12} = \iint_{I II} \frac{1}{4\pi\epsilon_0} \frac{\rho_1 \rho_2}{r} dV_1 dV_2 = \int_I \rho_1 \varphi_2 dV_1 = \int_{II} \rho_2 \varphi_1 dV_2$$

这一项可以大于零也可以小于零, 表示的是相互作用. 现在我们试探电荷体系在外场中的电势能有了一个新的看法, 它其实表示试探电荷体系与场源电荷体系的相互作用能, 作为相互作用, 它也可以反过来用场源电荷体系在试探电荷体系产生的场中的电势能来计算, 或是对称地, 用两者和的一半来计算.

最后让我们来看一下点电荷的自能. 实际上在之前点电荷体系相互作用能体系中我们已经指出, 其表达式中电荷 q_i 上电势 ϕ_i 与真实电势 φ_i 的区别就在于没有计及点电荷在自己位置处产生的电势 φ_{ii} , 从而漏掉的项其实就是点电荷的自能:

$$I_i = \frac{1}{2} q_i \varphi_{ii}$$

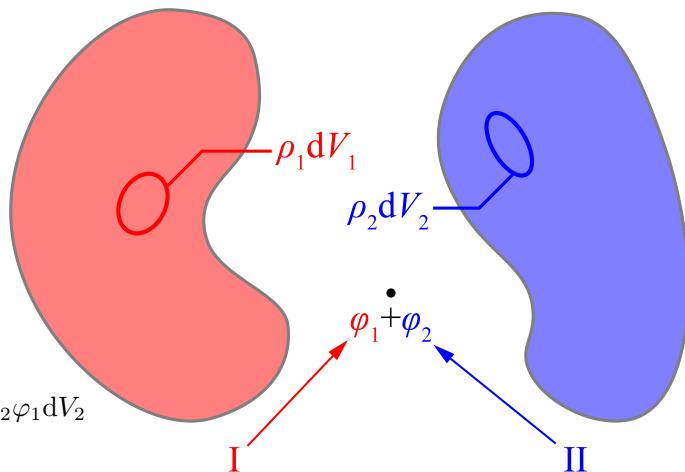


图 18.9: 自能与相互作用能

¹⁶场只能和带荷的物质发生作用, 这也是唯一的使得场能量变化的方法

这些项不会影响我们的实际计算,因为点电荷总被我们认为是没有内部改变的理想体系,我们经常把微观基本粒子做这样的抽象,因为它们都是小到可以忽略其尺寸.当粒子速度远小于光速,且不涉及到高能粒子反应的情况下,它们也的确不发生任何内部的改变,自能精确地保持不变.所以只需要计算扣除点电荷自能的相互作用能表达式,就能够知道有多少势能被存储在场中,就能够用它来写出电荷分布发生改变时释放出来的能量.

点电荷的自能如何计算呢?这取决于我们采取的模型.如果我们考虑的是一个电荷均匀分布在表面的球体,那么自能为:

$$I = \frac{1}{2}Q \cdot \frac{1}{4\pi\varepsilon_0} \frac{Q}{r}$$

对于电子,如果我们认为电子的静质量全部来源于它携带的场的惯性,那么能够写出:

$$I = \frac{e^2}{8\pi\varepsilon_0 r} = m_e c^2$$

事实上我们并不能确定电子内部电荷分布,所以我们更常直接定义**经典电子半径**(classical electron radius) 为:

$$\frac{e^2}{4\pi\varepsilon_0 r_e} = m_e c^2 \Rightarrow r_e = \frac{1}{4\pi\varepsilon_0} \frac{e^2}{m_e c^2} = 2.82 \times 10^{-15} \text{ m}$$

然而这仅仅是一个方便计算的实用常数,并不能代表电子的真实情况!事实上用不带电荷的粒子与电子发生散射实验就会发现,电子的尺寸比这要小得多,上面的计算结果甚至比质子的半径还要大几倍!而实验更是发现电子在多种意义上就是一个点粒子,根本没有内部组成结构,十分稳定不能衰变.那么显然从上式发现这也面临着自能发散的问题.双重矛盾如何解决呢?

解决方案是量子力学的引入,电子不是点粒子,在原子尺度下电子就已经必须要由波函数描述了,电子没有同时确定的位置和速度,在原子外形成“电子云”.狄拉克(P. Dirac)于1928年提出**狄拉克方程**(Dirac equation),把电子像光子那样处理为一个传播的“场”,从而从根本上摒弃了电子作为点电荷的观点,电子的质量被理解为场的质量或者静能,而电子的电荷量被理解为与电磁场(光子)的耦合常数.从而把电荷作为一种物质和场物质彻底分割开来独立描述.还同时一并给出了电子的自旋与预言了电子的反物质:**正电子**(positron)的存在.它们都被天然而巧妙的包含在一个统一的代数形式中,极大程度地推动了物理学理论的进展.

但这没有给出问题的全部解释,对于点电荷自能发散的问题这种做法仍然没有给出绝对的答案.经典地看一个匀速运动的经典无穷小点电荷会携带着一个发散的场共同向前运动,这在量子场论语言下,变成了电子与光子相互作用的无限可能性求和,从而也理所当然地引进了各种各样的发散.对于这些发散物理学家们发展出了各种各样的处理方法,但直到今天还仅仅停留在现象解释的层面,背后是否有更加简单的解释方法还是一个谜.

18.3.3 电场能

我们已经十分熟悉,电荷之间的相互作用能实际上就是场的能量.我们现在将给出这一说法的定域化描述.记得一个电荷连续分布体系的相互作用能为:

$$I = \frac{1}{2} \int_V \rho \varphi dV$$

而又根据场与电荷之间的关系:

$$\nabla^2 \varphi = -\frac{\rho}{\varepsilon_0}$$

我们把以上积分写为:

$$I = -\frac{\varepsilon_0}{2} \int_V \varphi \nabla^2 \varphi dV$$

再根据著名的¹⁷第一格林等式(Green's first identity):

$$\nabla \cdot (\phi \nabla \psi) = \phi \nabla^2 \psi + \nabla \phi \cdot \nabla \psi$$

我们再取 ϕ, ψ 都为 φ , 从而:

$$\varphi \nabla^2 \varphi = \nabla \cdot (\varphi \nabla \varphi) - (\nabla \varphi)^2$$

带入以上积分, 我们得到:

$$\begin{aligned} I &= \frac{\varepsilon_0}{2} \int_V (\nabla \varphi)^2 dV - \frac{\varepsilon_0}{2} \oint_{\partial V} \varphi \nabla \varphi \cdot dA \\ &= \int_V \frac{\varepsilon_0}{2} \mathbf{E}^2 dV + \frac{\varepsilon_0}{2} \oint_{\partial V} \varphi \mathbf{E} \cdot dA \end{aligned}$$

我们总是考虑电荷分布有限的情况. 此时最初的积分体积只要包含所有电荷即可, 那么我们最后得到的这个恒等式就十分的有意思, 它将一个体系的相互作用转化为了一个量的体积分和另一个量在包含所有电荷的面上的面积分的形式. 我们如果取体积为无穷大的全空间, 那么考虑到随着远离电荷体系 R 处 φ 以至多 $\frac{1}{R}$ 的方式衰减(若体系总电荷量为零则衰减更加剧烈), 而 \mathbf{E} 以至多 $\frac{1}{R^2}$ 的方式衰减, 而积分面积以 R^2 的方式增加, 故整个积分以至多 $\frac{1}{R}$ 的方式衰减, 在无穷远处这个积分的第二项就会趋于零.

从而我们得到了:

$$I = \int \frac{\varepsilon_0}{2} \mathbf{E}^2 dV$$

从而发现相互作用能可以用场来定域地计算. 电场的能量密度(energy density) 为:

$$w_e = \frac{\varepsilon_0}{2} \mathbf{E}^2$$

场的能量不再由电荷造成, 而是由场自己的场强去定义, 以后还会进一步证明, 场在一点处对电荷有力的作用, 将会使场的动量密度发生改变, 而对电荷有做功, 将会影响这一点处的能量密度.

18.4 电荷体系

讨论若干电荷体系对于之后问题的理解是大有帮助的. 这些包括:

18.4.1 电偶极子

我们都应该知道一个电荷体系如果带电, 那么尽管在离电荷体系很近处电场分布不均匀, 但离开分布电荷足够远处的场强却总是趋于一个点电荷的电场. 常见的情况发生在原子核中, 原子核由质子和中子组成, 质子带电而中子不带电. 原子核内电荷分布总是不均匀的, 但在考虑外界电子受力时总是将其产生的场视为库仑场. 这是因为电子典型半径: 波尔半径与原子核的尺寸比差了 5 个数量级. 在这个距离处电场完全可以看成是库仑场, 它与库仑场的偏离将作为一个可以忽略的小量而不去考虑.

但类似的, 我们将考虑到很多不带净电荷的电荷分布体系, 它们在近处产生一个不均匀的电场, 而在远处产生的电场

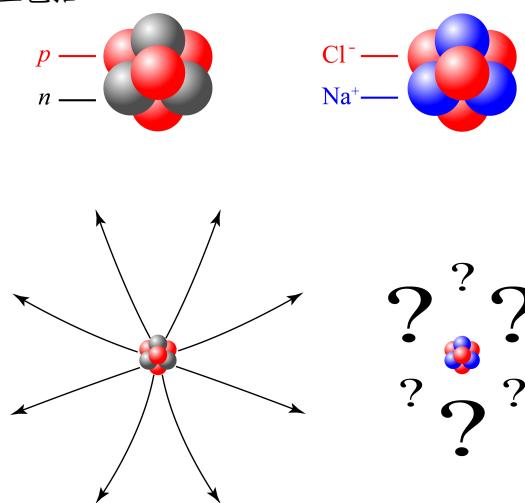


图 18.10: 远方的场

¹⁷而第二格林等式(Green's second identity) 在格林函数法解边值问题时将十分有用:

$$\nabla \cdot (\phi \nabla \psi - \psi \nabla \phi) = \phi \nabla^2 \psi - \psi \nabla^2 \phi$$

却总是存在着某些规律。这个问题有重要的实用价值。比如我们如果从 NaCl 晶体中取出一小块来，它的总电量总是为零，但由于晶体的变形等因素导致正负电荷产生的电场并不能完全抵消，这个电场将会与晶体的动力学耦合，极大地影响晶体的动力学特征。又比如很多分子实际上属于极性分子(polar molecule)，由于原子的电负性(electronegativity)差异，原子间的键合具有极性，在很多情形下一个原子对公用电子对的吸引能力更强，导致电子集中在一个原子一侧。从而负电荷的中心向这个原子位移。只要分子结构上的对称性不会使得键的极性不会相互抵消，整个分子就会显示出极性来，它是溶液中溶剂溶解机理的关键，也会影响物质的熔点、沸点、极化率等等关键的物理特性。

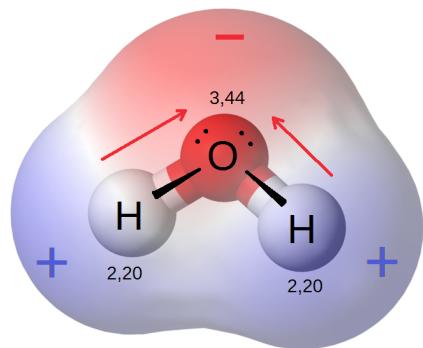


图 18.11: 水分子极性

所以我们关心总电荷量为零的体系在很远处产生的电场，理论上我们可以采取更完整的多极展开(multipole expansion)的做法，在很多情况下电偶极项(dipole term)成为了这种情形下的领头项(leading order)。我们定义小范围分布电荷体系中所有正电荷的“中心”和负电荷的“中心”：

$$\mathbf{r}_+ = \frac{\sum_{q_i > 0} q_i \mathbf{r}_i}{Q}, \quad Q = \sum_{q_i > 0} q_i$$

$$\mathbf{r}_- = \frac{\sum_{q_i < 0} q_i \mathbf{r}_i}{-Q}, \quad -Q = \sum_{q_i < 0} q_i$$

而体系的电偶极矩(dipole moment)就被定义为由负电荷指向正电荷的矢量：

$$\mathbf{d} = Q(\mathbf{r}_+ - \mathbf{r}_-)$$

或者写为：

$$\mathbf{d} = \sum_i q_i \mathbf{r}_i$$

求和遍及所有电荷，无论正负。由于总电荷量是零，不难验证这个矢量与坐标原点的选取无关，为了纪念物理学家德拜(P. Debye)在分子物理领域的相关开创性工作，它一般使用以德拜命名的单位，它是一个高斯制单位：

$$\begin{aligned} 1 \text{ D} &= 10^{-18} \text{ statC} \cdot \text{cm} \\ &= \frac{1}{299792458} \times 10^{-21} \text{ C} \cdot \text{m} \\ &= 3.34 \times 10^{-30} \text{ C} \cdot \text{m} \\ &= 0.208 \text{ e} \cdot \text{\AA} \end{aligned}$$

为何定义这样的一个矢量呢？我们考察在原点附近分布了一个电荷体系 $\{q_i, \mathbf{r}_i\}$ ，那么在很远处产生的电势应该为：

$$\varphi(\mathbf{r}) = \sum_i \frac{1}{4\pi\epsilon_0} \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}$$

而对于分母中的矢量的模，我们考虑到 $|\mathbf{r}| \gg |\mathbf{r}_i|$ ，可以做如下近似：

$$\begin{aligned} \frac{1}{|\mathbf{r} - \mathbf{r}_i|} &= [(r - r_i)^2]^{-\frac{1}{2}} = (r^2 - 2\mathbf{r}_i \cdot \mathbf{r} + r_i^2)^{-\frac{1}{2}} \\ &\approx (r^2 - 2\mathbf{r}_i \cdot \mathbf{r})^{-\frac{1}{2}} = \frac{1}{r} \left(1 - 2\frac{\mathbf{r}_i \cdot \mathbf{r}}{r^2}\right)^{-\frac{1}{2}} \\ &\approx \frac{1}{r} \left(1 + \frac{\mathbf{r}_i \cdot \mathbf{r}}{r^2}\right) = \frac{1}{r} + \frac{\mathbf{r}}{r^3} \cdot \mathbf{r}_i \\ &= \frac{1}{r} + \frac{\mathbf{e}_r}{r^2} \cdot \mathbf{r}_i \end{aligned}$$

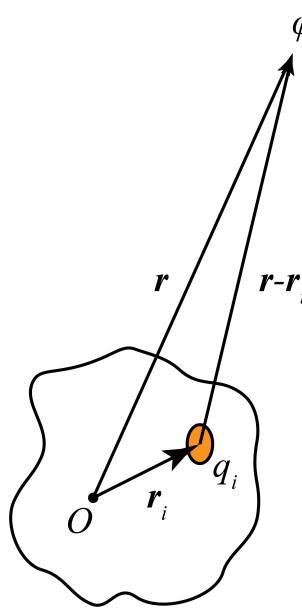


图 18.12: 偶极展开

代入求和表达式, 得:

$$\varphi(\mathbf{r}) \approx \frac{1}{4\pi\epsilon_0} \frac{\sum_i q_i}{r} + \frac{1}{4\pi\epsilon_0} \frac{\mathbf{e}_r \cdot \sum_i q_i \mathbf{r}_i}{r^2}$$

第一项求和由于总电荷量为零放弃了成为领头项的机会, 而第二项凸显了出来, 根据我们之前定义的电偶极矩, 它实际上就是:

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{e}_r \cdot \mathbf{d}}{r^2} \quad (r \gg d)$$

我们取了等号, 条件是我们考察的点的矢径 r 远大于电荷分布的平均尺寸 d . 而理论上常常构造一种十分独特的体系, 我们让一对电量为 $+Q$ 与 $-Q$ 严格相等的点电荷相离 $\mathbf{r}_+ - \mathbf{r}_- = \mathbf{l}$, 再令 $Q \rightarrow \infty$ 而 $\mathbf{l} \rightarrow \mathbf{0}$, 而保持 $\mathbf{p} = Q\mathbf{l}$ 的值保持不变, 其极限就是抽象的模型: 点电偶极子(point dipole), 它在除原点外, 全空间产生的电势都可以用上式来表示. 这也是我们以后讨论的中心. 如果在点电偶所在空间建立球极坐标系, 而把电偶极矩矢量指向的方向作为极轴, 我们可以写出电势为:

$$\varphi(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{p \cos \theta}{r^2}$$

这是一个与半径平方反比的电势, 它与角度有关. 利用球坐标下梯度算符的表示:

$$\nabla = \mathbf{e}_r \frac{\partial}{\partial r} + \frac{1}{r} \frac{\partial}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi}$$

我们计算出点电偶极子的电场来:

$$E_r(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{2p \cos \theta}{r^3}$$

$$E_\theta(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{p \sin \theta}{r^3}$$

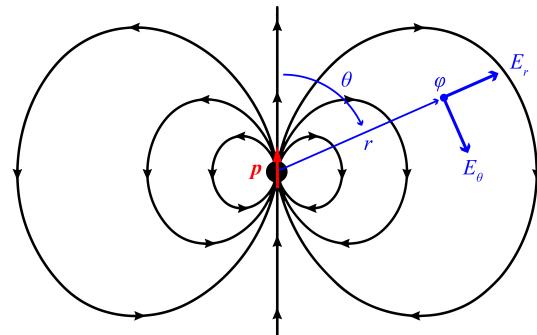


图 18.13: 点电偶极子的场

可见电场倒是三次反比了. 它比点电荷产生的场随距离减少地更快. 理论研究经常会把以上两个分量式组合为一个矢量式, 不难验证¹⁸

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{3(\mathbf{e}_r \cdot \mathbf{p})\mathbf{e}_r - \mathbf{p}}{r^3}$$

如果考虑电偶极子在外场中的情形, 采用两电荷分布极限为点电偶极子的结果, 我们很容易发现它的电势能为:

$$V = Q(\varphi_+ - \varphi_-) = Q\mathbf{l} \cdot \nabla\varphi = -\mathbf{p} \cdot \mathbf{E}$$

注意到根据电势能的含义, 这个能量是不包含构成电偶极子的两电荷之间的相互作用能的, 而在电偶极子在电场中发生位移的过程中我们也应该不能让电偶极子发生结构的改变: 即, \mathbf{p} 的大小不能改变, 但 \mathbf{p} 的方向, 位置都可以发生改变, 随着这两个剩余自由度发生变化, 电场对电偶极子做的虚功 $-\delta V = \delta[\mathbf{p} \cdot \mathbf{E}(\mathbf{r})]$ 就被理解为力矩与受力:

$$\delta\mathbf{r} = \mathbf{0}, \delta\mathbf{p} = \delta\theta \times \mathbf{p} : -\delta V = \mathbf{M} \cdot \delta\theta \Rightarrow \mathbf{M} = \mathbf{p} \times \mathbf{E}$$

$$\delta\mathbf{r} \neq \mathbf{0}, \delta\mathbf{p} = \mathbf{0} : -\delta V = \mathbf{F} \cdot \delta\mathbf{r} \Rightarrow \mathbf{F} = \mathbf{p} \cdot \nabla\mathbf{E}$$

其中表达式 $\mathbf{p} \cdot \nabla\mathbf{E}$ 的含义为取沿 \mathbf{p} 方向的矢量场 \mathbf{E} 的方向导数. 以后会提到导体或电介质颗粒将会在外电场下感应或极化出一个电偶极子来:

$$\mathbf{p} = \alpha\mathbf{E}$$

¹⁸这个表达式对原点以外的点适用, 如果考虑原点的奇异电场特性, 并引入狄拉克德尔塔函数(Dirac delta function), 则更准确的表达式为:

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{3(\mathbf{e}_r \cdot \mathbf{p})\mathbf{e}_r - \mathbf{p}}{r^3} - \frac{\mathbf{p}}{3\epsilon_0} \delta^3(\mathbf{0})$$

把此式代入受力公式, 我们发现:

$$\mathbf{M} = 0 \quad ; \quad \mathbf{F} = \frac{\alpha}{2} \nabla E^2$$

这也就是我们常说的电场有能够吸引轻小物体的本领的解释, 我们发现这些小物体在外电场中的受力方向总是指向电场强度大小增加的方向的, 而与电场强度方向关系不大.

18.4.2 电荷密度

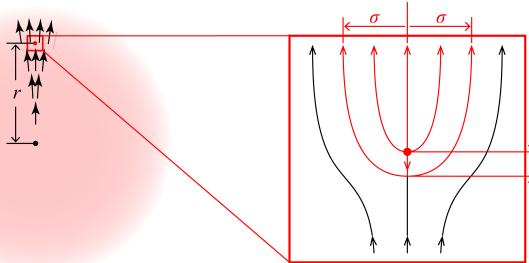


图 18.14: 连续分布点电荷产生平均电场

我们经常会用到连续分布的电荷这样的说法, 把电荷分布写成一个 $\rho(\mathbf{r})$ 的形式, 表示每一个体积元 dV 中含有的电荷量为 ρdV . 这种做法在分析理论问题时发挥了很重要的作用, 但是它却并不实际. 实际电荷分布在很小的尺度看来永远是不连续的, 经典情况下一般可以视为由基本电荷量的点粒子产生的. 那么正如前述, 在这样的情况下电荷密度被理解为:

$$\rho = \sum_i n_i q_i$$

更难理解的此时电场强度的定义. 让我们考虑一个均匀带点的球, 此时球外的电场强度自然可以简单的按 $\frac{Q}{4\pi\epsilon_0 r^2}$ 来计算. 若内部的电场若按照完全的连续电荷分布来计算, 给出了:

$$\mathbf{E} = \frac{\rho \mathbf{r}}{3\epsilon_0}$$

的结果. 但这个结果忽略了电荷的离散性, 我们如果认为这个球是由电荷数密度为 n , 电荷量为 q 的大量点电荷构成. 那么设在半径为 r 处的电场与单个点电荷在距中心为 r' 处的电场相等. 我们可以得到:

$$\frac{nqr}{3\epsilon_0} = \frac{q}{4\pi\epsilon_0 r'^2}$$

如果一个电荷的平均占据体积视为球体, 半径为 R , 那么电荷数密度与它的关系为:

$$n \cdot \frac{4}{3}\pi R^3 = 1$$

这样就得到:

$$\frac{r'}{R} = \sqrt{\frac{R}{r}}$$

或者用 r 区域内包含的电荷数量 $N = \frac{4}{3}\pi r^3 / \frac{4}{3}\pi R^3$, 得到:

$$r' = \frac{R}{N^{1/6}}$$

还有一种看法是考察一个电荷发出的电场线在远方形成匀强电场以后的截面:

$$\frac{nqr}{3\epsilon_0} \cdot \pi\sigma^2 = \frac{q}{\epsilon_0}$$

我们很容易发现:

$$\sigma = 2r$$

也就是说,离开中心区域越远,包含的总电荷数量越多,该处点电荷产生的电场相对与由内部大量电荷积累而来的大的几乎均匀的电场相比就足够地小.点电荷使得外电场偏离均匀的空间范围就远小于电荷与电荷之间的间距,使得电荷之间的空间就可以视为一个几乎均匀的场.

最后我们总结,在用堆放点电荷的方式来得到一个电荷密度时,尽管微观来看产生的电磁场仍然很不均匀,但我们可以抽象出来每一点的平均场强的概念.它的意义有二:一是下一节即将证明的该点局部微元体内电场的平均值.二是如果保证电荷密度不变电荷无限分散的极限情形下该点处的实际场强.它的算法依然是:

$$\mathbf{E} = \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{\rho e_R}{R^2} dV'$$

而既然电场算法与真实的连续分布的电荷没有区别.电势,局部电荷受力体密度就仍然可以这样计算:

$$\varphi = \int \frac{1}{4\pi\epsilon_0} \frac{\rho}{|\mathbf{R}|} dV' , \quad \mathbf{f} = \rho \mathbf{E}$$

18.4.3 极化强度 *

往往,平常状态下的物质的真实情况是,所有局部都不是由单一的点电荷单元堆积而成.就比如液体的水,它的重复单元是电中性的水分子.尽管电中性,对外表现出偶极子,在原场情况下近似为点电偶极子,便是其电势电场的领头项.故我们需要考虑的实际情况是:单位体积内有 $n(\mathbf{r})$ 个偶极矩为 $\mathbf{p}(\mathbf{r})$ 的点电偶极子.这里 \mathbf{p} 不会是常矢量,首先其方向往往随着空间位置逐渐在改变,而每个电偶极子的电偶极矩大小亦可能因为极化程度的不同而有所区别,下一章我们要介绍的位移极化就是这样的情形.我们将数密度 n 与电偶极矩 \mathbf{p} 相乘,得到单位体积内的电偶极矩,即**极化强度**(polarization density):

$$\mathbf{P} = n\mathbf{p}$$

对于连续情况,我们发现电势应该用电偶极子的电势进行体积分来得到:

$$\varphi = \int \frac{1}{4\pi\epsilon_0} \frac{\mathbf{P} \cdot \mathbf{e}_R}{R^2} dV' , \quad \mathbf{R} = \mathbf{r} - \mathbf{r}'$$

然而我们发现以下等式是成立的:

$$\begin{aligned} \nabla' &= e_x \frac{\partial}{\partial x'} + e_y \frac{\partial}{\partial y'} + e_z \frac{\partial}{\partial z'} \\ \nabla' \cdot \frac{\mathbf{P}}{|\mathbf{R}|} &= \frac{\nabla' \cdot \mathbf{P}}{|\mathbf{R}|} + \mathbf{P} \cdot \nabla' \frac{1}{|\mathbf{R}|} \\ \nabla \frac{1}{|\mathbf{R}|} &= \nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|} = -\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} = -\frac{\mathbf{e}_R}{R^2} , \quad \nabla' \frac{1}{|\mathbf{R}|} = \nabla' \frac{1}{|\mathbf{r} - \mathbf{r}'|} = -\frac{\mathbf{r}' - \mathbf{r}}{|\mathbf{r}' - \mathbf{r}|^3} = \frac{\mathbf{e}_R}{R^2} \\ \Rightarrow \quad \frac{\mathbf{P} \cdot \mathbf{e}_R}{R^2} &= \nabla' \cdot \frac{\mathbf{P}}{|\mathbf{R}|} - \frac{\nabla' \cdot \mathbf{P}}{|\mathbf{R}|} \end{aligned}$$

这样我们就得到在区域 V' 内存在极化矢量 \mathbf{P} 时在全空间产生的电势的公式:

$$\begin{aligned} \varphi &= \int_{V'} \frac{1}{4\pi\epsilon_0} \left(\nabla' \cdot \frac{\mathbf{P}}{|\mathbf{R}|} - \frac{\nabla' \cdot \mathbf{P}}{|\mathbf{R}|} \right) dV' \\ &= \oint_{\partial V'} \frac{1}{4\pi\epsilon_0} \frac{\mathbf{P} \cdot dA'}{|\mathbf{R}|} + \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{-\nabla' \cdot \mathbf{P} dV'}{|\mathbf{R}|} \\ &= \oint_{\partial V'} \frac{1}{4\pi\epsilon_0} \frac{\sigma_P dA'}{|\mathbf{R}|} + \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{\rho_P dV'}{|\mathbf{R}|} \end{aligned}$$

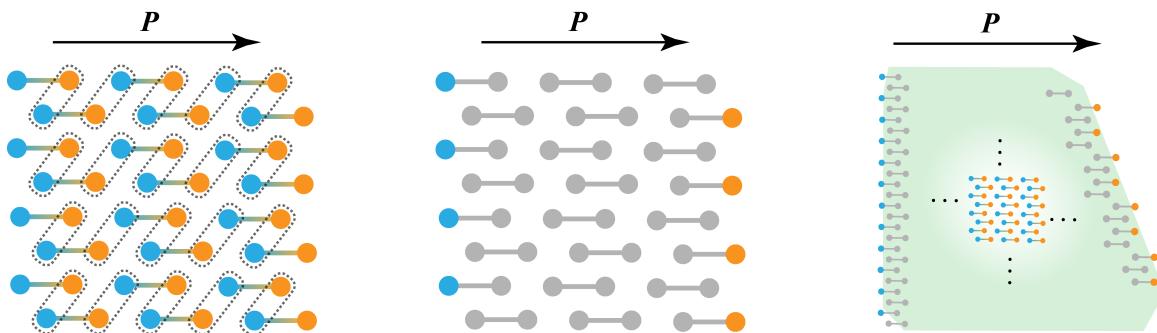


图 18.15: 极化强度导致表面电荷分布

从以上表达式看得出来, 最后电势有一种等效的求法, 就是认为在体内有一个 $\rho_P = -\nabla' \cdot \mathbf{P}$ 的电荷密度分布, 而认为在体表面的面元 dA 上带有面电荷 $dQ = \mathbf{P} \cdot d\mathbf{A}$, 或者认为有面电荷密度 $\sigma_P = \mathbf{P} \cdot \mathbf{n}$, 其中 \mathbf{n} 为体积表面的外法向量. 与其说这是一种等效, 不如说这只是对实际情况的另一种理解. 如上图所示, 这是一块均匀极化的介质, 我们这样理解介质表面上会出现的电荷: 对介质内部的电荷进行合适的分组使得一对一对的正负电荷产生的场互相抵消. 这样便只剩下了表面的电荷能产生场了. 在几何学上不难证明, 如果电偶极子长 l , 两电荷为 q , 体密度为 n , 那么在与极化矢量平行的方向看两端未配对电荷的面密度恰好为 $\sigma = nql = np = P$. 再考虑面可能的倾斜便容易发现以上极化电荷面密度 σ_P 的公式. 类似地, 体电荷密度 ρ_P 对应于沿某方向的 \mathbf{P} 分量在变化的情形. 具体的物理图像也请读者加以分析与思考.

我们有把握预期空间各处的电场强度的公式, 也可以通过上述方式等效, 毕竟我们可以通过电势计算电场:

$$\mathbf{E} = \oint_{\partial V'} \frac{1}{4\pi\epsilon_0} \frac{\sigma_P dA'}{|\mathbf{R}|^2} \mathbf{e}_R + \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{\rho_P dV'}{|\mathbf{R}|^2} \mathbf{e}_R$$

然而, 若要开始考虑电偶极子在场中受到的外力, 那就势必要考虑连续分布的电荷体系产生的电场和微观非均匀分布的电荷体系的局部场之间的区别. 我们会惊人地发现以上电场只能作为平均意义上的**宏观场**(macroscopic field) 而区别于电偶极子具体感受到的**局域场**(local field). 首先让我们来计算一个概念, 一个球体内的平均场强 $\bar{\mathbf{E}}$. 它会计算球体内电荷的受力提供关键线索. 以原点为中心, 半径为 R 的球内积分直接得到:

$$\bar{\mathbf{E}} = \int_{\text{Ball}} \mathbf{E} dV / \frac{4}{3}\pi R^3$$

直接计算会具有相当的难度. 这里采用的技巧是在分子积分中引入 ρ' 常数, 把积分看做原来的电荷体系产生的电场与均匀分布在这个球体内部的电荷体系之间的相互作用力¹⁹. 这样的球体产生的电场强度为:

$$\mathbf{E}' = \begin{cases} \frac{\rho' r}{3\epsilon_0} \mathbf{e}_r & (r < R) \\ \frac{\rho' R^3}{3\epsilon_0 r^2} \mathbf{e}_r & (r > R) \end{cases}$$

从而上式被等价写作:

$$\bar{\mathbf{E}} = \int_{\text{Ball}} \rho' \mathbf{E} dV / \frac{4}{3}\pi\rho' R^3 = \int_{\text{Ball}} \mathbf{E} \nabla \cdot \mathbf{E}' dV / \frac{4}{3}\pi\rho' R^3 = \int_{\text{all}} \mathbf{E} \nabla \cdot \mathbf{E}' dV / \frac{4}{3}\pi\rho' R^3$$

最后一步 all 改为对全空间积分. 由牛顿第三定律看出, 上式通过某种化简²⁰后可以得到:

$$\int_{\text{all}} \mathbf{E} \nabla \cdot \mathbf{E}' dV + \int_{\text{all}} \mathbf{E}' \nabla \cdot \mathbf{E} dV = 0$$

¹⁹ 数学上相当于把 \mathbf{r} 视作 \mathbf{r} 的散度的 $1/3$, 即:

$$\nabla \cdot \mathbf{r} = \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 3$$

²⁰ 而数学上的证明需要用到张量分析, 过程略去.

于是把原电场对应的电荷密度记做 ρ , 待求的量变成:

$$\bar{\mathbf{E}} = - \int_{\text{all}} \rho \mathbf{E}' dV / \frac{4}{3} \pi \rho' R^3 = \frac{1}{4\pi\epsilon_0} \frac{- \int_{\text{Ball}} \rho r dV}{R^3} + \int_{\text{all-Ball}} \frac{1}{4\pi\epsilon_0} \frac{\rho(-\mathbf{e}_r)}{r^2} dV$$

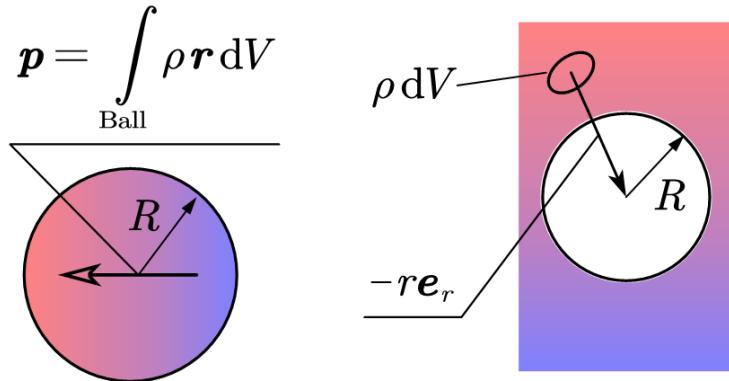


图 18.16: 球内平均电场的两项积分

通过以上结果我们发现, 在一个电荷分布体系中的一个球体内的平均电场取决于两项积分. 第一项不是别的, 就是直接取决于球内电荷分布产生的电偶极矩 \mathbf{p} . 第二项则代表外界电荷直接对球心处产生的电场²¹.

接下来可以进行以下推演: 如果外界电荷不是那么紧密地靠近球面, 那么我们可以通过球面把电荷体系分为两部分. 现在我们来关心内部这些电荷的受力. 依然从电场考虑, 由于外部电荷离球面较远, 故可以近似认为在球内产生一个均匀的电场 \mathbf{E}_{ex} , 甚至于它就可以近似地等于对球心处的场强, 这样我们就把之前的式子写为:

$$\bar{\mathbf{E}} = \frac{1}{4\pi\epsilon_0} \frac{-\mathbf{p}}{R^3} + \mathbf{E}_{\text{ex}} \Rightarrow \mathbf{E}_{\text{ex}} = \bar{\mathbf{E}} + \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p}}{R^3}$$

如果让求平均的体积非常小, 而电荷又总是连续化, 那么 $\bar{\mathbf{E}}$ 又应当与之前通过极化电荷分布计算的体内点场强 \mathbf{E} 无异. 我们已经观察到一个十分关键的信息: 考虑球内电荷受力时应当使用的外界的场 \mathbf{E}_{ex} , 它与忽略电荷不连续性而计算出来的“平均场” $\bar{\mathbf{E}}$ 存在一个差. 最后考虑到极化强度的定义 $\mathbf{P} = \lim_{R \rightarrow 0} \mathbf{p} / \frac{4}{3}\pi R^3$, 最终我们得到:

$$\mathbf{E}_{\text{ex}} = \bar{\mathbf{E}} + \frac{\mathbf{P}}{3\epsilon_0}$$

这个细节对理解介质的极化十分重要. 我们将会在下一章重新回到这个问题.

18.4.4 若干对称带电体系的电场

最后还有一些常见的带电体系的电场是值得作为现成结论去记忆的. 它们易于计算的原因在于或多或少具有某些对称性. 在此逐一加以回忆与讨论.

1. 薄球壳: 以球心为原点, 三维旋转对称性. 过球心的镜面对称性.

$$\varphi = \begin{cases} \frac{1}{4\pi\epsilon_0} \frac{Q}{R} & (r < R) \\ \frac{1}{4\pi\epsilon_0} \frac{Q}{r} & (r > R) \end{cases}, \quad \mathbf{E} = \begin{cases} \mathbf{0} & (r < R) \\ \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \mathbf{e}_r & (r > R) \end{cases}$$

我们遇到了一个问题: 球面上的电势和电场强度是如何的? 我们这时候在定义上会遇到发散困难(divergency difficulty): 论电势, 问题不大, 从球内球外的电势数据(下面的势曲线18.17)来看, 球面上的电势连续, 但是

²¹这意味着如果球内没有电荷, 那么对球内电场做平均的结果恰好得到球心处电场. 这种性质称作调和(harmonic).

其变化率不连续(不光滑). 所以很自然地球面上电势定义为双边的极限值即球内电势 $\frac{1}{4\pi\varepsilon_0}\frac{Q}{R}$. 有限面电荷情况的电势都是不存在发散困难的. 因为在面上局部划出来一个近似均匀带电 σ , 半径为 a 小圆盘, 则盘心电势:

$$\varphi = \int_0^a \frac{1}{4\pi\varepsilon_0} \frac{\sigma \cdot 2\pi r dr}{r} = \frac{\sigma a}{2\varepsilon_0}$$

是一个有限值. 而圆盘以外的电荷对圆心的电势显然也是有限的. 故无发散困难. 反观电场强度, 这情况就很不一样了. 首当其冲地从以上公式上看, 球内球外的电场显然不连续而在球面上具有不同的极限. 其更深层次的原因在于: 面电荷的电场强度积分本质上是发散的, 有奇异性(sigularity)的. 考虑一个均匀带电 σ , 圆心角为可近似的小角 $\Delta\alpha$, 半径 a 的扇形. 圆心处的场强:

$$E = \int_0^a \frac{1}{4\pi\varepsilon_0} \frac{\sigma \cdot \Delta\alpha r dr}{r^2} = \frac{\sigma\Delta\alpha}{4\pi\varepsilon_0} \ln \frac{a}{0} = \infty$$

规避掉发散²²的思想有两种. 一是通过主值积分(principal value integral)的方式去做截断(cutoff), 在需要确定场强的点处挖去一个半径为 a 的球, 计算其他部分在该点处的场强并最后让 $a \rightarrow 0$; 二是通过平均来理解发散的场强, 如之前所说的, 点电荷模型与连续分布的电荷模型是统一的, 后者计算出来的场强就是前一种不均匀场强的球体平均. 故只需要在面电荷情况也做平均就能规避发散点的计算而利用球内其他可以计算的点来计算场强. 容易验证, 两种方法最后的结果都是法向电场为内外场强的平均数, 球面上的场强分量由于对称性为零:

$$\mathbf{E} = \frac{Q}{8\pi\varepsilon_0 R^2} \mathbf{e}_r, \quad (r = R)$$

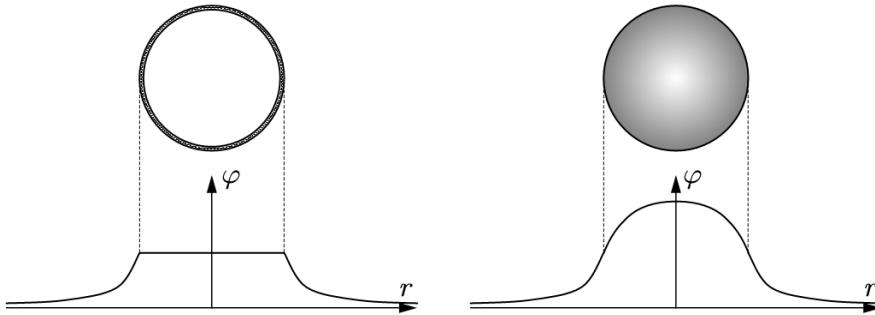


图 18.17: 薄球壳与球体的势曲线

2. 球体: 以球心为原点, 三维旋转对称性. 过球心的镜面对称性.

$$\varphi = \begin{cases} -\frac{\rho}{6\varepsilon_0} r^2 + C & (r < R) \\ \frac{1}{4\pi\varepsilon_0} \frac{Q}{r} & (r > R) \end{cases}, \quad \mathbf{E} = \begin{cases} \frac{\rho}{3\varepsilon_0} r \mathbf{e}_r & (r < R) \\ \frac{1}{4\pi\varepsilon_0} \frac{Q}{r^2} \mathbf{e}_r & (r > R) \end{cases}$$

均匀带电球体没有发散困难. 但注意在无穷远电势为零的规范下为了使得以上电势连续, 附加常数 C 应当为:

$$C = \frac{3Q}{8\pi\varepsilon_0 R}$$

3. 无限大带电平面: 保持平面位置不变的平移, 旋转对称性. 带电平面镜像对称. 竖直面镜面对称.

$$\varphi = -\frac{\sigma}{2\varepsilon_0} |z|, \quad \mathbf{E} = \frac{\sigma}{2\varepsilon_0} \text{sgn}(z) \mathbf{e}_z$$

其中 sgn 是符号函数(sign function), 即当自变量 z 取正时为 $+1$, 为负时取 -1 . $\text{sgn}(z) = z/|z|$. 从计算上看, 面上电场依然有奇异性. 面外则通过积分的方式定义的电场奇异性消失. 但是电势即使在面外此时也

²²称作正规化(regularization).

有奇异性了,不能通过电荷积分得到电势的值,只能通过与电场的关系,所以此时无法把无穷远作为势能的零点,这里的无穷远也自动被分离为下半空间无穷远和上半空间无穷远,所有涉及到无穷大带电平面的问题这样两个无穷远都必须要被区别对待.

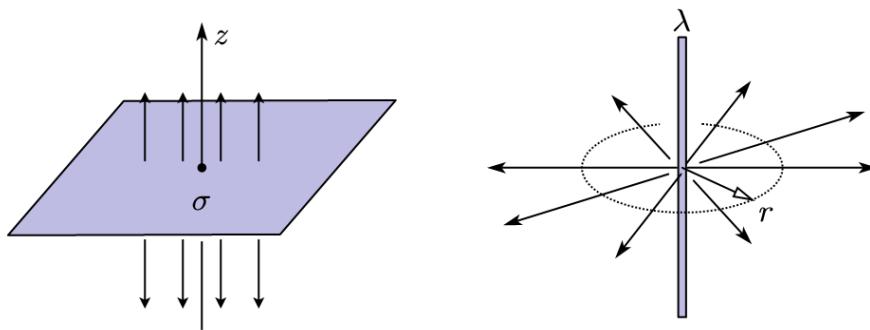


图 18.18: 无限大平面与无限长直线

4. 无限长带电直线:保持直线位置不变的平移,旋转对称性.水平面镜面对称.过直线的竖直面镜面对称.

$$\varphi = -\frac{\lambda}{2\pi\epsilon_0} \ln r, \quad \mathbf{E} = \frac{\lambda}{2\pi\epsilon_0 r} \mathbf{e}_r$$

请尤其注意电势前的负号.同样地,电场在导线外是没有奇异性的,但是电势有奇异性,电势是通过与电场关系定义的,不能把无穷远电势视作零.与上一种电荷分布之所以具有不能把无穷远视作势能零点的共同点,是因为电荷分布的范围不是有限.试想在无穷远外都有电荷了怎么可能不影响无穷远的电势分布?

如果是一个半径为 R 的,带电面密度为 σ 的圆柱薄壳,那么内部也将没有电场,等势.这一点从高斯定律很容易能看出来.以及一个均匀带电的圆柱体的电势电场分布都留给读者计算.

5. 圆环:旋转对称性.过环水平面镜面对称.过轴线的竖直面镜面对称.

这样一个体系只有轴线上的电势和电场才有比较简单的表达式:

$$\varphi = \frac{\lambda}{2\epsilon_0} \frac{1}{\sqrt{R^2 + z^2}}, \quad \mathbf{E} = \frac{\lambda}{2\epsilon_0} \frac{z}{(R^2 + z^2)^{\frac{3}{2}}} \mathbf{e}_z$$

但是实际应用场合中,环心附近的场也是非常需要的.其计算方法其实就是之前介绍的多极展开,不过现在需要保留到二阶项才能得到领头项:

$$\begin{aligned} \varphi &= \int \frac{1}{4\pi\epsilon_0} \frac{dQ}{|\mathbf{r} - \mathbf{R}|} \\ &= \int_0^{2\pi} \frac{1}{4\pi\epsilon_0} \frac{\lambda R d\theta}{\sqrt{(R \cos \theta - \rho)^2 + R^2 \sin^2 \theta + z^2}} \end{aligned}$$

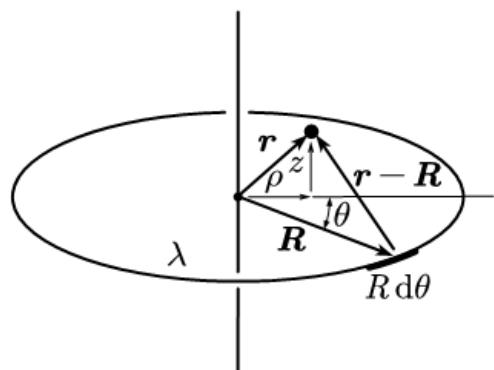


图 18.19: 圆环的中心附近电场

$$\begin{aligned}
&= \int_0^{2\pi} \frac{1}{4\pi\varepsilon_0} \frac{\lambda d\theta}{\sqrt{1 + \frac{z^2 + \rho^2 - 2R\rho \cos\theta}{R^2}}} \\
&= \frac{\lambda}{4\pi\varepsilon_0} \int_0^{2\pi} \left(1 + \frac{z^2 + \rho^2 - 2R\rho \cos\theta}{R^2}\right)^{-\frac{1}{2}} d\theta \\
&\approx \frac{\lambda}{4\pi\varepsilon_0} \int_0^{2\pi} \left[1 - \frac{1}{2} \frac{z^2 + \rho^2 - 2R\rho \cos\theta}{R^2} + \frac{3}{8} \left(\frac{z^2 + \rho^2 - 2R\rho \cos\theta}{R^2}\right)^2\right] d\theta \\
&\approx \frac{\lambda}{4\pi\varepsilon_0} \int_0^{2\pi} \left(1 + \frac{\rho \cos\theta}{R} - \frac{z^2 + \rho^2}{2R^2} + \frac{3\rho^2 \cos^2\theta}{2R^2}\right) d\theta \\
&= \frac{\lambda}{2\varepsilon_0} \left(1 - \frac{z^2}{2R^2} + \frac{\rho^2}{4R^2}\right)
\end{aligned}$$

于是原点附近的场强就是这个势的梯度:

$$E_\rho = -\frac{\partial\varphi}{\partial\rho} = -\frac{\lambda}{4\varepsilon_0 R^2} \rho, \quad E_z = -\frac{\partial\varphi}{\partial z} = -\frac{\lambda}{2\varepsilon_0 R^2} z$$

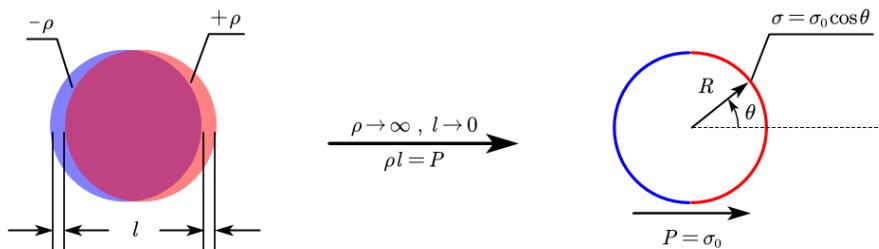


图 18.20: 均匀极化的球

6. 均匀极化球体. 旋转对称性, 过轴的平面的镜面对称性. 垂直轴的平面的镜面反对称性²³.

设球内极化强度为 P . 由之前的极化强度与电荷密度的关系知道, 这相当于一个表面带电 $\sigma = \sigma_0 \cos\theta$ 的球壳. 但是为了计算其内外的场强, 我们需要把极化强度还原为单位体积内携带的偶极子的总和. 每个偶极子可视为一对点电荷. 从而形成两个电荷密度分别为 $+\rho$ 和 $-\rho$ 的球体, 球心沿极化方向岔开 l . 只需计算极限情况:

$$\rho \rightarrow \infty, \quad l \rightarrow 0, \quad \rho l = P$$

定义球内的总电偶极矩 $\mathbf{p} = \mathbf{P} \cdot \frac{4}{3}\pi R^3$ 由球对内对外的场强公式计算得, 球内外的场强为:

$$\mathbf{E} = \begin{cases} -\frac{\mathbf{P}}{3\varepsilon_0} & (r < R) \\ \frac{1}{4\pi\varepsilon_0} \frac{2p \cos\theta}{r^3} \mathbf{e}_r + \frac{1}{4\pi\varepsilon_0} \frac{p \sin\theta}{r^3} \mathbf{e}_\theta & (r > R) \end{cases}$$

无奇异性, 且也可以得到电势公式:

$$\varphi = \begin{cases} \frac{P}{3\varepsilon_0} r \cos\theta & (r < R) \\ \frac{1}{4\pi\varepsilon_0} \frac{p \cos\theta}{r^2} & (r > R) \end{cases}$$

²³即需要镜面反射后进行正负电荷互换才能与变换前一致.

第十九章 导体与介质

19.1 导体与静电平衡

19.1.1 绝缘体与导体

微观地看,物质由原子或分子等组成,其中导电现象通常发生在不同情况下:

1. 真空导电:一般不会说真空具有导电性(conductive),因为真空中是没有载流子(charge carrier)的¹.的确,在阴极射线管(cathode ray tube)中,加热一个阴极,并辅以合适的偏置电压可以造成真空中的电流.或是用一束能量足够的光子去轰击金属表面造成电子逸出,甚至纯粹由于阴极表面尖处十分强的电场导致电子直接克服逸出功发射出来.三种现象分别称为热发射(thermionic emission),光电效应(photoelectric effect)与场发射(field emission).这种电荷的定向移动现象被统一地称为运输现象(transport phenomenon).由于真空运输的独特性质,比如电子不会受到散射,平均自由程远大于仪器尺寸,与凝聚态物理中的一些概念对应,这被称为弹道运输(ballistic transport).

2. 绝缘体漏电:大多数非金属晶体,或是不含离子的液体与气体,原子核在晶体中都限制在点阵格子的特定位置,在气体,液体中则是可以活动的原子分子周围.电子分为两类,,一类是原子的内层电子,它极为稳定地存在于原子周围,离子晶体中的几乎完全被阴离子夺取的电子也属于这种情况.它们与原子核一起构成原子实(atomic core).而价层电子(valence shell electron),它们用来成键,将原子连接形成分子或者晶体,一般被定域在原子间的特定区域.所有这些电子的特点都是束缚态(bond state).它们不能在介质中自由传导,一个电子的区域到另一个电子的区域间存在势垒(potential barrier),经典物理认为电子的动能不足以穿过这些势垒.但是,量子理论则认为电子的波函数可以通过隧道效应(tunnelling)以小概率在不同区域间转移.这就为有电场的情况下电荷的平均定向移动提供了可能.这种现象称为漏电(leakage).

3. 绝缘体击穿:在十分高的电场下,原子与原子间,或是分子中显不同电性的部分之间的典型电压将大于阻碍电子转移的势垒.在晶体被击穿时此时电子在本质上可以被视为在周期性势能场与线性势能场的叠加中运动的经典粒子.不难看出电子不仅可以脱离单个势阱的束缚,还可以在这样的电场力下做持续的加速运动.气体液体则是在电场力作用下原子,分子被撕裂为离子而继续在电场力下做加速运动.此时碰撞就发生了:碰撞的结果往往是使得更多的原子分子发生电离.这样就会雪崩式的累积,固液气中的这些现象都称作雪崩击穿(avalanche breakdown).尤其是气体中的放电现象在不同电流条件下分别造成汤森放电(Townsend discharge),电晕放电(corona discharge),辉光放电(glow discharge),电弧放电(arc discharge)的不同情况.

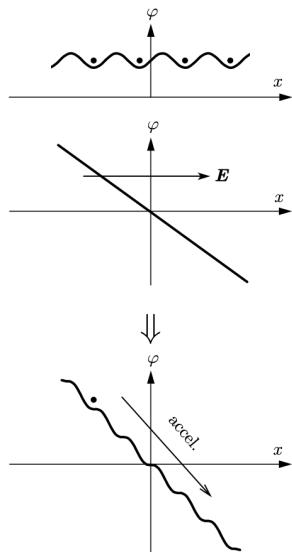


图 19.1: 晶体击穿

¹量子场论认为真空存在电子对的产生湮灭涨落从而具有一定的导电性.

4. 自然的导体: 我们将材料区分为绝缘体(insulator) 与导体(conductor), 最重要的依据就是在弱场下是否天然地具有能产生电荷输运的载流子. 所以导体就是一类事先就具有一定数密度 n 的一种或多种载流子 q 的材料, 它们往往在有电场 E 在场时发生位移, 且与散射造成的形式阻力平衡, 造成平均的匀速运动, 即定向漂移(directional drift). 常见的金属以最外层的自由电子为载流子, 半导体以热激发或掺杂形成的电子或空穴为载流子, 电解质溶液或部分气态或液态的等离子体以阴阳离子为载流子. 它们都是典型的导体. 区别于前三种情况与最后的超导体.

5. 超导体: 低温下的很多奇异的宏观现象大大拓宽了人们对量子理论在物理学各个层次中应用的认识. 其中很典型的一个就是超导(superconductivity) 现象. 最早在 1911 年由昂内斯(H. K. Onnes) 在研究低温下汞的电阻率时惊人的发现在 4.2K 下电阻率消失至零, 这就是超导的发现. 但是直到约十年后才引起足够重视. 在 1933 年迈斯纳(F. W. Meissner) 才发现超导不仅仅是电阻率的消失, 还伴随着完全抗磁性(perfect diamagnetism): 就好像理想导体内部无电场线, 在超导体中交变的磁场也只能存在于表面极小的深度内(趋肤深度), 也称迈斯纳效应(Meissner effect). 后来通过量子理论的蓬勃发展, 在 1957 年提出的BCS 理论(Bardeen-Cooper-Schrieffer theory) 终于解释了超导的成因, 原来是一对库珀电子对(Cooper pair) 通过与声子(晶格振动)相互作用而造成了没有任何散射的量子模式. 完整地解释了直到 1986 年所有低温超导(low-temperature superconductivity) 的现象. 但是随着 35K 超导的镧钡铜氧化物的发现, 宣告人类进入高温超导(high-temperature superconductivity) 纪元. 新的超导材料不断被发现, 新的理论也不断被提出, 目前最“高温”的材料是汞钡钙铜氧化物, 超导临界温度 135K. 为什么有高温超导? 高温超导的临界温度如何进一步提高? 直到今天这也还是方兴未艾的研究领域. 其在科研, 科技, 生产, 生活方面的应用是不可估量的.

我们将把研究的范围主要放在导电的导体和不导电的绝缘体上. 前者分析其静电平衡, 即本章要集中讲的情形; 下一章的稳恒电流和之后章节的拟稳的一些情况则是导体研究的其他场合. 对于绝缘体我们本章也介绍其介电特性, 光作为电磁波在介质中的传播则本章做一个抛砖引玉, 主要内容在光学色散理论中讲解.

19.1.2 导体的特点

这里说的导体, 主要指固体形式的金属.

首先界定我们研究的问题范围: 静电平衡(electrostatic equilibrium) 问题是一种稳态(steady state). 我们研究的问题由各式各样的, 有限的块状导体 $C_1, C_2 \dots C_n$ 构成, 这些导体带电情况由体电荷密度 ρ 和面电荷密度 σ 描述. 但是导体 C_i 上的总电荷量不一定为零:

$$Q_i = \int_{C_i} \rho dV + \int_{\partial C_i} \sigma dA \neq 0$$

这个电量 Q_i 一般也是可以在一定范围内人为控制的, 即给导体充电(charge). 利用电容的性质很容易可以做到这一点. 在为导体充电后由于电量守恒, 其总电量就不能改变了, 但是其电荷分布一般是未知的.

除了它们往往还有人为指定的各种点电荷 q_j 或者电荷分布 $\rho(r)$ 存在于导体外. 那么, 容易想像, 当导体外突然产生电荷分布时(如原来中和的电荷被重新分开), 或者导体被置入电荷体系中, 本质上就是当导体外的场环境发生改变时, 导体内部显然不可能保持电场强度始终为零. 只要有电场就会引起电流, 它必然使得内部和表面的电荷分布发生改变 $\rho(t), \dot{\sigma}(t) \neq 0$. 这些状态就是暂态(transient state). 但是只要外界的场不再变动, 经过特征的弛豫时间(relaxation time) 后, 一般很短只有 $10^{-14}s$ 量级, 就会接近, 直到变成稳态.

通过以上推理, 足以看出, 静电平衡下, 导体内部应当没有电场. 这就是静电平衡理论的初始命题. 它的初步推演泛化出以下的九宫格式的结论, 分别讨论导体内, 导体表面, 导体外三处的电荷分布, 电场分布, 和电势分布三个特性:

一般特征	电荷	电场	电势
导体内	$\rho = 0$	$E = 0$	$\varphi = C$
导体表面	$\sigma \neq 0$	奇异	$\varphi = C$
导体外	$\rho \neq 0$	$E \neq 0$	$\varphi \neq C$

表 19.1: 静电平衡的初级结论

其实也就是说,因为导体内部没有电场,所以导体内部也没有形成电场的电荷(高斯定理可证),也使得导体成为一个等势体(equipotential body),导体表面成为一个等势面(equipotential surface).注意,这里说的“内部”,“表面”,“外部”包含两种情况.导体内部就是金属成分的空间,外部一般是真空与外加电荷分布的空间,表面是它们的分界面,但是下图左的情况,导体的外部延伸到了无穷远,这种为开外界(open exterior),但是中和右两种情况,导体的外部是有限的闭外界(closed exterior).第二种典型情况这个单连通的闭外界还被称作空腔(cavity).除非特殊说明,我们一般默认导体的内部不会延伸到无穷远(总是闭的).

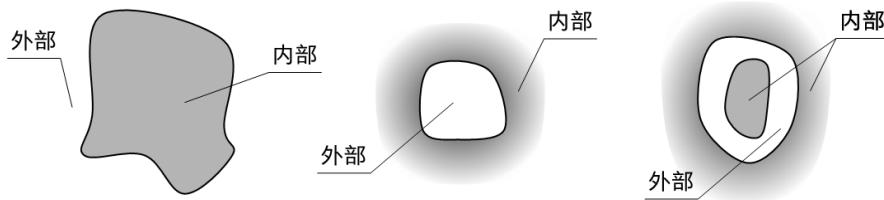


图 19.2: 导体内外

在导体表面发生了电场计算的奇异性,因为这里是面电荷分布.我们考察细节就会进一步发现以下进一步的结论: **导体表面的面电荷只朝外侧发出(或吸收)电场线,其方向垂直于表面(电场线垂直于等势面),且疏密程度正比于电荷密度(即高斯定理)**.即:

$$\mathbf{E} = -\nabla\varphi = \frac{\sigma}{\epsilon_0}\mathbf{n}$$

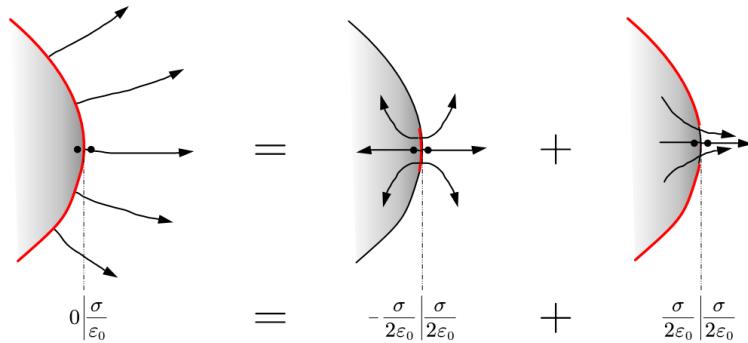


图 19.3: 导体表面

但是正如上图所示,如果在十分接近表面的内外各取一点,就会发现如果只选取相对这两点足够大的一块面电荷(但又足够小以至于可以视作平面),它对面左右的电场强度只贡献以上值的一半.那么另一半必然由去掉这一块电荷的另外的部分所贡献.所以如果要讨论面上的场强,通过上一节的球体平均的理解方

式, 它就自然地被定义为内外场强和的一半, 即:

$$\mathbf{E} = \frac{\sigma}{2\epsilon_0} \mathbf{n}$$

对于导体静电平衡问题的计算求解, 也包括模拟求解, 本质上都建立在以下基本原理的基础上:

独立叠加原理: 静电平衡问题可以拆解为独立的在导体外部一个个连通区域内的静电边值问题.

所谓的**边值问题**(boundary value problem) 本是偏微分方程的术语, 在这里特指下面说的静电情况. 每一种情况中, 我们无需分析导体的内部的电荷, 电场, 电势. 只需要求解外部的电场 $\mathbf{E}(\mathbf{r})$ 和电势 $\varphi(\mathbf{r})$. 导体外部的电荷分布 $\rho(\mathbf{r})$ 却是事先给定的. 还给定了每一种导体表面的, 要么是电势 V_i (从而导体内部的电势也是这个值), 把这种导体称作第一类导体, 对应条件称为第一类边界条件; 要么是给定了这个面上的总电量 Q_i , 把这种导体称作第二类导体, 对应条件称为第二类边界条件. 而每一个导体外部的区域的边界, 就回到了各个导体的表面 $\Sigma_1, \Sigma_2 \dots \Sigma_n, \Pi_1, \Pi_2 \dots \Pi_m$ 构成的集合. 但是表面上的电荷分布也是事先不知道的. 由于电荷, 电场都可以从电势中得到, 故只暂时考虑求电势, 我们把静电边值问题整理一下:

已知: 导体外连通区域 D 以第一类导体表面 $\Sigma_1, \Sigma_2 \dots \Sigma_n$ 和第二类导体表面 $\Pi_1, \Pi_2 \dots \Pi_m$ 为边界. 已知区域 D 内电荷分布 ρ , 已知第一类导体表面 Σ_i 上的电势 V_i . 已知第二类导体表面 Π_j 上的电量 Q_j :

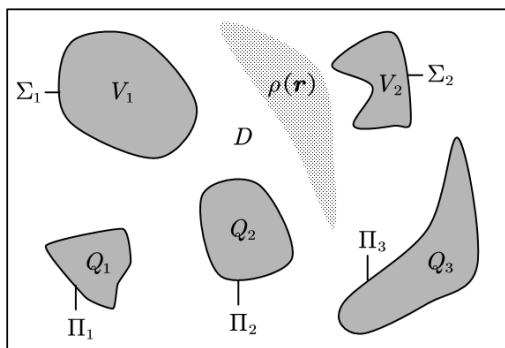
$$\partial D = \left(\bigcup_i \Sigma_i \right) \cup \left(\bigcup_j \Pi_j \right)$$

$$\rho = \rho(\mathbf{r}) \quad , \quad \mathbf{r} \in D$$

$$\varphi(\mathbf{r})|_{\Sigma_i} = V_i \quad , \quad \int_{\Pi_j} \nabla \varphi \cdot d\mathbf{A} = Q_j / \epsilon_0 \quad , \quad \varphi(\mathbf{r})|_{\Pi_j} = C$$

未知: 导体表面电荷的具体分布 $\sigma(\mathbf{r})|_{\Sigma_i, \Pi_j}$

待求: D 内的电势分布 $\varphi(\mathbf{r})|_D$



但是还要注意这种边值问题又分为**外问题**(outside problem) 和**内问题**(inside problem) 两类. 外问题是区域 D 延伸到了无穷远, 这种情况的边值问题提法就完全如上所述. 但是内问题中区域 D 存在一个最外的表面 Σ_0 , 它把整个 D 和其他所有金属表面与内部包含在其内部. Σ_0 的外部亦是金属. 对于内问题, 我们很容易发现 Σ_0 上的总带电量为内部所有带电量的和的相反数, 且仅仅由内部体系和它产生的电场线不会延伸到 Σ_0 外部. 所以我们暂时可以以这一类问题中 Σ_0 上的电势为零点:

$$\varphi(\mathbf{r})|_{\Sigma_0} = 0$$

这样所有已知未知的电势皆是相对 Σ_0 而言的电势差.

对于每一个具体的边值问题, 指导我们求解的重要思想是**广义叠加原理**(general superposition principle) 和**唯一性定理**(uniqueness theorem). 从语义上看广义叠加原理显然与之前的单纯的叠加原理有着直接的联系. 其实它更是来自于数学上具有线性结构的系统的解的叠加原理. 这个叠加原理是这样表述的:

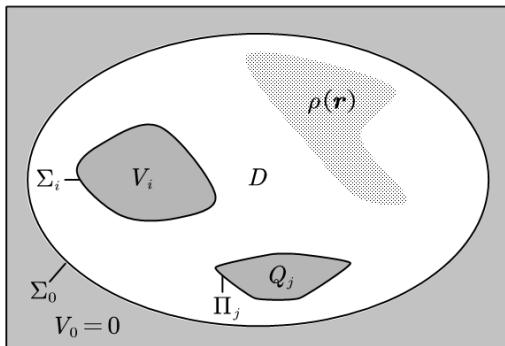


图 19.4: 边值问题

广义叠加原理: 两个边值问题如果 D, Σ_i, Π_j 一致. 第一个问题具有第一类边界条件 V_{1i} 和第二类边界条件 Q_{1j} 和内部电荷分布 $\rho_1(\mathbf{r})|_D$. 第二个问题具有第一类边界条件 V_{2i} 和第二类边界条件 Q_{2j} 和内部电荷分布 $\rho_2(\mathbf{r})|_D$. 两个问题的解分别为 $\varphi_1(\mathbf{r})|_D, \varphi_2(\mathbf{r})|_D$. 则同样以 D 为区域, Σ_i, Π_j 为 D 两类导体边界, 但是以 $V_{1i} + V_{2i}$ 和 $Q_{1j} + Q_{2j}$ 为两类边界条件的问题的解一定是:

$$\varphi(\mathbf{r}) = \varphi_1(\mathbf{r}) + \varphi_2(\mathbf{r})$$

这个定理的证明是显然的: 首先 φ 作为一个边值问题的解, 当且仅当它满足以下的一个全局条件和若干边值条件:

$$\nabla^2 \varphi = -\frac{\rho}{\varepsilon_0}$$

$$\varphi(\mathbf{r})|_{\Sigma_i} = V_i \quad , \quad \int_{\Pi_j} \nabla \varphi \cdot d\mathbf{A} = Q_j \quad , \quad \varphi(\mathbf{r})|_{\Pi_j} = C$$

那么既然 φ_1 和 φ_2 都是之前两个问题的解, 这就意味着:

$$\nabla^2 \varphi_1 = -\frac{\rho_1}{\varepsilon_0} \quad , \quad \nabla^2 \varphi_2 = -\frac{\rho_2}{\varepsilon_0}$$

$$\varphi_1(\mathbf{r})|_{\Sigma_i} = V_{1i} \quad , \quad \int_{\Pi_j} \nabla \varphi_1 \cdot d\mathbf{A} = Q_{1j} \quad , \quad \varphi_1(\mathbf{r})|_{\Pi_j} = C$$

$$\varphi_2(\mathbf{r})|_{\Sigma_i} = V_{2i} \quad , \quad \int_{\Pi_j} \nabla \varphi_2 \cdot d\mathbf{A} = Q_{2j} \quad , \quad \varphi_2(\mathbf{r})|_{\Pi_j} = C$$

那么定义 $\varphi = \varphi_1 + \varphi_2$, 它就自然也能满足:

$$\nabla^2 \varphi = -\frac{\rho_1 + \rho_2}{\varepsilon_0}$$

$$\varphi(\mathbf{r})|_{\Sigma_i} = V_{1i} + V_{2i} \quad , \quad \int_{\Pi_j} \nabla \varphi \cdot d\mathbf{A} = Q_{1j} + Q_{2j} \quad , \quad \varphi(\mathbf{r})|_{\Pi_j} = C$$

从而就的确是新问题的一个解. 但是这个解是否唯一? 答案是肯定的. 这就是著名的唯一性定理, 可以这么表述:

唯一性定理: 如果同一个边值问题具有两个解 φ_1, φ_2 . 那么 $\varphi_1 - \varphi_2 = 0$.

其证明的核心思想是利用上一章介绍的第一格林等式². 首先注意到如果 φ_1, φ_2 都是同一个边值问题的解. 那么命 $\varphi = \varphi_1 - \varphi_2$ 就是以下边值问题的解:

$$\nabla^2 \varphi = -\frac{\rho - \rho}{\varepsilon_0} = 0$$

$$\varphi(\mathbf{r})|_{\Sigma_i} = 0 \quad , \quad \int_{\Pi_j} \nabla \varphi \cdot d\mathbf{A} = 0 \quad , \quad \varphi(\mathbf{r})|_{\Pi_j} = C$$

上面这个等号右侧都是零的方程称作齐次方程(homogeneous equation). 那么问题就被转化为证明齐次方程的解必为零解. 想起第一格林等式:

$$\nabla \cdot (\phi \nabla \psi) = \phi \nabla^2 \psi + \nabla \phi \cdot \nabla \psi$$

同样是在 ϕ, ψ 都取 φ 的情况:

$$\varphi \nabla^2 \varphi = \nabla \cdot (\varphi \nabla \varphi) - (\nabla \varphi)^2$$

²其实使用简单的“电荷发出电场线, 沿电场线方向电势降低”的观点也是可以论证这个定理的. 请读者思考过程.

现在把 φ 就视作刚刚定义的 $\varphi_1 - \varphi_2$, 并把这个公式运用于 D 中的任意一点, 由于电势满足的条件 $\nabla^2\varphi = 0$, 就有:

$$(\nabla\varphi)^2 = \nabla \cdot (\varphi\nabla\varphi)$$

等号右侧会让人联想到奥-高定理. 事实上如果把它应用到 D : 左右两侧就能化为:

$$\int_D (\nabla\varphi)^2 dV = \int_D \nabla \cdot (\varphi\nabla\varphi) dV = \oint_{\partial D} \varphi\nabla\varphi \cdot dA = \oint_{(\cup_i \Sigma_i) \cup (\cup_j \Pi_j)} \dots = \sum_i \oint_{\Sigma_i} \varphi\nabla\varphi \cdot dA + \sum_j \oint_{\Pi_j} \varphi\nabla\varphi \cdot dA$$

在第一类导体边界上, 由于 $\varphi = 0$ 所以积分为零. 在第二类导体边界上, 首先 φ 也是常数所以可以提出到积分外面, 即:

$$\oint_{\Pi_j} \varphi\nabla\varphi \cdot dA = C \cdot \oint_{\Pi_j} \nabla\varphi \cdot dA = C \cdot 0 = 0$$

从而得到恰好等号的右侧就是 0. 这就相当于证明了这个电势 φ 满足:

$$\int_D (\nabla\varphi)^2 dV = 0$$

一个总是不小于零的 $(\nabla\varphi)^2$, 经过积分得到了零的结果. 这就说明它恒等于零:

$$\nabla\varphi = \mathbf{0}$$

从而电势就是一个常数, 由于在第一类边界条件下得取零, 故只能是:

$$\varphi = 0$$

以上两个原理是静电平衡问题能够被分析与解决的基础. 以下几类问题中, 区域 D 有高度对称性, 且内部都没有电荷分布 ρ , 那么其解决甚至用不上这两个原理, 下一节的电像法就非常依赖这两个原理了. 不过这里的几类情况足以解决很多实际的问题:

19.1.3 常见简单体系

1. 平行的无限大导体板

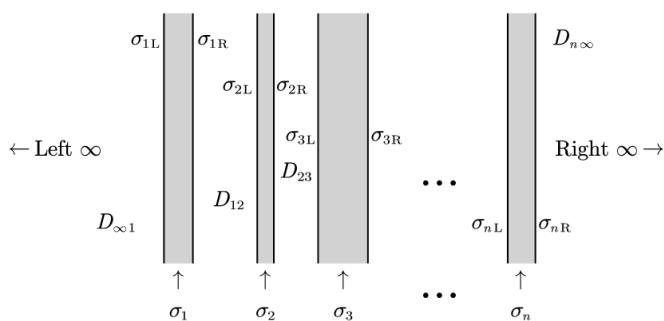


图 19.5: 平行板问题

我们先考虑无限大的导体平面板, 它可能具有一定的厚度. 把这样的 n 块板子从左至右平行放置, 如图19.5. 这样就把空间隔离为中间的 $D_{12}, D_{23}, \dots, D_{n-1,n}$ 等区域.

我们没有必要严格写出这个问题的解. 把这个体系的解需要满足的性质罗列出来会更实用. 首当其冲地我们写出每一块板上的电量需要满足的关系. 如果板非常薄, 那么左右两侧的带电面密度 σ_{iL} 和右侧 σ_{iR} 就可以合成为统一的 σ_i . 那么以下两个特征是显著的:

$$1. \sigma_{iR} + \sigma_{i+1L} = 0$$

$$2. \sigma_{1L} = \sigma_{nR} = (\sigma_1 + \dots + \sigma_n)/2$$

而在第 i 与第 $i+1$ 块板之间的区域 $D_{i,i+1}$ 内, 如果研究场强, 以向右为正. 它应该满足两个关系:

$$3. E_{i,i+1} = \sigma_{iR}/\epsilon_0 = -\sigma_{i+1L}/\epsilon_0$$

$$4. E_{i,i+1} = (\sigma_1 + \dots + \sigma_i - \sigma_{i+1} - \dots - \sigma_n)/2\epsilon_0$$

请注意, 在以上关系式中, 1, 3 关系式仅仅依赖于上一节我们推出过的边界条件. 是无条件成立的. 但是 2, 4 却无法从边界条件中推导出来. 所以这两个式子的成立是有条件的: 那就是整个体系“仅仅”由这些带电的板组成. 事实上, 如果把第 1 块板和第 n 向左向右平移到无限远, 可以想象不改变板上的电荷分布也能保持静电平衡. 但是这样中间就只剩下 2, 3, …, $n-1$ 这些板, 它们就可以造成典型的仅仅符合 1, 3 而不符合 2, 4 的带电体系. 所以 2, 4 的正确性要求“无穷远处不能有电荷分布”. 此时我们可以通过叠加原理来解释 4 的正确性. 又反过来通过 4 计算出来的电场强度用 3 来计算各板左右两侧的带电量. 最后就能推导出 2 式来.

对于等式 2, 我们发现这实际上代表一种左右对称性: 在这些所有板的外侧区域 $D_{\infty 1}, D_{n \infty}$ 观察, 这些所有板等效于单块 $\sigma_1 + \dots + \sigma_n$ 的板. 从而左右的场强都是:

$$E_{\infty} = \frac{\sigma_1 + \dots + \sigma_n}{2\epsilon_0}$$

如果把上面的模型替换为电容器电路, 各个区域 D 变成电容内部的场强区, 而每一块导体板内部的区域被改成一根导线, 并可以通过外界引入电量. 那么需要注意, 这个体系额外多了一个性质: 一般情况下各个电容器外侧的电量都一定是零. 这是因为电容器初始状态必然不带电, 而充放电过程是拟稳³的, 故两极板积累的电量必然相反. 故按照上面的公式计算的 E_{∞} 就是零. 所以电容器极板外侧一般不允许有场强. 除非遇到电容器中间被放置了外电荷的情况, 比如在中间引入电压可以调控的栅极, 对于这种“异型”电容的问题处理方法就又回到了之前我们说的四个结论上.

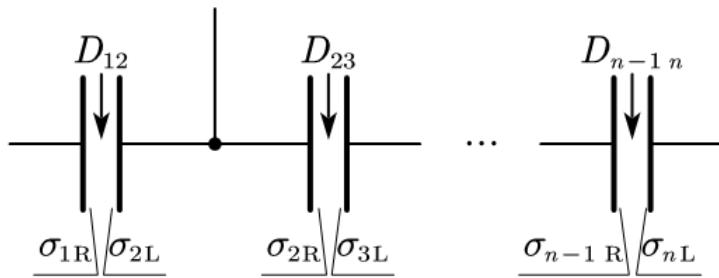


图 19.6: 电容器

对于一个面积为 S , 极板间距为 l 的平板电容器(parallel plate capacitor), 正常工作的时候根据之前的结论, 两板正对面带相反的电荷, 电场线起始于带正电的板而终止于带负电的板. 如果忽略其边缘效应, 即 $S \gg l$ 情况下, 我们可以认为其去掉边缘以后的正对面积与 S 的相对误差非常小, 电容就被定义为两板带正电或负电的绝对值与两板电势差的比值, 它等于:

$$C = \frac{Q}{U} = \frac{\epsilon_0 S}{l}$$

当平板电容器带电时, 其产生的能量可以通过静电能公式或计算外界做功, 也可以用电场能量计算得到:

$$E = \frac{1}{2}CU^2 = \frac{1}{2}QU = \frac{1}{2}\frac{Q^2}{C}$$

这个公式是普遍的, 对任何符合上一个式子定义的电容都是成立的.

2. 同心导体球与球壳

导体球与球壳反倒简单一些, 它们面上的总带电量不存在发散的情况, 而且也不需要考虑无穷远处的边界条件. 用各内表面总带电量 Q_i 和外表面带电量 Q_o 描述, 之前的四个条件现在也需要改写为:

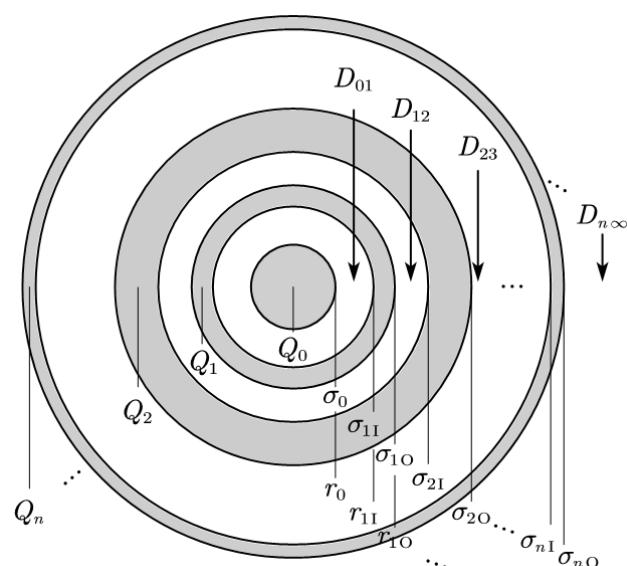


图 19.7: 球壳或柱壳

³即符合节点电流方程等, 见之后章节的说明.

1. $Q_{iO} + Q_{i+1I} = 0$
2. $Q_{nO} = Q_1 + \dots + Q_n$
3. $E_{i+1} = \frac{Q_{iO}}{4\pi\varepsilon_0 r^2} = -\frac{Q_{i+1I}}{4\pi\varepsilon_0 r^2}$
4. $E_{i+1} = \frac{Q_1 + \dots + Q_i}{4\pi\varepsilon_0 r^2}$

两个球壳构成的电容器,如果内外球壳半径分别为 r_I 和 r_O ,那么电容为:

$$C = 4\pi\varepsilon_0 \frac{r_O r_I}{r_O - r_I}$$

3. 共轴的导体圆柱与圆柱壳

我们先考虑无限长的圆柱.那么依然参考上图19.7.圆柱带电应用单位长度的电量 λ 来描述,四个关系式变为:

1. $\lambda_{iO} + \lambda_{i+1I} = 0$
2. $\lambda_{nO} = \lambda_1 + \dots + \lambda_n$
3. $E_{i+1} = \frac{\lambda_{iO}}{2\pi\varepsilon_0 r} = -\frac{\lambda_{i+1I}}{2\pi\varepsilon_0 r}$
4. $E_{i+1} = \frac{\lambda_1 + \dots + \lambda_i}{2\pi\varepsilon_0 r}$

两个柱壳构成的电容器,如果内外球壳半径分别为 r_I 和 r_O ,其长度为 l 且忽略边缘效应,即 $l \gg r_O > r_I$ 那么电容为:

$$C = 2\pi\varepsilon_0 l \ln \frac{r_O}{r_I}$$

19.2 电像法

下面考虑两种实际的边值问题的解.第一种问题中区域 D 为半无限大空间,即,其边界为一个无限大导体的表面.第二种问题中区域 D 要么是一个球面外的延伸到无穷远的区域(外问题),要么是球面内的区域(内问题),不管哪种,区域边界都是导体的球面表面,只是前一种是外表面,后一种是内表面.

我们提供的方法为电像法.其意义是,先考虑一个较简单的,在空间中 \mathbf{R} 处放置一个点电荷 q :

$$\rho(\mathbf{r}) = q\delta(\mathbf{r} - \mathbf{R})$$

的情况下,边值问题的求解.其中 $\delta(\mathbf{r} - \mathbf{R})$ 为**狄拉克德尔塔函数**(Dirac delta function),读者大可以把它视作某物理量的某种极限化的密度分布的抽象⁴,例如物理量总值为 1,分布在 $\mathbf{R} = \mathbf{0}$ 附近的一个非常小的范围内,那么其密度就是:

$$\delta(\mathbf{r}) = \begin{cases} +\infty & (\mathbf{r} = \mathbf{0}) \\ 0 & (\mathbf{r} \neq \mathbf{0}) \end{cases} \quad \& \quad \int_{\text{Ball}(\mathbf{0}, \epsilon)} \delta(\mathbf{r}) dV = 1$$

那么此前的密度函数 $\rho(\mathbf{r}) = q\delta(\mathbf{r} - \mathbf{R})$ 就可以表示点电荷的电荷密度分布.

⁴ 实则这样的广义函数数学上称作一种**分布**(distribution),它是通过积分形式的运算定义的:

$$f(\mathbf{r}) \circ g(\mathbf{r}) := \int f(\mathbf{r})g(\mathbf{r}) dV$$

$$\delta(\mathbf{r} - \mathbf{R}) \circ g(\mathbf{r}) := g(\mathbf{R})$$

19.2.1 半无限大空间的电像法

下面考虑在无限大导体平板处建立坐标系, $z > 0$ 就是待研究的 D 区域. 现在在 \mathbf{R} 处放置一个正点电荷 q , 容易证明, 由于导体板延伸到无穷远, 所以导体板的表面自然成为第一类边界条件的表面且其电势为零 $V = 0$. 表面上就会产生一个面电荷分布 $\sigma(x, y)$.

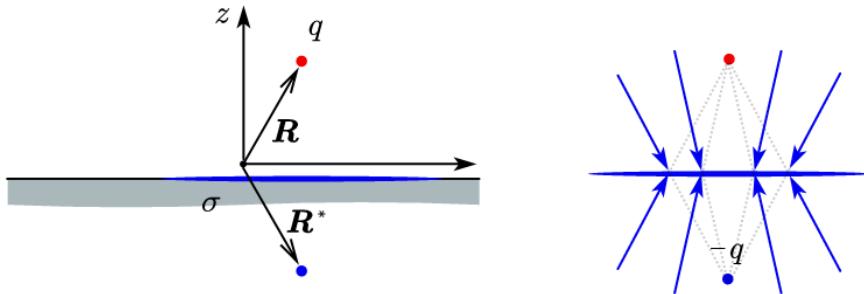


图 19.8: 半无限空间的电像

见上图右, 单独考虑这个面电荷产生的电场线, 由于在 $z < 0$ 区域内为导体, 合电场强度必然为零. 故这个区域面电荷的电场线要与原来的 \mathbf{R} 处的 q 的电场线相抵消. 故为指向 \mathbf{R} 处的汇聚的电场线. 但是, 由于这个面电荷产生的电场的上下的镜面对称性, 这就导致我们可以直接证明它对 $z > 0$ 的 D 区域产生的电场线, 实际上就相当于在 \mathbf{R} 的镜面对称点 \mathbf{R}^* 处放置一个像电荷(image charge), 值为 $-q$ 而产生的电场.

从而 $z > 0$ 区域的电场, 实际上就是由在 \mathbf{R} 的原电荷和在 \mathbf{R}^* 的像电荷共同产生的电场. 或者说, 满足以下方程的唯一解:

$$\nabla^2 \varphi = -\frac{q\delta(\mathbf{r} - \mathbf{R})}{\varepsilon_0} (z > 0) , \quad \varphi|_{z=0} = 0$$

就是:

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \frac{q}{|\mathbf{r} - \mathbf{R}|} - \frac{1}{4\pi\varepsilon_0} \frac{q}{|\mathbf{r} - \mathbf{R}^*|}$$

那么我们如何求解下列更加普遍的第一类边值问题⁵呢?

$$\nabla^2 \varphi = -\frac{\rho(\mathbf{r})}{\varepsilon_0} (z > 0) , \quad \varphi|_{z=0} = V$$

这个解答是简单的: 就是对以上点电荷电像法的情形做叠加原理. 我们把以上问题中, 如果点电荷为单位电荷时, 对应的解称作参数为 \mathbf{R} 处的格林函数(Green function):

$$\delta(\mathbf{r} - \mathbf{R}) \longrightarrow \varphi(\mathbf{r}) = G(\mathbf{R}, \mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \left(\frac{1}{|\mathbf{r} - \mathbf{R}|} - \frac{1}{|\mathbf{r} - \mathbf{R}^*|} \right)$$

由于任意的电荷分布可以看成是点电荷的叠加, 下式 dV 为 \mathbf{R} 改变对应的体积元:

$$\rho(\mathbf{r}) = \int \rho(\mathbf{R}) \delta(\mathbf{r} - \mathbf{R}) dV$$

那么这个电荷分布产生的电势, 就是各个点电荷的格林函数的叠加:

$$\rho(\mathbf{r}) \longrightarrow \varphi(\mathbf{r}) = \int \rho(\mathbf{R}) G(\mathbf{R}, \mathbf{r}) dV$$

这样在导体面上产生的电势依然是零, 为了符合边界条件, 只需要整体加上常数 V , 最终结果为:

$$\varphi(\mathbf{r}) = V + \int \frac{1}{4\pi\varepsilon_0} \left(\frac{1}{|\mathbf{r} - \mathbf{R}|} - \frac{1}{|\mathbf{r} - \mathbf{R}^*|} \right) \rho(\mathbf{R}) dV$$

⁵ 第二类边值问题此时会有问题: 首先导体面上的总电荷量就是上方的自由电荷量总值的相反数而不可以任意给定, 而最终电势的结果也会产生一个未定常数的不确定性.

19.2.2 球面外与球面内的电像法

有了半无限区域的电像法求解边值问题做基础, 我们对球面内外区域的边值问题的求解就有了线索. 同样的, 在球心处建立原点和坐标系, 求解这两个问题的关键在于先求解以下两个最简单的第一类边值问题:

$$\nabla^2 \varphi = -\frac{q\delta(\mathbf{r} - \mathbf{R})}{\varepsilon_0} \quad (|\mathbf{r}| > a) \quad , \quad \varphi|_{|\mathbf{r}|=a} = 0$$

$$\nabla^2 \varphi = -\frac{q\delta(\mathbf{r} - \mathbf{R})}{\varepsilon_0} \quad (|\mathbf{r}| < a) \quad , \quad \varphi|_{|\mathbf{r}|=a} = 0$$

分别对应外问题和内问题. 即下图左和右:

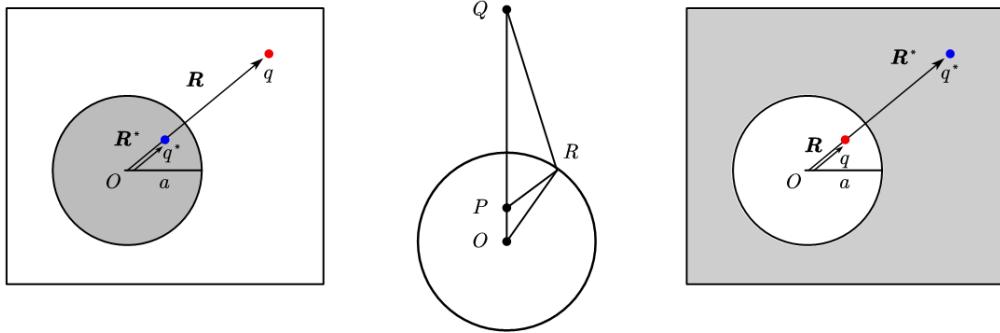


图 19.9: 球面的电像

而以上两个问题的电像解, 其实直接就都是:

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \frac{q}{|\mathbf{r} - \mathbf{R}|} + \frac{1}{4\pi\varepsilon_0} \frac{q^*}{|\mathbf{r} - \mathbf{R}^*|}$$

其中像电荷与原电荷在同一条半径上, 且其大小 q^* 和到原点的距离 R^* 与圆的半径 a 分别满足:

$$RR^* = a^2$$

$$-\frac{q^*}{q} = \frac{a}{R} = \frac{R^*}{a}$$

证明这两个解的正确性的关键在于原电荷与像电荷的位置对圆是构成了一对反演点的关系. 见上图中间的几何关系, 不难证明:

$$\overline{OP} \cdot \overline{OQ} = a^2 = \overline{OR} \Rightarrow \triangle OPR \sim \triangle ORQ$$

于是根据相似关系不难得到:

$$\frac{\overline{OP}^{1/2}}{\overline{PR}} = \frac{\overline{OQ}^{1/2}}{\overline{QR}}$$

所以只要让两个电荷分别位于 P, Q 处, 且电荷量反号且大小之比为 $\overline{OP}^{1/2} : \overline{OQ}^{1/2}$, 那么它们在 R 处的电势和就是零. 实际上上面的电势解就是取了这样的两个电荷的电势的和, 故就能满足原来的方程的边界条件, 而这样的电势恰好在给定区域 D 内也能满足对应的微分方程: 即 $\nabla^2 \varphi$ 只得到 R 处电荷 q .

这就证明了以上写出来的解是正确的. 根据唯一性定理, 它也是唯一的.

下面考虑在普遍的第一类和第二类边值问题下外问题的解. 先考虑第一类边值问题, 即:

$$\nabla^2 \varphi = -\frac{\rho(\mathbf{r})}{\varepsilon_0} \quad (|\mathbf{r}| > a) \quad , \quad \varphi|_{|\mathbf{r}|=a} = V$$

同样的利用以上解写出格林函数:

$$G(\mathbf{R}, \mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left(\frac{1}{|\mathbf{r} - \mathbf{R}|} - \frac{a/|\mathbf{R}|}{|\mathbf{r} - \mathbf{R}^*|} \right)$$

于是只需要写出下列积分, 就可以满足问题前一半的拉普拉斯方程:

$$\int \rho(\mathbf{R}) G(\mathbf{R}, \mathbf{r}) dV$$

这样一个势在边界上必然取零的值. 为了使得最后的解满足边界条件, 自然去思考叠加原理: 让外界 $\rho = 0$ 且边界上具有电势 V 的电势显然为:

$$\varphi = \frac{VR}{r}$$

故, 最后的解就是:

$$\varphi = \frac{VR}{r} + \int \frac{1}{4\pi\epsilon_0} \left(\frac{1}{|\mathbf{r} - \mathbf{R}|} - \frac{a/|\mathbf{R}|}{|\mathbf{r} - \mathbf{R}^*|} \right) \rho(\mathbf{R}) dV$$

但是如果换为第二类边值问题:

$$\nabla^2 \varphi = -\frac{\rho(\mathbf{r})}{\epsilon_0} (|\mathbf{r}| > a) \quad , \quad \int_{|\mathbf{r}|=a} \nabla \varphi \cdot d\mathbf{A} = Q, \quad \varphi|_{|\mathbf{r}|=a} = C$$

这就需要先考虑格林函数对应的势在边界上产生的总面电荷量, 他其实就是像电荷的值:

$$\delta(\mathbf{r} - \mathbf{R}) \longrightarrow \int_{|\mathbf{r}|=a} \nabla \varphi \cdot d\mathbf{A} = -\frac{a}{|\mathbf{R}|}$$

于是单独对格林函数进行积分得到的电势就使得面上带上电荷:

$$\rho(\mathbf{r}) \longrightarrow \int_{|\mathbf{r}|=a} \nabla \varphi \cdot d\mathbf{A} = - \int \frac{a}{|\mathbf{R}|} \rho(\mathbf{R}) dV$$

那么通过叠加原理, 为了使得面上电荷量变为 Q , 就需要引入一个电势:

$$\varphi = \frac{1}{4\pi\epsilon_0} \frac{Q + \int \frac{a}{|\mathbf{R}|} \rho(\mathbf{R}) dV}{r}$$

综上, 最后的解为:

$$\varphi = \frac{1}{4\pi\epsilon_0} \frac{Q + \int \frac{a}{|\mathbf{R}|} \rho(\mathbf{R}) dV}{r} + \int \frac{1}{4\pi\epsilon_0} \left(\frac{1}{|\mathbf{r} - \mathbf{R}|} - \frac{a/|\mathbf{R}|}{|\mathbf{r} - \mathbf{R}^*|} \right) \rho(\mathbf{R}) dV$$

球面外问题中, 所有电势的结果都是 $\rho(\mathbf{r})$ 和球面上的面电荷分布通过叠加原理产生的电势.

最后考虑内问题. 我们介绍过, 此时只有唯一的一种情况: 即内表面取为电势为零的第一类边界条件. 其问题表述和解必然为:

$$\nabla^2 \varphi = -\frac{\rho(\mathbf{r})}{\epsilon_0} (|\mathbf{r}| < a) \quad , \quad \varphi|_{|\mathbf{r}|=a} = 0$$

$$\varphi = \int \frac{1}{4\pi\epsilon_0} \left(\frac{1}{|\mathbf{r} - \mathbf{R}|} - \frac{a/|\mathbf{R}|}{|\mathbf{r} - \mathbf{R}^*|} \right) \rho(\mathbf{R}) dV$$

这个电势就是内部的电荷 $\rho(\mathbf{r})$ 和面上的感应面电荷共同对球内叠加产生的电势的和. 对于球外的电势的叠加必然为零.

19.2.3 立体谐函数 *

将一个中心在原点 $\mathbf{r} = \mathbf{0}$, 半径为 a 的球置于外场中讨论静电平衡为上一节中外问题的设置. 但是上一节选择了在距离球心 R 处放置了一个点电荷. 从而得到了球面在点电荷产生的外场下形成的格林函数 $G(\mathbf{R}, \mathbf{r})$. 实用问题中通常把外场按照**立体谐函数**(solid harmonics) 在原点进行分解, 类似于泰勒展开. 只不过立体谐函数更适合于球坐标:

$$R_{lm}(\mathbf{r}) = \sqrt{\frac{(l-m)!}{(l+m)!}} r^l P_{lm}(\cos \theta) e^{im\varphi}$$

$$\begin{aligned} P_{lm}(x) &= (-)^m (1-x^2)^{m/2} \frac{d^m P_l(x)}{dx^m} \\ P_l(x) &= \frac{(-)^l}{2^l l!} \frac{d^l}{dx^l} (1-x^2)^l = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l \end{aligned}$$

这样的复杂定义其实非常自然地缘起于球坐标下拉普拉斯方程的解. 除了以上**正规**(regular) 部分, 还有一组对应的在原点发散的**非正规**(irregular) 解:

$$I_{lm}(\mathbf{r}) = \sqrt{\frac{(l-m)!}{(l+m)!}} r^{-(l+1)} P_{lm}(\cos \theta) e^{im\varphi}$$

非正规解作为电势代表的场由放在原点的电荷, 电偶极子乃至更高阶的对象产生. 因为我们在原点放置了一个导体球, 所以它受到的外场不可能含有非正规成分.

以上完备谐函数组非常复杂, 实际我们考虑的问题通常对球坐标 φ 的改变是对称的, 例如球外点电荷对原点附近的外场, 若以原点和点电荷的连线方向为极轴定义 θ, φ , 那么这个外场就不含任何立体谐函数中 $m \neq 0$ 的成分. 此时仅考虑 $m = 0$ 的那些函数是完备的:

$$R_l(\mathbf{r}) = \frac{1}{2^l l!} \cdot r^l \cdot \left(\frac{d}{\sin \theta d\theta} \right)^l \sin^{2l} \theta$$

$$I_l(\mathbf{r}) = \frac{1}{2^l l!} \cdot \frac{1}{r^{l+1}} \cdot \left(\frac{d}{\sin \theta d\theta} \right)^l \sin^{2l} \theta$$

对于较小的 l 我们罗列若干对称立体谐函数:

l	正规	非正规
0	1	$\frac{1}{r}$
1	$r \cos \theta$	$\frac{\cos \theta}{r^2}$
2	$\frac{r^2}{2}(3 \cos^2 \theta - 1)$	$\frac{1}{2r^3}(3 \cos^2 \theta - 1)$
3	$\frac{r^3}{2}(5 \cos^3 \theta - 3 \cos \theta)$	$\frac{1}{2r^4}(5 \cos^3 \theta - 3 \cos \theta)$
4	$\frac{r^4}{8}(35 \cos^4 \theta - 30 \cos^2 \theta + 3)$	$\frac{1}{8r^5}(35 \cos^4 \theta - 30 \cos^2 \theta + 3)$
\vdots	\vdots	\vdots

表 19.2: 对称立体谐函数表

值得注意的是,正规的立体谐函数总是可以写成多项式的形式.如果以 $\theta = 0$ 的方向为 z 轴,那么 $r \cos \theta = z$, $r^2 = x^2 + y^2 + z^2$.这样就得到了对称情况下的谐和多项式(harmonic polynomials),例如:

$$R_0(\mathbf{r}) = \mathcal{P}_0(x, y, z) = 1 \quad (19.1)$$

$$R_1(\mathbf{r}) = \mathcal{P}_1(x, y, z) = z \quad (19.2)$$

$$R_2(\mathbf{r}) = \mathcal{P}_2(x, y, z) = \frac{2z^2 - (x^2 + y^2)}{2} \quad (19.3)$$

$$R_3(\mathbf{r}) = \mathcal{P}_3(x, y, z) = \frac{2z^3 - 3z(x^2 + y^2)}{2} \quad (19.4)$$

$$R_4(\mathbf{r}) = \mathcal{P}_4(x, y, z) = \frac{35z^4 - 30z^2(x^2 + y^2) - 3(x^2 + y^2)^2}{8} \quad (19.5)$$

$$\vdots \quad (19.6)$$

不难验证它们都会符合拉普拉斯方程:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \mathcal{P}_l(x, y, z) = 0$$

就是这样的正规立体谐函数中的某项代表的电势,如果作为外场施加在置于原点的接地导体球上,应该得到怎样的静电平衡结果呢?首先我们明确这种场合下第一类边值问题表述为:

$$\nabla^2 \varphi = 0 (|\mathbf{r}| > a) \quad , \quad \varphi|_{|\mathbf{r}|=a} = 0 \quad , \quad \varphi|_{|\mathbf{r}| \rightarrow +\infty} = \lambda R_l(\mathbf{r})$$

由于正规立体谐函数电荷摆放在无穷远处,而球面上感应电荷产生的电势在无穷远处影响一次反比于半径,故此时边界条件变成了一个渐进边值条件(asymptotic boundary condition),它规定了半径趋近于无穷远时电势函数的变化趋势.

这个方程的解非常平庸地等于同阶的非正规立体谐函数与原来的正规谐函数的叠加:

$$\varphi(\mathbf{r}) = \lambda R_l(\mathbf{r}) + \mu I_l(\mathbf{r})$$

这是非常显然的.同阶的正规与非正规谐函数关于角度的部分变化趋势是一致的,为了满足在球面上电势为零的边值条件,我们只需要把半径取 a 并视作系数,再要求两个系数抵消,这只需要要求:

$$a^l \lambda + \frac{\mu}{a^{l+1}} = 0$$

从而马上得到 $\mu = -a^{2l+1} \lambda$.或者更简单地,我们把以下电势作为这一类问题的基础解:

$$\varphi = \frac{V_0}{2l!} \cdot \left[\left(\frac{a}{r} \right)^{l+1} - \left(\frac{r}{a} \right)^l \right] \cdot \left(\frac{d}{\sin \theta d\theta} \right)^l \sin^{2l} \theta$$

它代表怎样的电场呢?我们尤其关注 $l = 0, l = 1$ 两种简单情况.此时电势为:

$$l = 0 \quad : \quad \varphi = \frac{V_0 a}{r} - V_0$$

$$l = 1 \quad : \quad \varphi = \frac{V_0 a^2 \cos \theta}{2r^2} - \frac{V_0}{2a} r \cos \theta$$

作为外问题, $l = 0$ 电势中的 $-V_0$ 就必须理解为外场,而 $\frac{V_0 a}{r}$ 则理解为球面上的感应电荷产生的场,很显然,这表示一个均匀带电的球面外的场,如图19.10左的第一种情况.

作为外问题, $l = 1$ 电势中的 $-\frac{V_0}{2a} r \cos \theta = -E_0 z$ 就必须理解为外场,而 $\frac{V_0 a^2 \cos \theta}{2r^2} = \frac{E_0 a^3 \cos \theta}{r^2}$ 则理解为球面上的感应电荷产生的场,很显然,外场此时成为了 $+z$ 方向的匀强电场 $E_0 e_z$,而球面上的感应电荷产生的电势恰好与一个电偶极子产生的电势一致,电偶极矩为 $\mathbf{p} = 4\pi a^3 \varepsilon_0 E_0$.如图19.10左的第二种情况.

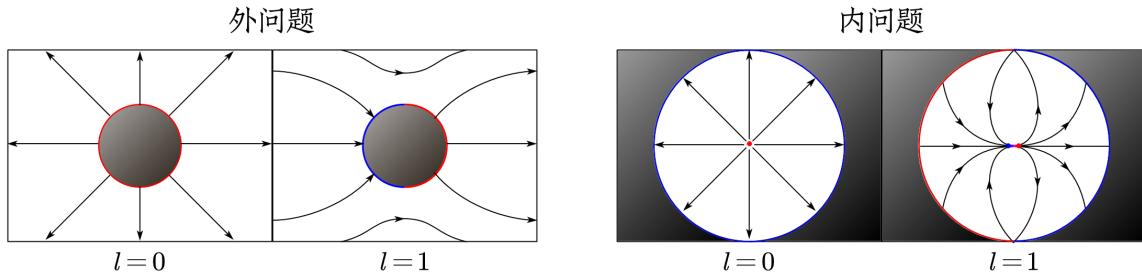


图 19.10: 低阶立体谐函数静电平衡

有意思的是,这样的解也可以适用于内问题.此时球面上的感应电荷在原点产生的场应该展开为正规项,而非正规项对应放在球心的外电荷,此前的解两项将交换意义:

作为内问题, $l = 0$ 电势中的 $\frac{V_0 a}{r}$ 就必须理解为外场,而 $-V_0$ 则理解为球面上的感应电荷产生的场,很显然,这相当于球心放了一个 $Q = 4\pi\varepsilon_0 a V_0$ 的点电荷造成的静电平衡,如图19.10右的第一种情况.

作为内问题, $l = 1$ 电势中的 $\frac{V_0 a^2 \cos\theta}{2r^2} = \frac{p_0 \cos\theta}{4\pi\varepsilon_0 r^2}$ 就必须理解为外场,而 $-\frac{V_0}{2a} r \cos\theta = -\frac{p_0}{4\pi\varepsilon_0 a^3} z$ 则理解为球面上的感应电荷产生的场,很显然,外场此时由 $+z$ 方向的电偶极子 $p_0 e_z$ 产生,而球面上的感应电荷产生的电势恰好是一个匀强电场 $E = \frac{p_0}{4\pi\varepsilon_0 a^3}$. 如图19.10右的第二种情况.

立体谐函数的意义在于,我们可以在原点附近去以球坐标的方式用正规的那些函数去展开一个电势,或者在一个球面外用非正规的那些函数去描写球面内的电荷体系产生的电势.故它天然地与多级展开挂钩.我们记得上一章我们用多级展开的思想来处理了偏移中心电荷的电势,多展开几项就是:

$$\frac{1}{4\pi\varepsilon_0} \frac{q}{|\mathbf{r} - \mathbf{R}|} = \frac{q}{4\pi\varepsilon_0} \left[\frac{1}{r} + \frac{R \cos\theta}{r^2} + \frac{R^2(3\cos^2\theta - 1)}{2r^3} + \frac{R^3(5\cos^3\theta - 3\cos\theta)}{2r^4} + \dots \right]$$

轻易就能发现,展开后各项,原来对应总电荷,电偶极矩,电四极矩,电八极矩等等,现在恰好就是一个个立体谐函数.它便是这样自然地产生于电势的多极展开中.个中原理在于以下展开式,又称**勒让德多项式的母函数**(generatin function of Legendre polynomials):

$$\frac{1}{\sqrt{1 - 2xt + t^2}} = \sum_{n=0}^{\infty} P_l(x)t^l$$

而其中 $P_l(x)$ 就是此前使用的**勒让德多项式**(Legendre polynomials).我们此前是通过**罗德里格公式**(Rodrigues' formula) 引入它的:

$$P_l(x) = \frac{(-1)^l}{2^l l!} \frac{d^l}{dx^l} (1 - x^2)^l = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l$$

我们简要介绍一种通过母函数证明罗德里格公式的方法:

容易发现,根式 $\sqrt{1 - 2xt + t^2}$ 出现在以下方程的解中:

$$tr^2 - 2r + 2x - t = 0 \quad \Rightarrow \quad r = \frac{1 - \sqrt{1 - 2xt + t^2}}{t}$$

这样 t 也可以用 r 表示:

$$t = \frac{2(x - r)}{1 - r^2}$$

但是 r 对 t 的泰勒展开为 $r = x + (x^2 - 1)t/2 + O(t^2)$. 故重新命:

$$s = r - x = \frac{1 - tx - \sqrt{1 - 2xt + t^2}}{t} = f(t)$$

这样用 s 表示 t 就是:

$$t = \frac{2s}{(s + x)^2 - 1} = g(s)$$

这样就存在两个互反函数 $f \circ g = id$. 且两者泰勒展开第一项都为一次项:

$$f(t) = f_1 t^1 + f_2 t^2 + \dots, \quad f_1 = \frac{x^2 - 1}{2}$$

$$g(s) = g_1 s^1 + g_2 s^2 + \dots, \quad g_1 = \frac{2}{x^2 - 1}$$

关于这样的函数有一个著名的拉格朗日反演定理(Lagrange inversion theorem), 如果用下角标表示泰勒展开的系数:

$$f_n = \frac{1}{n} \left(\frac{id}{g} \right)_{n-1}^n$$

写的更明确一点就是:

$$\frac{d^n}{dt^n} f(t) \Big|_{t=0} = \frac{d^{n-1}}{dA^{n-1}} \left(\frac{s}{g(s)} \right)^n \Big|_{s=0}$$

这个定理的证明如下: 把 $g(s)$ 代入 f 的泰勒展开式:

$$s = \sum_{m=1}^{\infty} f_m g^m(s)$$

两边对 s 求导数后乘以 $s^n/g^n(s)$:

$$\begin{aligned} \left(\frac{s}{g(s)} \right)^n &= s^n \sum_{m=1}^{\infty} f_m \frac{[g^m(s)]'}{g^n(s)} \\ &= s^n f_n \frac{[g^n(s)]'}{g^n(s)} + s^n \sum_{m \neq n}^{1 \rightarrow \infty} \frac{m}{m-n} f_m [g^{m-n}(s)]' \\ &= n f_n s^n \frac{g'(s)}{g(s)} + s^n \left(\sum_{m \neq n}^{1 \rightarrow \infty} \frac{m f_m}{m-n} g^{m-n}(s) \right)' \\ &= n f_n s^n \frac{g_1 + 2g_2 s + \dots}{g_1 s + g_2 s^2 + \dots} + s^n \left(\sum_{l=-n+1}^{\infty} \frac{l f_{n+l}}{l} (g_1 s + g_2 s^2 + \dots)^l \right)' \\ &= n f_n s^{n-1} \left(1 + \frac{g_2}{g_1} s + \dots \right) + s^n \left(\dots + \frac{c_{-2}}{s^2} + \frac{c_{-1}}{s} + c_0 + c_1 s + c_2 s^2 + \dots \right)' \end{aligned}$$

后一项中, 任意 s 的幂函数的导数都不会带来 s^{-1} 的结果项. 故等号右侧唯一的 s^{n-1} 项就是 $n f_n s^{n-1}$. 从而可以提取左侧的 s^{n-1} 项系数得到右侧的 $n f_n$. 原命题得证.

将这个结果重新表示 $f(t)$ 的泰勒展开, 得到:

$$f(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!} \frac{d^{n-1}}{dA^{n-1}} \left(\frac{s}{g(s)} \right)^n \Big|_{s=0}$$

现在, 把此前的 $g(s)$ 代入:

$$\begin{aligned} \frac{1 - tx - \sqrt{1 - 2xt + t^2}}{t} &= \sum_{l=1}^{\infty} \frac{t^l}{2^l l!} \frac{d^{l-1}}{dA^{l-1}} [(s+x)^2 - 1]^l \Big|_{s=0} \\ &= \sum_{l=1}^{\infty} \frac{t^l}{2^l l!} \frac{d^{l-1}}{dx^{l-1}} (x^2 - 1)^l \end{aligned}$$

两边再次对 x 求导数:

$$\begin{aligned} -1 + \frac{1}{\sqrt{1 - 2xt + t^2}} &= \sum_{l=1}^{\infty} \frac{t^l}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l \\ \Rightarrow \quad \frac{1}{\sqrt{1 - 2xt + t^2}} &= \sum_{l=0}^{\infty} P_l(x) t^l, \quad P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l \end{aligned}$$

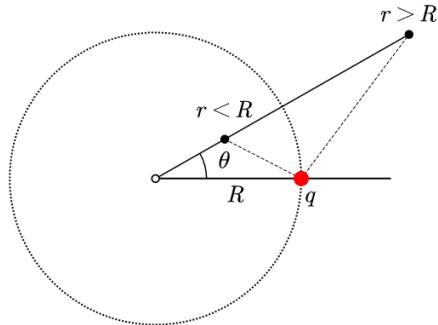


图 19.11: 电势的勒让德展开

有了这个公式就可以把点电荷的电势进行展开,如图19.11所示,在半径 r 很小时原点附近应当视 $x = r/R$ 展开为 r 的正规项,但是在半径 r 很大时的远方应当视 $x = R/r$ 展开为 r 的非正规项:

$$\varphi = \frac{q}{4\pi\epsilon_0} \frac{1}{\sqrt{R^2 + r^2 - 2Rr \cos\theta}} = \begin{cases} \sum_{l=0}^{\infty} \frac{qr^l P_l(\cos\theta)}{4\pi\epsilon_0 R^{l+1}} & (r < R) \\ \sum_{l=0}^{\infty} \frac{qR^l P_l(\cos\theta)}{4\pi\epsilon_0 r^{l+1}} & (r > R) \end{cases}$$

现在让我们来尝试一下新的工具用来解球面内外放置点电荷的第一类边值问题的结果.对于外问题,在 $r = a$ 处放置球面并接地,在 $r = R$ 处放置点电荷 q 并沿其方向建立球极坐标.则球面感受到电势:

$$\varphi_q = \sum_{l=0}^{\infty} \frac{qr^l P_l(\cos\theta)}{4\pi\epsilon_0 R^{l+1}} = \sum_{l=0}^{\infty} \frac{qa^l}{4\pi\epsilon_0 R^{l+1}} P_l(\cos\theta) \left(\frac{r}{a}\right)^l$$

根据此前的讨论,球面上的感应电荷就会在 $r > a$ 产生一个电势:

$$\varphi_{q^*} = - \sum_{l=0}^{\infty} \frac{qa^l}{4\pi\epsilon_0 R^{l+1}} P_l(\cos\theta) \left(\frac{a}{r}\right)^{l+1} = \sum_{l=0}^{\infty} \frac{(-qa/R)(a^2/R)^l P_l(\cos\theta)}{4\pi\epsilon_0 r^{l+1}}$$

恰好对应一个放置在 $r = R^*$ 处的 q^* 产生的电势:

$$q^* = -\frac{a}{R}q, \quad R^* = \frac{a^2}{R}$$

内问题的验证在这里留给读者作为一个练习.

19.3 电介质

之前介绍过,绝缘体只有考虑漏电或是大电场下击穿时,导电才得以发生.但这绝对不意味着忽略导电现象以后绝缘体对于静电场没有任何响应.这时发生的一个重要的现象为**极化**(polarization).它会造成很多独特现象:从在头发上摩擦过的塑料笔杆对小纸屑的吸引⁶,到电气工业大电容器的制作,更广义地说,在交变电场中介质的极化最终还能决定电磁波或光在介质中的传播,下一章我们将首次小小讨论以下这个问题.我们把明显能产生极化的介质称作**电介质**(dielectrics).极化的含义可以从微观和宏观两个角度来考虑,背后涉及物理学的多个方面.我们分别进行讨论.最后揭示微观角度与宏观角度极化规律的联系.

19.3.1 微观角度理解极化

固体或晶体,尤其是离子晶体或部分共价晶体,每一个原子实际上本身就带电,其极化机理非常复杂,需要用到爱因斯坦,玻恩,黄坤等人的陆续完善的理论.或是放弃研究微观机理,通过宏观的方式引入唯象的描述方法.我们下面研究的主要是由原子,分子构成的气体或者液体的极化.

在这样的场合下,每一个单元被单独研究,原子或分子整体是没有电性的,但是正如之前所说的,下式决定的原子或分子的电偶极矩不必为零:

$$\mathbf{p} = \int \rho \mathbf{r} dV$$

如果从非量子的经典物理观点(包括半经典的玻尔模型)来看,即使简单如氢原子,看作轻的电子绕近似不动的质子做匀速圆周运动,每一个瞬时也会形成一个电偶极矩.但是它随时间快速转动而导致在静电场意义下其平均值为零.事实上,量子力学认为无论哪一个时刻,氢原子都处于电偶极矩沿任何方向概率都相

⁶用金属屑的静电平衡也可以实现,不过金属屑密度大现象更不明显.

等的定态(stationary state), 并没有在随时间演变. 以无论是从理论上来看还是从实际上来看, 孤立氢原子基态下总是不体现电偶极矩的:

$$\bar{p} = 0$$

具有这样性质的分子就称作无极分子或非极性分子(nonpolar molecule), 反之则称作有极分子或极性分子(polar molecule) 一般可以通过化学原理给出分子的成键和立体构型加以判断. 例如 CO 分子和 H₂O 具有极性而 CO₂ 分子和 CH₄ 则是无极的.

现在把无极分子放在外电场中, 情况就会发生戏剧性的转变: 把无极分子感受到的局部的电场视作匀强场, 这个外场记做 \mathbf{E}' . 即使是简单如氢原子, 其平均电偶极矩也不应当再视作零. 这一个现象有若干解释方法:

- 最唯象的解释方法把氢原子视作一个由劲度系数 k 的弹簧连接的作为正电负电中心的质点构成的体系, 带电量为 $\pm e$. 但是忽略正负电荷之间的电相互作用力, 只用去考虑受到的外力与等效电相互作用出来的线性弹性力的平衡, 那么自然平衡时从负电荷到正电荷具有位移:

$$\mathbf{r} = \frac{e\mathbf{E}'}{k}$$

从而得到分子平均电偶极矩:

$$\bar{p} = er = \frac{e^2}{k} \mathbf{E}'$$

- 一个具有可证伪性的结果由以下考虑描述. 通过量子理论的计算, 把原子外的电子看作电子云, 它是一个总电量为 $-e$ 的连续电荷密度分布:

$$\rho = \frac{-e}{\pi a_0^3} e^{-\frac{2r}{a_0}}$$

其中 $a_0 = 5.3 \times 10^{-11} \text{ m}$ 为玻尔半径. 进一步假设这个电荷分布为彻底刚性的. 在外电场作用下, 仅仅是原来在原点的质子现在位移到了 \mathbf{r} 处. 质子对电子云的力就是电子云对质子的力, 它与外电场力平衡. 这样计算可以得到:

$$\mathbf{r} = \frac{3\pi\varepsilon_0 a_0^3}{e} \mathbf{E} \quad , \quad \bar{p} = 3\pi\varepsilon_0 a_0^3 \mathbf{E}'$$

- 以上结果仅仅是与实验测量结果数量级一致, 但是差了 6 倍, 即实验结果表明:

$$\bar{p} = 18\pi\varepsilon_0 a_0^3 \mathbf{E}'$$

这是因为上一个做法中认为电子云不会发生任何改变, 反倒是质子位移了的做法不符合事实. 正确的做法反而需要依然以质子为原点去解库仑场与弱匀强电场下电子的薛定谔方程的微扰解. 这样才可以给出电子云的新分布. 同样的计算也会带来著名的斯托克效应(Stark effect). 即玻尔模型对应的第 n 能级(第 $n - 1$ 激发态) 现在将分裂为 $2n - 1$ 个能级. 进而影响谱线.

不管用何种方式解释, 以上现象就是第一种极化机理, 称作位移极化(distortion polarization). 它的结果为每一个分子产生了一个量子平均的电偶极矩, 在弱场下这个电偶极矩应当正比于电场强度:

$$\bar{p} = \alpha \mathbf{E}'$$

其中 α 称作分子极化率(molecular polarizability). 它是表征微观极化难易程度的重要参数. 对于位移极化, 分子极化率是一个量子行为的结果, 从氢原子的极化率就可以看出, 因为玻尔半径 a_0 的计算依赖于普朗克常数的取值:

$$\alpha = 18\pi\varepsilon_0 a_0^3 = \frac{9(4\pi\varepsilon_0)^4 \hbar^6}{2m_e^3 e^6} = 7.42 \times 10^{-41} \text{ F} \cdot \text{m}^2$$

而对于有极分子, 分子即使在没有外场作用时, 由于其电荷分布本身就会造成一个电偶极矩, 称作**固有电偶极矩**(permanent dipole moment), 我们记做 p_0 , 但是由于无规则的热运动, 其方向 e 是随机的, 所以作为每个分子的电偶极矩矢量 $\mathbf{p} = p_0 e$ 的统计平均, 其结果依然应当为 $\mathbf{0}$:

$$\bar{\mathbf{p}} = \mathbf{0}$$

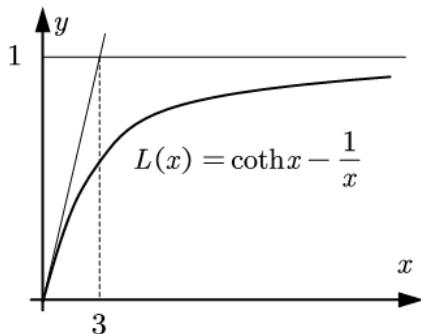


图 19.12: 郎之万函数 $\bar{p} = p_0 L\left(\frac{pE'}{kT}\right)$, $L(x) = \coth x - \frac{1}{x}$

现在施加外电场 E' 以后, 如果认为分子依然保持刚性, 即 p_0 不变, 但方向可以自由转动, 则与电场保持同向的电偶极子的势能得到降低, 电偶极子在动力学上倾向于与外电场同向排列. 但是其热运动越剧烈, 转动动能就越大, 不同电偶极子相互影响就会使得其排列趋于杂乱无章. 忽略可能的量子效应, 由玻尔兹曼分布可以得到在温度为 T 时分子沿电场方向的平均电偶极矩:

$$\bar{p} = p_0 \left(\coth \frac{pE'}{kT} - 1 \right)$$

一般把上面这个函数称作**郎之万函数**(Langevin function), 即:

这个函数的图像如19.12所示. 它有两条渐近线, 分别对应两种具体的物理情形:

如果低温 (T 很小) 或者强场 (E' 很大) 但不至于漏电或击穿. 使得宗量 $x = pE'/kT \gg 1$, 那么就使用水平 $y = 1$ 渐近线, 即:

$$\bar{p} \approx p_0$$

对应几乎所有电偶极子都沿电场方向排列的**饱和极化**(saturated polarization) 情况.

如果高温 (T 很大) 或者弱场 (E' 很小). 使得宗量 $x = pE'/kT \ll 1$, 那么就使用原点切线 $y = x/3$ 的渐近线, 即:

$$\bar{p} \approx \frac{p_0^2}{3kT} E'$$

写成矢量式, 我们发现与位移极化得到了相似的结果:

$$\bar{\mathbf{p}} = \alpha \mathbf{E}' , \quad \alpha = \frac{p_0^2}{3kT}$$

这被称作**郎之万-德拜定律**(Langevin-Debye law). 同样地把 α 称作分子极化率. 而这种极化机理称作**取向极化**(orientation polarization). 取向极化的极化率是一个热学行为的结果, 从极化率反比于温度这一点就可以看出来.

关于取向极化和位移极化我们总结以下几点:

一是要注意两者的不同: 前者是外电场诱导产生或改变固有电偶极矩, 是一个量子行为, 几乎不依赖于温度. 后者是外电场诱导固有电偶极矩选取优势方向, 是一个热学现象, 定性解释不依赖于量子特性. 但是严格说来温度对前者也会有非常微小的影响, 低温下量子特性对后者也会有不小的改变.

二是要注意两者可以同时出现: 无极分子当然只能采取位移极化. 但有极分子的取向极化和位移极化是并存的. 所以分子极化率的完整计算方法为分别考虑两种极化方式带来的系数的和:

$$\alpha = \alpha_{\text{dis}} + \alpha_{\text{ort}}$$

三是要注意具体情况主要考虑哪一种极化方式: 一般说来, 有极分子的极化率中取向极化远大于位移极化的贡献⁷, 故有极分子的极化主要是取向极化. 但是无极分子, 位移极化就取代取向极化称为唯一需要

⁷使分子转向导致的能量差小, 使电子在不同能级间跃迁需要的能量差大, 前者易后者难.

考虑的极化因素了,故静电场下其分子极化率一般比有极分子要小一个量级.但是对于高频电场,**弛豫时间**(relaxation time)成为了另一个需要考虑的重要因素.大约对于超过红外线频率的电场下,取向的改变就难以跟上电场的变化频率了,这导致介质的极化转而成为了电子的位移极化,考虑介质对光的色散和吸收都是基于这一点的.

19.3.2 宏观角度理解极化

之前章节已经介绍过极化强度 \mathbf{P} 的定义与相关的性质.宏观上的极化规律指的是,如果介质内部某点处局域地存在(平均意义下的)场强 \mathbf{E} ,那么该点就局域地产生一个极化强度 \mathbf{P} ,做为 \mathbf{E} 的响应函数:

$$\mathbf{P} = f(\mathbf{E})$$

这样的函数实际上或是应用上都不一定是线性的,但是在各向同性介质,弱场,低频的三个条件下,理论或实验都能给出线性的结果,一般写为:

$$\mathbf{P} = \chi_e \varepsilon_0 \mathbf{E}$$

这样 χ_e 就是只与材料和环境(温度,压强)有关的,而与外加电场大小与方向无关的无量纲标量常数.称作**极化率**(susceptibility).

类似于导体的静电平衡问题,介质的极化是如何达到平衡的?,这一个问题也应当动态地去思考:对比导体中的平衡条件 $\mathbf{E} = \mathbf{0}$,介质中的平衡条件也不过是一个最终才能达到的结果,达到平衡前的弛豫过程,等号不一定要成立.导体未平衡时,内部正是因为有电场才驱使电荷定向运动从而转移,从而电场才被改变,在改变中逐渐趋于零.那么介质也是类似:如果把一块未被极化的介质置于外电场 \mathbf{E}_0 中,一开始 $\mathbf{P} = \mathbf{0}$ 将小于预期的 $\chi_e \varepsilon_0 \mathbf{E}_0$ 值,那么极化就会慢慢发生, \mathbf{P} 开始增大,位移极化和取向极化在逐渐发生,宏观看会产生**极化电流**(polarization current),不难验证其电流密度为 $\mathbf{J}_P = \partial \mathbf{P} / \partial t$.但是由于 \mathbf{P} 的变化,势必会在介质表面甚至内部产生极化电荷分布 σ_P, ρ_P .这些电荷产生的场将部分抵消原来的外电场,称作**退极化场**(depolarization field).于是 \mathbf{E}_0 也将动态的变化到最终的 \mathbf{E} ,使得以上平衡方程成立.这样极化强度就不再增加.

所以同样的以极化平衡条件 $\mathbf{P} = \chi_e \varepsilon_0 \mathbf{E}$ 为出发点,应当期望可以类似于导体静电平衡那样得到有效的求解带电介质的静电平衡问题.但是由于平衡条件更复杂,我们的方法也需要进行改进.事实上不难验证改进的方法取 $\chi_e \rightarrow +\infty$ 就会退化为导体的情况.所以改进的方法反而更具有普遍性.

最核心的一点是引入辅助的**电位移**(electric displacement)矢量来简化问题.这个矢量被定义为:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$$

用 \mathbf{D} 的值反过来计算反映极化的 \mathbf{P} ,这会带来两个好处,第一是,极化规律的平衡条件被等价地写作:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \chi_e \varepsilon_0 \mathbf{E} = \varepsilon_r \varepsilon_0 \mathbf{E} = \varepsilon \mathbf{E}$$

同样的 $\varepsilon_r = 1 + \chi_e$ 也会作为一个基本常数,与原来的极化规律的结构复杂度一致.这个常数称作**相对介电常数**(relative permittivity),其与真空介电常数 ε_0 的乘积 $\varepsilon = \varepsilon_r \varepsilon_0$ 则称作**绝对介电常数**(absolute permittivity)或简单称作**介电常数**(permittivity).合理的极化率应当为正数,而根据相对介电常数的定义,它则要是大于 1 的正数.而介电常数的实用程度大于极化率的实用程度.

第二个关键的好处是,我们已经知道, \mathbf{E} 与电荷通过高斯定理联系,而 \mathbf{P} 与极化电荷也可以通过高斯定理联系:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}, \quad \nabla \cdot \mathbf{P} = -\rho_P$$

代入刚才的电位移的定义马上就可以发现:

$$\nabla \cdot \mathbf{D} = \rho_f$$

即, 电位移通过高斯定理联系的电荷就是“自由”电荷. 这里的“自由”不是指导体中或导体表面的电荷, 恰恰相反, 如果把导体静电平衡问题中的导体也视作 $\epsilon_r \rightarrow +\infty$ 的特殊电介质, 这些电荷反倒要被排除. ρ_f 的含义通过推导上式使用的关系式 $\rho_f = \rho - \rho_P$ 就很容易看出来了, 表示除去介质极化产生的面电荷或体电荷以外的一切电荷, 一般都是置于介质外部区域的给定电荷, 但对于介质情况它也可以分布在介质内.

对于带介质的静电平衡问题的求解就有了两种思路, 第一种思路仅仅适用于比较简单的情况, 这种情况不需要借助我们引入的电位移矢量 D 就能得到结果. 分别是把无限大块状介质和介质球放在匀强电场 E_0 中的求解:

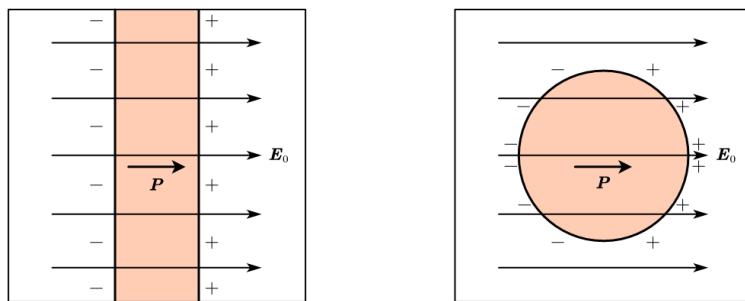


图 19.13: 两种简单情况

这两个问题的核心在于大胆猜测介质内部是均匀极化的. 设极化强度为 P . 对于平板介质块, 极化电荷在两侧产生的电荷面密度 $\pm\sigma_P$ 就是:

$$\sigma_P = P$$

从而在介质内部形成的退极化场就是:

$$\mathbf{E}_P = -\frac{\mathbf{P}}{\epsilon_0}$$

从而极化规律导致:

$$\mathbf{P} = \chi_e \epsilon_0 \mathbf{E} = (\epsilon_r - 1) \epsilon_0 \left(\mathbf{E}_0 - \frac{\mathbf{P}}{\epsilon_0} \right)$$

这就解出了平衡时的极化强度:

$$\mathbf{P} = \frac{\epsilon_r - 1}{\epsilon_r} \epsilon_0 \mathbf{E}_0$$

我们关心的是介质内部的电场强度:

$$\mathbf{E} = \mathbf{E}_0 - \frac{\mathbf{P}}{\epsilon_0} = \frac{\mathbf{E}_0}{\epsilon_r}$$

这个结论很贴切地反应了 ϵ_r 的物理意义: 它是简单的平板情况下介质对电场的削弱. 故对于导体情况, 内部电场强度在静电平衡时被削弱为零, 故可以视作 $\epsilon_r \rightarrow +\infty$.

注意到这样的 $\pm\sigma_P$ 在介质块外场强为零, 故不影响外界. 从而如果把多块介质块叠放, 不同介质块就相当于按照不同的相对介电常数去削弱同一个电场. 也就是说如图19.14的三个介质块上方下方的导体板上的电荷面密度如果为 σ , 就会有:

$$E_0 = \frac{\sigma}{\epsilon_0}, \quad \epsilon_{r1} E_1 = \epsilon_{r2} E_2 = \epsilon_{r3} E_3 = E_0$$

但是要注意, 这个 E_0 不一定要与旁边的空间中的 E' 相等, 即在 S 和 S' 面上的电荷面密度没有必要是相等的, 同一个表面的带电荷量也不一定均匀. 然而 $E_i, i = 1, 2, 3$ 也没有必要等于 E_0 , 这不会与电场的环路定理矛盾, 因为在介

质块的边缘电场实际上已经不能视作匀强电场,而这一点作为边缘效应被我们忽略了.只有以下等势条件式是必须成立的:

$$E_1 d_1 + E_2 d_2 + E_3 d_3 = E' d = U$$

除了严格把以上量都解出来,我们讨论这个体系的静电平衡时经常会把它看做电容,每一个电容的性质和值由以下关系描述:

$$Q = CU \quad , \quad C = \frac{\epsilon S}{d}$$

注意电容器的并联等效电容应当满足直接求和的关系,串联反倒需要用电阻并联的方式来计算:

$$C_{\text{series}} = C_1 + C_2 \quad , \quad C_{\text{parallel}} = C_1 \parallel C_2 = \frac{C_1 C_2}{C_1 + C_2}$$

于是上一个问题的总电容就是:

$$C = (C_1 \parallel C_2 \parallel C_3) + C'$$

现在让我们关心一下在三个介质中的电位移矢量,就会发现:

$$D_1 = D_2 = D_3 = \epsilon_0 E_0 = \sigma$$

可以发现电位移与忽略电介质时的电场其实就差一个物理常数 ϵ_0 . 它的存在可以理解为仅仅为了统一量纲.

现在让我们看介质球的极化问题. 此时均匀的极化为球的表面带来的电荷面密度就变成了 $\sigma_P = P \cos \theta$, 而如上一章所述,这样的电荷分布带来的退极化场为:

$$\mathbf{E}_P = -\frac{\mathbf{P}}{3\epsilon_0}$$

仅为块状情况的三分之一. 可以想见,其极化会更加的强. 事实上平衡方程与解变为:

$$\mathbf{P} = (\epsilon_r - 1)\epsilon_0 \left(\mathbf{E}_0 - \frac{\mathbf{P}}{3\epsilon_0} \right) \Rightarrow \mathbf{P} = \frac{3(\epsilon_r - 1)}{\epsilon_r + 2}\epsilon_0 \mathbf{E}_0$$

那么球内的电场就被削弱到了:

$$\mathbf{E} = \mathbf{E}_0 - \frac{\mathbf{P}}{3\epsilon_0} = \frac{3}{\epsilon_r + 2}\mathbf{E}_0$$

可见此时 ϵ_r 就没有体现出它的物理意义来: 它并不能代表任意情况下介质对外加电场的削弱的比例,这里被削弱的比例变成了:

$$\epsilon = \frac{\epsilon_r + 2}{3}$$

从而这种情况下研究电位移表面上看上去没有意义:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} = \frac{3\epsilon_r}{\epsilon_r + 2}\epsilon_0 \mathbf{E}_0$$

其实不然,下面我们来看带电介质的静电平衡问题的第二种思路,它也是普遍问题的处理思路,主要就是需要借助电位移矢量来考虑问题的.

首先,在同一种介电常数为 ϵ 的介质中,我们抛弃极化矢量 \mathbf{P} ,用场强度 \mathbf{E} 和电位移 \mathbf{D} 可以完备地表达出极化平衡以后的场与荷(下面默认去除极化电荷,平衡的时候极化电荷的值是有唯一解的未知量)关系,它由三个关系构成:

$$\mathbf{D} = \epsilon \mathbf{E}$$

$$\begin{cases} \nabla \cdot \mathbf{D} = \rho \\ \nabla \times \mathbf{E} = 0 \end{cases} \quad \text{或} \quad \begin{cases} \oint_{\partial V} \mathbf{D} \cdot d\mathbf{A} = \int_V \rho dV \\ \oint_{\partial A} \mathbf{E} \cdot dl = 0 \end{cases}$$

第一个式子称作电介质的**本构方程**(constitutive equation). 后两个式子就是介质中的高斯定律和环路定律. 实际上还缺少一个势与它们的关系, 势与场之间的关系不需要做任何改变:

$$\mathbf{E} = -\nabla \varphi$$

从而很容易得到势与荷之间的关系, 现在变为:

$$\nabla^2 \varphi = -\frac{\rho}{\varepsilon}$$

那么现在我们考虑电介质的静电平衡问题的表述方法. 与导体静电平衡的区别在于, 除了各真空区域 $\varepsilon = \varepsilon_0$ 的电势分布待解以外, 各个介质内部的电势分布也是待解的, 我们甚至可以把外电荷引入介质的内部. 这样我们索性就把空间按大块介质分成不同不相交的区域 D_i 的并:

$$\mathbb{R}^3 = \bigcup_i D_i$$

每一个区域内部的介电常数为 ε_i . 而待解的电势为 φ_i , 如果该区域内有外电荷分布 $\rho(\mathbf{r})$, 则在介质内部需要满足:

$$\nabla^2 \varphi(\mathbf{r})|_{\mathbf{r} \in D_i} = -\frac{\rho(\mathbf{r})}{\varepsilon_i}$$

但是凭以上方程不足以确定整个问题, 一般还需要配合**边界条件**(boundary relation). 介质静电平衡问题的边界条件有它独特的重要性, 一般会做如下表述:

如图19.15, 在介质 i, j 的分界面上, 把 i 指向 j 为正方向的场的法向分量的电位移记为 D_{ni}, D_{nj} . 而切向分量以其自然方向为正方向的电场强度记做 $E_{\tau i}, E_{\tau j}$. 如果面上没有外电荷, 只有极化电荷, 那么根据高斯定律和环路定律, 容易得到:

$$D_{ni} = D_{nj}, \quad E_{\tau i} = E_{\tau j}$$

通常称作: **法向电位移连续, 切向电场强度连续.**

但是面上如果有 σ_{ij} 的外电荷, 第一个式子就必须改为:

$$D_{nj} - D_{ni} = \sigma_{ij}$$

用这样的方程来确定平板电容器或其他具有高对称性电容器中的场分布非常方便. 即使是之前的球在匀强电场中极化的问题, 抛弃极化矢量而取而代之利用电位移也不会造成困难. 此时我们假设表面上的电荷分布为 $\sigma = \sigma_0 \cos \theta$. 那么对面内面外产生的退极化场场强为:

$$\mathbf{E}_P = \begin{cases} -\frac{\sigma_0}{3\varepsilon_0} \mathbf{e} & (r < R) \\ \frac{1}{4\pi\varepsilon_0} \frac{2p \cos \theta}{r^3} \mathbf{e}_r + \frac{1}{4\pi\varepsilon_0} \frac{p \sin \theta}{r^3} \mathbf{e}_\theta & (r > R) \end{cases}$$

其中 $p = 4\pi\sigma_0 R^3 / 3$. 现在取在 $\theta = 0, r = R$ 处的分界面, 将退极化场与外场叠加, 得法向内外场强分别为:

$$E_1 = E_0 - \frac{\sigma_0}{3\varepsilon_0}, \quad E_2 = E_0 + \frac{1}{4\pi\varepsilon_0} \frac{2p}{R^3} = E_0 + \frac{2\sigma_0}{3\varepsilon_0}$$

再由法向电位移连续条件, 即得到:

$$\varepsilon_r \varepsilon_0 E_1 = \varepsilon_0 E_2 \Rightarrow \sigma_0 = \frac{3(\varepsilon_r - 1)}{\varepsilon_r + 2} \varepsilon_0 E_0$$

代入之前的电场表达式就得到了平衡时的场强分布.

但是含介质问题的静电平衡问题难度有甚于导体的静电平衡问题:在导体静电平衡问题中我们发展了格林函数法,并对区域 D 的边界为平面和球面两种情况分别引入了电像法.但是对于介质问题,可以证明平面情况依然可以使用电像法,但是球面的静电平衡问题外电荷已经不存在单个的简单的电像.此时原则上静电平衡的求解需要用到偏微分方程求解的通用思路,这已经超出本教材的范围,在此不再加以介绍.

19.3.3 微观与宏观的联系

我们已经得到微观和宏观的极化规律并使用后者解决了一部分问题:

$$\bar{p} = \alpha E' , \quad P = \chi_e \varepsilon_0 E$$

当然如果单个分子的平均电偶极矩为 \bar{p} ,而这些分子以数密度 n 组成介质,那么介质的极化强度满足以下关系式也是正确无误的:

$$P = n\bar{p}$$

但因此而认为微观的分子极化率是通过以下关系式确定宏观的极化率 χ_e 的话就不够准确了:

$$\chi_e = \frac{n\alpha}{\varepsilon_0}$$

这是因为 E' 和 E 略有区别.但是我们先指出,以上做法对于稀疏的材料,比如气体,通常是足够准确的了.但是一旦分子稠密起来,比如低温下的液氦或是分子晶体固体,就必须考虑 E' 和 E 的区别.前者是造成分子极化的场强,而后者则是在宏观小,微观大尺度下平均化的局域场强.两者的区别实际上就在于偶极子自己的场强在前者计算中需要略去,因为自己的场强不可能极化自己.而后者则需要加以考虑.定量上的差值实际上我们在上一章末也进行过计算,关系式为:

$$E' = E + \frac{P}{3\varepsilon_0}$$

这样就得到更准确的关系式:

$$\chi_e = \frac{n\alpha}{\varepsilon_0 - n\alpha/3}$$

可见,如果 $n\alpha \ll \varepsilon_0$,那么之前的近似计算法就可以适用,此时总是有 $\chi_e \ll 1$,从而只会得到 $\varepsilon_r \approx 1$ 介质的介电特性与真空区别不大.所以对于所有介电特性明显的电介质,实际的联系微观分子极化率与宏观极化率的公式必然是以上式子,或者写为相对介电常数:

$$\varepsilon_r = \frac{\varepsilon_0 + 2n\alpha/3}{\varepsilon_0 - n\alpha/3}$$

这两个公式称为克劳修斯-莫索蒂关系(Clausius-Mossotti relation).我们对于稀薄介质的判定条件进一步加以说明.如果采用氢原子的位移极化模型,我们就知道 $\alpha = 18\pi\varepsilon_0 a_0^3$,代入条件 $n\alpha \ll \varepsilon_0$ 就得到了:

$$na_0^3 \ll 1$$

如果把每个分子占据的体积近似为 a_0^3 ,这就是说,单位体积内被分子占据的体积必须是个小量,主要的体积必须是被真空占据才能称作“稀薄”.这一点也符合物理的直觉.但是要注意,如果考虑取向极化,分子极化率要大一个量级,那么就要求 n 更小一个量级才能符合“稀薄”的条件.

19.4 再议静电能

导体静电平衡时的总静电能计算并不会带来任何问题.之前的用荷与势的公式和用场的计算公式依然适用:

$$I = \frac{1}{2} \int_V \rho \varphi dV = \int_V \frac{1}{2} \epsilon_0 \mathbf{E}^2 dV$$

唯一要注意的问题是别在使用电像法的场合,把像电荷也代入第一个电荷乘电势的公式去计算了.这很可笑,第一可笑是即使出现了像电荷,它也仅仅是作为一种替代:用来替代某一部分感应电荷在某一特定区域产生的场强与电势.但绝对不是说这个虚假的电荷和它那个位置的实际的电势会对这个像电荷产生一个实际的能量.第二可笑是因为在更多的静电平衡问题中原则上没法用电像法,就更加不存在使用上面的积分计算静电能时没有考虑像电荷上的能量的问题了.

还有一个细节可以简化:实际上导体静电平衡问题中除了第一类或第二类区域,现在统一记为 D_i 中可以有 $\rho(r)$ 分布以外,其余部分就是各个导体,每一个导体都等势,故具有带电量 Q_j 和电势 V_j .那么总的静电能就可以更具体地表达为:

$$I = \sum_i \frac{1}{2} \int_{D_i} \rho \varphi dV + \sum_j \frac{1}{2} Q_j V_j$$

如果实际电荷分布实际由微观的点电荷形成,那么在平均场意义下的上式应当代表所有点电荷之间的相互作用能的和.而点电荷的自能并不包含在上式中:它一般也不会改变.

但是如果遇到电介质极化的情况,问题就大不相同了!因为此时伴随着分子的极化,分子内部的能量也会改变.位移极化改变了分子的自能是不言而喻的.但是取向极化何以改变分子内部的自能?分子仅仅是在外电场下选取了优势方向而已.事实上,注意到极化率是温度的函数,就非常容易理解,此时有效的能量其实不是真正的系统内能,而是热力学理论中的自由能,它代表体系能量最低原理与熵最大原理的一种平衡:

$$F = U - TS$$

当分子取向杂乱无章时,极化强度 $\mathbf{P} = \mathbf{0}$,此时熵最大,自由能最小.可以取为自由能的零点.但是一旦产生极化, $\mathbf{P} \neq \mathbf{0}$,此时熵减小,自由能就因此而增加.

这样就不难理解,讨论介质极化时,此时的能量(自由能)应当具有四个组成部分:

1. 非宏观静电形式的,形成极化导致的极化能(polarization energy).
2. 单纯看介质极化自己产生的,介质内外的电场能,即介质自己的静电自能.
3. 不考虑介质时其他电荷产生的全空间的电场能,即其他电荷分布的静电自能.
4. 介质与其他电荷之间的静电相互作用能.

其中后三个部分构成了我们之前意义上的“静电能”.但是它并不是能量的全部.考虑一个能量守恒的过程时,应当有外界做的功等于体系内部所有能量的变化.如为一个电容器充电的过程.很容易发现,如果介质内部采取取向极化,在一个等温的极化过程中介质是要向外界放热的,因为体系的熵在降低.故外界需要做的功,并不是要与体系的内能挂钩,而是直接与自由能挂钩:

$$W = \Delta U - Q = \Delta U - T \Delta S = \Delta F$$

所以从这个意义上说,我们此时定义的静电能,应当包括第一项极化能,而直接就是指自由能:

$$I = F$$

我们从平板电容器的极化来思考其表达式,因为这种情况场是简单的匀强场,而且区域限制在简单的长方体的内部.我们不妨预先假设,极化能的能量密度应当为极化强度的二次函数:

$$w_p = \frac{1}{2} \alpha \mathbf{P}^2$$

做这个假设的原因是因为自由能必然可以对 \mathbf{P} 泰勒展开, 而空间的各向同性决定了它必然不含一阶的项 (一维情况就是必然是偶函数), 那么处于弱场近似可以忽略高阶项. 其实待会也可以发现, 这是唯一能够得到之前的极化规律的能量密度的取法.

那么在电容器完成极化后, 真正的静电能, 就应当为上式代表的极化能与后三项代表的之前意义上的“静电能”的和, 其能量密度为:

$$w = w_p + w_0 = \frac{1}{2}\alpha\mathbf{P}^2 + \frac{1}{2}\varepsilon_0\mathbf{E}^2$$

此时我们可以将介质的极化看作是一种自由能最小的结果: 需要将外场 \mathbf{E}_0 视作一种不变的约束, 而极化强度被视作自由能中的变量, 在极化平衡时应当有以下自由能密度函数最小:

$$w = \frac{1}{2}\alpha\mathbf{P}^2 + \frac{1}{2}\varepsilon_0(\mathbf{E}_0 - \frac{\mathbf{P}}{\varepsilon_0})^2$$

那么由导数为零的方式得到:

$$\alpha\mathbf{P} = \mathbf{E}_0 - \frac{\mathbf{P}}{\varepsilon_0}$$

对比之前的极化规律, 我们就非常容易看出来:

$$\alpha = \frac{1}{\chi_e\varepsilon_0}$$

从而得到极化能量密度的表达式:

$$w_p = \frac{\mathbf{P}^2}{2\chi_e\varepsilon_0}$$

这一点我们在用稀薄气体模型配合位移极化中的最唯象的弹簧模型进行一个验证. 位移极化使得每个弹簧偏离平衡位置 \mathbf{r} 以后, 弹簧势能造成的能力密度就是极化能的本质:

$$w_p = n \cdot \frac{1}{2}k\mathbf{r}^2$$

再结合宏观极化强度, 微观电偶极矩与位移值三者的关系:

$$\mathbf{P} = np = ner$$

得到:

$$w_p = \frac{\mathbf{P}^2}{2ne^2/k}$$

而由之前讨论过的平衡方程得到分子极化率:

$$\alpha = \frac{e^2}{k}$$

从而宏观极化率, 在稀薄条件下为:

$$\chi_e = \frac{n\alpha}{\varepsilon_0}$$

把上两式去代换之前的极化能, 就得到了:

$$w_p = \frac{\mathbf{P}^2}{2\chi_e\varepsilon_0}$$

所以这就是均匀各向同性的介质的极化能公式. 如果与之前的“静电能”合并, 就得到了体系真正的静

电能,如果利用极化规律就得到:

$$\begin{aligned}
 w &= \frac{\mathbf{P}^2}{2\chi_e\epsilon_0} + \frac{1}{2}\epsilon_0\mathbf{E}^2 \\
 &= \frac{(\chi_e\epsilon_0\mathbf{E})^2}{2\chi_e\epsilon_0} + \frac{1}{2}\epsilon_0\mathbf{E}^2 \\
 &= \frac{1}{2}(1+\chi_e)\epsilon_0\mathbf{E}^2 \\
 &= \frac{1}{2}\epsilon_r\epsilon_0\mathbf{E}^2 \\
 &= \frac{1}{2}\epsilon\mathbf{E}^2 \\
 &= \frac{1}{2}\mathbf{D} \cdot \mathbf{E}
 \end{aligned}$$

可见其实相比纯粹的静电场,能量密度就是单纯的增大了 ϵ_r 倍.

以上静电能也具有用电荷来计算的变式.在一块状均匀介质的内部区域 D_i ,由于此时电势梯度依然是场强 \mathbf{E} ,但是场强的散度与极化电荷挂钩,故如果考虑电位移 $\mathbf{D} = \epsilon\mathbf{E}$,其散度就会是外电荷,从而得到:

$$\nabla^2\varphi = -\frac{\rho}{\epsilon}$$

通过上一章类似的积分公式,在这块介质内部总是有:

$$\frac{1}{2}\int_{D_i}\rho\varphi dV = \int_{D_i}\frac{1}{2}\epsilon\mathbf{E}^2 dV + \oint_{\partial D_i}\frac{1}{2}\varphi \cdot \epsilon\mathbf{E} \cdot d\mathbf{A}$$

前一项的积分是我们熟知的,后一项在朝外为正的表面上的积分可以拆解为各个介质分界面上的项的求和,如果是介质 1, 2 之间的分界面,两项积分合成为:

$$\int_{\Sigma_{12}}\frac{1}{2}\varphi(\mathbf{D}_1 - \mathbf{D}_2) \cdot d\mathbf{A}_{12} = -\int_{\Sigma_{12}}\frac{1}{2}\varphi(D_{2n} - D_{1n})dA$$

由之前说过的介质边界条件,两个法向电位移的差就是在分界面上具有的外电荷面密度,故以上积分被简单化为:

$$-\frac{1}{2}\int_{\Sigma_{12}}\sigma_{12}\varphi dA$$

综上,在分块介质情况下,体系的总静电能就等价于求体内与分界面上两类外电荷与电势产生的以下积分值:

$$I = \sum_i \frac{1}{2} \int_{D_i} \rho\varphi dV + \sum_j \frac{1}{2} \int_{\Sigma_j} \sigma\varphi dA$$

最后,容易验证,即使有电介质,电容器具有的总能量的表达式形式依然不变,为:

$$E = \frac{1}{2}CU^2 = \frac{Q^2}{2C} = \frac{1}{2}QU$$

这一点不难理解:虽然这个能量并不必真正意义上内部的“内能”,但是它作为自由能,的的确确要与充电过程中外电路需要做的功相联系.而在这个意义上通过外电路电压对积累的电荷量积分就会得到完全一致的结果.只是要注意,如果涉及取向极化,电容器充放电的过程就总是伴随着与外界的吸放热.

第二十章 稳恒电流

20.1 稳恒电流描述与形成

20.1.1 德鲁特模型

电荷的定向移动形成电流(current). 就好像一缸气体在慢慢挪动那样, 电荷的定向移动并不是纯粹的匀速运动, 而是与无规则的热运动相叠加. 1900 年前后德鲁特 (P. Drude) 和洛伦兹 (H. A. Lorentz) 等人提出德鲁特模型(Drude model) 来解释金属中的导电现象. 主要观点是金属内部自由运动的电子类似于理想气体那样做自由的运动, 称为自由电子气(free electron gas). 我们用电子电量 $-e$, 质量 m , 电子数密度 n , 和弛豫时间(relaxation time) τ , 平均速度 v 来表示其特征. 弛豫时间就是电子做匀速直线运动, 与原子实两次碰撞之间的平均间隔时间. 与之相关的另一个量还可以是平均自由程(mean free path) λ . 容易想像, 典型的情形是, 常温下电子的平均速度是和气体分子的平均热运动速度那样, 一个非常巨大的速度, 而金属原子之间的距离又是那么地短, 导致电子发生十分频繁的碰撞. 而如果在金属中加一个电场 E , 它在两次碰撞内是只可以让电子速度改变一个十分微小的量的:

$$\mathbf{a} = -\frac{e\mathbf{E}}{m}$$

$$\mathbf{v}(0) \rightarrow \mathbf{v}(t) = \mathbf{v}(0) + \mathbf{a}t \rightarrow \mathbf{v}(\tau) = \mathbf{v}(0) + \mathbf{a}\tau$$

我们根据统计力学的思想, 计算在某一时刻电子速度对所有电子取热平衡分布的平均:

$$\mathbf{u} = \langle \mathbf{v} \rangle = \langle \mathbf{v}(0) + \mathbf{a}t \rangle = \mathbf{a}\tau$$

上式 t 代表距离上次碰撞每个电子幸存的时间. $\mathbf{v}(0)$ 代表上次碰撞后其速度. 一方面, 认为碰撞使得电子速度完全随机分布, 平均的结果为零. 另一方面, 认为电子的碰撞是一个泊松过程, 其碰后幸存时间的概率分布是一个指数分布:

$$p(t) = e^{-\frac{t}{\tau}}$$

从而平均的 t 为

$$\langle t \rangle = - \int_0^\infty t dp = pt|_0^\infty + \int_0^\infty e^{-\frac{t}{\tau}} dt = \tau$$

最后结合电流密度为:

$$\mathbf{j} = -nev$$

我们得到:

$$\mathbf{j} = \frac{ne^2\tau}{m}\mathbf{E}$$

的结果. 称为微观欧姆定律(microscopic Ohm's law), 其中系数被称为电导率(electrical conductivity):

$$\sigma = \frac{ne^2\tau}{m} ; \quad \mathbf{j} = \sigma\mathbf{E}$$

十分值得指出的是,作为德鲁特模型的另一个重要结论,金属的热导率(thermal conductivity)在理论中也可以给出一个估计值.我们都应该知道金属比绝大多数其他固体都拥有好得多的导热性能.这可以用热导率与傅里叶热传导定律(Fourier's law of thermal conduction)来描述:

$$\mathbf{q} = -\kappa \nabla T$$

其中 \mathbf{q} 为热流密度, κ 即为材料的热导率.水常温下热导率只有 $0.591\text{W}/(\text{m}\cdot\text{K})$,但纯的铜却能够达到 $401\text{W}/(\text{m}\cdot\text{K})$.金属的高热导率全都得益于轻盈的电子气,它能够迅速地把局部的热运动加剧传导开来.类比理想气体的非平衡态统计方法,我们给出:

$$\kappa = \frac{1}{3} n v \lambda c$$

式中 c 为每个电子的动能与温度的比.利用 $\lambda = v\tau$, 我们把热导率和电导率做比,便可以把较难确定的散射弛豫时间消去,得到:

$$\frac{\kappa}{\sigma} = \frac{mv^2 c}{3e^2}$$

德鲁特模型认为,作为类似于理想气体的电子气,理应有:

$$c = \frac{3}{2} k \quad ; \quad \frac{1}{2} mv^2 = \frac{3}{2} kT$$

从而得到:

$$\frac{\kappa}{\sigma T} = \frac{3}{2} \left(\frac{k}{e} \right)^2$$

历史上金属热导率和电导率的比值与温度相关的现象很早就被人们发现,一般温度升高时电导率会有十分明显的下降,也就是金属的电阻会上升.小灯泡在未发光时与正常发光时电阻就经常有 2 倍左右的差距.而 1853 年魏德曼 (G. Wiedemann) 和弗朗茨 (R. Franz) 观察到不同金属虽然导电导热性能差距悬殊,室温下两者之比却接近一个常数,称为魏德曼-弗朗茨定律(Wiedemann-Franz's law).而洛伦茨 (L. Lorenz)¹ 在 1872 则把这一经验规律确定到常数与绝对温度成正比的形式.最后德鲁特电子论将这个常数以微观常数的形式确定下来,等式右边现在约为 $1.11 \times 10^8 \text{W} \cdot \Omega/\text{K}^2$, 这与实验结果数量级是一致的,但却差了约 2 倍.真实的值被称为洛伦兹常数(Lorentz number):

$$\frac{\kappa}{\sigma T} = L = \frac{\pi^2}{3} \left(\frac{k}{e} \right)^2 = 2.44 \times 10^8 \text{W} \cdot \Omega/\text{K}^2$$

这个理论很成功,但最后的结果的不符合让人感到疑惑.还有一个令人感到疑惑的问题.就是对于电子气的电容.既然电子是完全独立于金属原子实的另一个热力学体系.按照经典统计物理.原子实的摩尔热容与离子晶体等类似,大概在 $3R$ 附近².而电子按只有平动自由度的理想气体考虑应贡献 $1.5R$ 热容.实验却否定了这一点,金属常温下热容仍然是在 $3R$ 附近.电子仿佛没有被热激发,然而电子又势必参与导热,因为金属的导热性能明显优于其他物质.那么以上推导过程中哪儿存在不合理之处?几年后到来的量子力学革命让人们认识到这一个经典模型从一开始就做了不止一个的与微观实际情况不符的假定,但又是十分巧合地,最终结果与真实值数量级自动一样了.

20.1.2 费米气观点 *

元素周期律引发了人们对原子核外电子排布规律的研究,人们惊奇地发现电子是费米子(fermion),符合泡利不相容原理(Pauli exclusion principle).这赋予电子独特的量子特性.具体来说,单位体积内如果电子数目越多,那么其最低平均能量就必须越大.因为能量最低的状态一旦被占据,其他电子就必须占据能量更

¹注意,丹麦物理学家路德维希·洛伦茨 (Ludwig Lorenz)(1829-1891) 与荷兰物理学家亨德里克·洛伦兹 (Hendrik Lorentz)(1853-1928) 是两个不同的人.有一个方程以它们两人的名字共同命名:Lorenz-Lorentz 关系.

²即督龙-裴替定律(Dulong-Petit Law).

高的状态,即使按照最低能量的方式去堆积(绝对零度时的行为),电子也将具有很高的平均能量.它符合类似位置-动量不确定性原理的反比率,电子浓度的减小了单电子占据的位置尺度,则它的动量就会增加,从而根据色散关系 $p^2 = 2mE$ 其能量也会升高.这一点使得我们去修改经典的德鲁特模型.相应的量子气体称为费米气(Fermi gas).

定量计算费米气的特性需要考虑电子的波动本性.我们暂时取金属为长宽高为 ABC 的长方体,那么如果将 N 个电子倒入这个容器,电子的数密度为:

$$n = \frac{N}{ABC}$$

让我们考虑一下电子对状态的填充,长方体相当于一个谐振腔,事实上给出了电子动量状态的量子化:

$$p_x A = ah, p_y B = bh, p_z C = ch$$

$$a, b, c \in \mathbb{N}$$

这是因为根据德布罗意关系 $p = \hbar k$ 而电子状态需要满足 $kL = 2m\pi$ 的缘故.从而我们发现电子的三个方向的动量都是量子化的.量子化的单位为:

$$\Delta p_x = \frac{h}{A}, \Delta p_y = \frac{h}{B}, \Delta p_z = \frac{h}{C}$$

现在把 N 个电子倒入动量空间中,那么电子在低温下将从最低能量态开始填充,由于电子数目巨大,最后电子将填充到能量为 ε_F 处.这样的一个填满电子的动量空间中的球体称为费米海(fermi sea),最终填充到的能量称为费米能级 ε_F ,对应的电子动量为费米动量 p_F .注意到一个动量空间中的一个标志状态的坐标内部实际上有两个独立的状态,它们表示即使这两个波的波矢 k 一样,它们代表电子的自旋也不一样³.从而我们写出:

$$2 \cdot \frac{4\pi p_F^3}{3} / (\Delta p_x \Delta p_y \Delta p_z) = N$$

从而得到:

$$p_F = \left(\frac{3Nh^3}{8\pi N} \right)^{\frac{1}{3}} = \left(\frac{3nh^3}{8\pi} \right)^{\frac{1}{3}}$$

在以上过程中总电子数和总体积被消去了,最后电子堆积到的动量值仅仅取决于各处电子的数密度而成为了强度量.这样一个由于泡利不相容原理所造成的动量值对应到速度上,对一般金属估计约为 10^6 m/s .而如果按非费米气计算,热运动速度应为 $\sqrt{kT/m} \sim 7 \times 10^4 \text{ m/s}$,我们发现经典结果是严重估计少了的.但电子的热容又估计多了.量子统计给出:

$$c = \frac{\pi^2}{2} k \cdot \frac{kT}{\varepsilon_F}$$

代入热导率公式,得:

$$\frac{\kappa}{\sigma} = \frac{\pi^2}{6} \cdot \frac{mv_F^2}{\varepsilon_F} \cdot \left(\frac{k}{e} \right)^2 \cdot T$$

恰好, $\varepsilon_F = \frac{1}{2}mv_F^2$,从而我们得到了正确形式的魏德曼-弗朗茨定律:

$$\frac{\kappa}{\sigma T} = \frac{\pi^2}{3} \left(\frac{k}{e} \right)^2$$

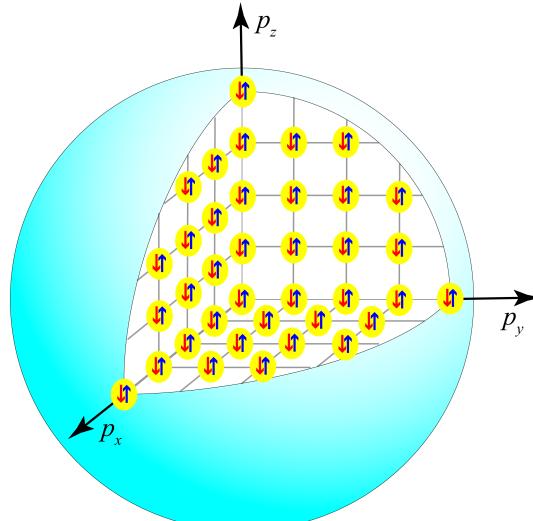


图 20.1: 电子态密度

³电子的微观描述是旋量波,也就是二分量波函数,具有两种可能的自旋.

我们最后做一个经典德鲁特模型与量子费米气模型的比较. 两个理论都承认以下基本公式的成立:

$$\sigma = \frac{ne^2\tau}{m} , \quad \kappa = \frac{1}{3}nv^2\tau c$$

先看电导率, 经典理论与量子理论对各个参数的估计除了 τ 其他都是一致的. 而经典更倾向于对 τ 的估计更小. 因为经典理论一般认为 $\tau = \lambda/v$, 而 λ 一般就按照原子实之间的平均距离来估计. 这一点之后就知道是不妥当的, 因为电子实际上可以在严格周期性的势能场中毫无散射的传播下去. 而使得电子能够被散射的其实是晶格的缺陷, 热振动等因素, 从而一般温度升高, τ 减小, 金属的导电性能就大大减弱. 这给出了金属电阻随温度升高而升高的结论. 这里经典理论对 λ 的过低估计恰好被对 v 的过低估计所抵消掉一部分, 从而最后 τ 的值在常温下差距也不大. 而对于热导率, 经典理论对 v 的过低估计又恰好被对 c 的过高估计修正, 精确到了仅仅相差一个常数, 除了经典理论说不清楚的 τ , 经典与量子的在数量级上是基本符合的.

值得一提, 热导率描述导热, 电导率描述导电, 而导电现象的附效应便是热产生. 也就是焦耳热(Joule heating). 微观的功率密度

$$\mathbf{j} = \sigma \mathbf{E} , \quad w = \mathbf{j} \cdot \mathbf{E}$$

也即:

$$w = \sigma \mathbf{E}^2 = \rho \mathbf{j}^2$$

其中电阻率(resistivity) ρ 为电导率的倒数. 上为微观焦耳-楞次定律(microscopic Joule-Lenz law).

20.1.3 能带论 *

费米气模型在量子力学迅猛发展的几十年后回头看又是过于 naïve 了. 人们发现电子终究是在晶格间传播的波动, 如果原子实与背景的电子给出了严格的周期性的势能, 那么如果电子的总机械能小于最大势能, 那么就成为了束缚态, 限制在一个原子附近运动而成为原子实的一部分. 而如果电子守恒机械能大于最大势能, 就成为不受任何阻碍的传导态. 动能虽在运动中有所波动, 但绝不会越来越小. 描述电子的波函数称为布洛赫波函数(Bloch wave function). 而电子如何去填充怎样的能级? 布洛赫(F. Bloch) 研究并发现了能带论(energy band theory) 来替代旧的费米气理论, 这个理论可以统一地描述导体与绝缘体.

能带论可以看成是电子在单原子外的束缚态特征与传导的自由电子气的特征的一种综合, 一方面电子在这样的一个周期性势场中的运动其守恒的机械能并不是可以取所有值, 而是分为可以连续取值的区间: 能带(band) 与根本取不到的能量区间: 带隙(gap) 所构成. 而电子作为波动用其所在的能带与波矢 k 来描述. 其波矢方向的物理意义不再重要, 因为每一个态一般都代表某种驻波与行波的混合, 而每一个态附近的群速度则可以代表这个态代表的真实电子运动速度. 另一方面这些能带的结构又恰好对应到了单原子的束缚态能级. 这使得我们经常是一望某元素的单原子核外电子排布, 便能从其是否具有最外层电子读出这种元素单质是否能导电的信息.

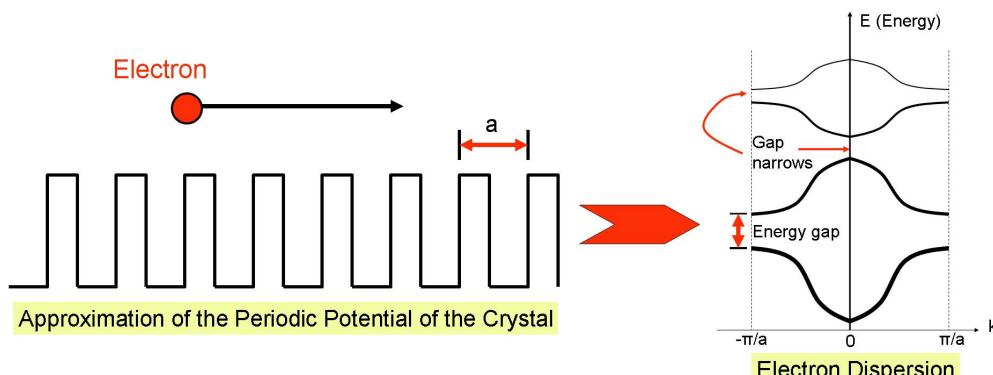


图 20.2: 电子在周期性势场中传播

某种原子形成晶体时,电子从最低的能带开始填充,内层电子填满了较低的那些能带,它们不参与导电,对应于原子实电子和在晶体共价键键合的价电子.最外层电子较少的原子,形成晶体时最后填充的那个能带没有填满,或者虽然填满但与下一个能带在能量区间上重合而能够导电.称为导体(conductor).对应的元素也就显示出明显的金属性⁴.而若原子外层电子数目多,则很有可能电子恰好填满一个能带后隔着一个能隙与空的上方能带相望,则所有态都稳定不变,材料所有电子都是束缚电子,最后填充的那个满带称为价带(valence band),代表原子间成共价键的价电子所在的能带.而上方的那个空带称为导带(conduction gap).意为一旦这个带中存在电子则会大大改善其导电性能.此时元素体现出非金属性,一般形成的都是绝缘体(insulator).而在金属性元素与非金属元素分界处则存在很多的类金属(metalloid),它们形成的晶体虽然与绝缘体一样,在最低能量时恰好把价带排满,而隔着一个带隙与空的导带相望.但带隙很小(小于4eV),导致常温的热激发和掺杂,抑或是光照激发都将在导带中激发大量载流子而具有可观的导电性能.这些材料称为半导体(semiconductor).

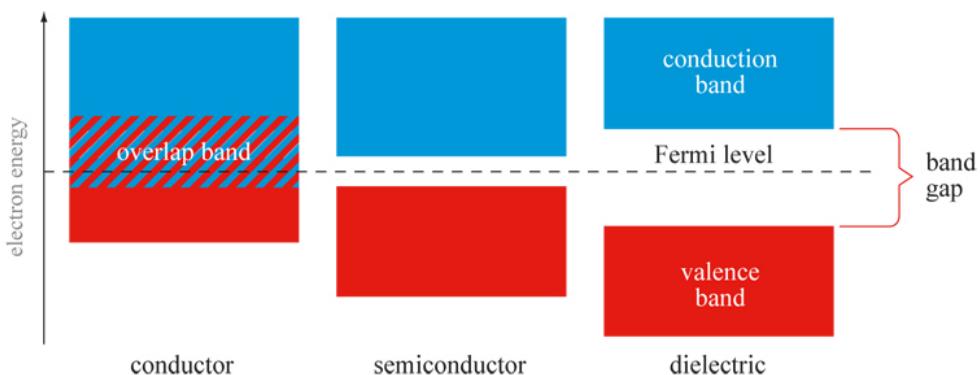


图 20.3: 能带与导电性

1 H														2 He			
3 Li	4 Be																
11 Na	12 Mg																
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Be		72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra		104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn						
57 La	58 Ce	59 Nd	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu			
89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr			

图 20.4: 金属、类金属与非金属元素

那么能带论如何给出电导率的计算公式?仍然是:

$$\sigma = \frac{ne^2\tau}{m}$$

与上一个费米气模型区别在于一点,就是这里的 m 不再是电子的裸质量(bare mass),而是由于与周期

⁴思考:氢是否具有金属性?参考https://en.wikipedia.org/wiki/Metallic_hydrogen.

性势场相互作用, 电子的色散关系(能量与波矢关系)被修正后给出的有效质量(effective mass). 即使对于金属, 这个质量与真实电子质量相差好几倍的现象也是十分普遍的.

20.1.4 惯性, 阻尼与回复力

在量子力学还未诞生的时期, 为了解释导体的导电, 绝缘体的介电现象, 并适用于任意交变电磁场在介质中的传播, 色散, 吸收与散射. 洛伦兹提出著名的**洛伦兹模型**(Lorentz model) 作为德鲁特模型的补充. 在这儿电子的惯性被重视, 电子与晶格的碰撞被简化, 可能的原子实对电荷的束缚被简化为线性回复力. 也就是洛伦兹用唯象的谐振子类比来解释电子在外场下的行为:

$$m\ddot{\mathbf{r}} + \gamma\dot{\mathbf{r}} + k\mathbf{r} = -eEe^{i\omega t}$$

这个模型的解和它对应的各种特性我们将在光学教材里详细讲解. 在这里我们仅仅讨论导体中的自由电子, 故 $k = 0$. 与导体导电相关的两个因素为:

一是阻尼系数 γ . 在直流电场下平衡时稳定的电子速度为:

$$\dot{\mathbf{r}} = -\frac{e}{\gamma}\mathbf{E}$$

对比更加本质的德鲁特等导电模型的电导公式, 容易发现这个唯象系数与电导率和基本参量关系为:

$$\gamma = \frac{ne^2}{\sigma} = \frac{m}{\tau}$$

于是我们可以写出在导体上突然加一个不随时间变化的匀强电场后电子的运动的微分方程:

$$m\ddot{\mathbf{r}} + \frac{m}{\tau}\dot{\mathbf{r}} = -e\mathbf{E}$$

它的解为:

$$\dot{\mathbf{r}} = -\frac{e\mathbf{E}\tau}{m}(1 - e^{-\frac{t}{\tau}})$$

我们十分自然地发现定义电子碰撞的弛豫时间恰好与导体对电场响应的弛豫时间相吻合. 这一点稍微需要一些讨论和修正. 见下.

第二点我们来讨论表示电子惯性的质量 m . 正是它导致了衰减因子 $e^{-\gamma t/m}$, 从而造成电路对电场响应的弛豫. 但如果讨论电路的弛豫, 或者说电流的惯性, 有一点是被我们过低地估计了, 便是自感现象. 电荷的加速运动形成变化的电流, 从而与之相伴的是变化的磁场. 这个变化的磁场反过来给电流一个反向的作用力. 由于该力正比于加速度, 故等价于增加了载流子的质量. 这里的弛豫时间对应直流暂态电路中的 $\tau' = L/R$. 如果一定要理解为电子的惯性质量, 那么一般比电子的裸质量要大好几个数量级 $m^* \gg m$. 从而一般有电路弛豫时间 $\tau = m^*/\gamma$ 不等于微观碰撞弛豫时间.

纯粹的导体是否有回复力项? 显然在以上电子运动方程中不含这一项. 但注意到如果有回复力项, 即使外场为零电子也会做振动. 真实情况会发生这样的现象吗? 答案是会, 即**等离子体振荡**(plasma oscillation). 原因是电子相对于原子实的位移实际上在金属的表面累积了电荷分布, 从而在内部激发了电场. 由于频率很高, 一般要到紫外波段的频率, 我们可以完全忽略阻尼的影响. 当电子有位移 \mathbf{r} 时将形成极化强度⁵:

$$\mathbf{P} = -ner$$

我们考虑块状金属板⁶的集体等离子体振荡, 那么由于金属表面的累积电荷在金属内部形成的电场为:

$$\mathbf{E} = \mathbf{E}_{in} = -\frac{\mathbf{P}}{\epsilon_0} = \frac{ner}{\epsilon_0}$$

⁵显然此时“极化强度”描述的非极化电荷, 而是自由电荷.

⁶的确, 取不同形状的物体这个频率会有所不同.

代入原方程:

$$m^* \ddot{\mathbf{r}} = -e \mathbf{E} = -\frac{ne^2 \mathbf{r}}{\varepsilon_0}$$

从而得到谐振子方程. 振荡频率为:

$$\omega_p = \sqrt{\frac{ne^2}{\varepsilon_0 m^*}} ; \quad \sigma = \varepsilon_0 \omega_p^2 \tau$$

最后我们考虑低频电磁波(内部实际场强)下金属的电子行为, 利用光学中讨论过的洛伦兹模型, 我们计算金属的复电容率, 最后得到:

$$\varepsilon_r(\omega) = 1 - \frac{\omega_p^2 \tau^2}{1 + \omega^2 \tau^2} ; \quad \sigma(\omega) = \frac{\sigma}{1 + \omega^2 \tau^2}$$

一般 $\omega_p \tau \gg 1$, 故低频下金属的介电常数实际上是绝对值很大的负数. 电导率随着频率的增加而减小.

20.1.5 稳恒电流与形成条件

作为电荷的流动, 电流密度与电荷密度间满足电荷守恒定律:

$$\nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = 0$$

而所谓**稳恒电流**(steady current) 指的是一定区域内导电物质形成的电荷, 电场, 电流分布. 其中电流的分布不能随时间变化, 自然的结论是电荷密度分布也不能随时间变化. 否则将产生随时间变化的电场, 电场造成电流的变化. 即上式必有:

$$\frac{\partial \rho}{\partial t} = 0 ; \quad \nabla \cdot \mathbf{j} = 0$$

即在稳恒电流中电流密度是一定没有散度的, 无头无尾的闭合曲线. 稳恒电流必须是**直流**(direct current) 而非**交流**(alternating current).

容易发现, 在稳恒电流中微观欧姆定律写成以下形式是不完整的:

$$\mathbf{j} = \sigma \mathbf{E}$$

在很多情况下, 上式并没有问题. \mathbf{E} 被理解为由电荷分布 ρ 形成的静电场⁷, 它是一定可以引入电势的:

$$\mathbf{E} = -\nabla \varphi$$

那么如果承认以上形式的微观欧姆定律就会引起矛盾: 电流线是环状闭合的, 电场线必须与电流线同向平行, 但电场线又不能是闭合的, 否则与静电场环路定理, 电势的可定义性矛盾. 实际上, 微观形式的欧姆定律必须被写成:

$$\mathbf{j} = \sigma(\mathbf{E} + \mathbf{K})$$

上式 \mathbf{K} 代表**非静电力场**(non-electrostatic field) 对电子产生的作用. 定义方式与电场量纲一致, 都是单位正电荷的受力. 这么写的不同之处就在于, 把驱动电流的力分成了两部分: 静电场 \mathbf{E} 是保守的, 可以引入势的, 在一个回路中不做功的, 不吞吐能量的场(注意电流本身就会发热, 这不是场造成的而是电荷定向流动碰撞晶格造成的). 但非静电力就是源

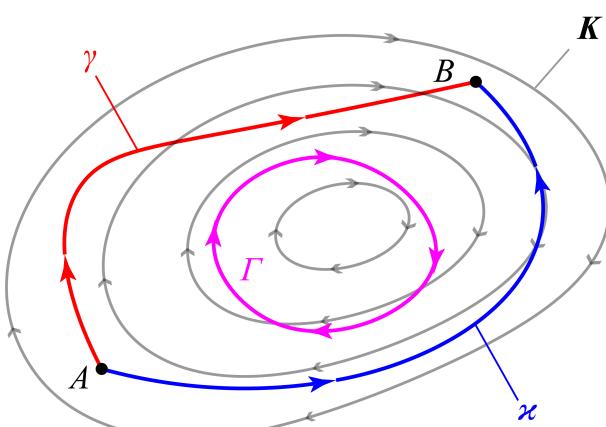


图 20.5: 非静电力场

⁷实际上, 不包括涡旋电场, 因为它被归于之后引入的非静电力. 换句话说, 这里的 \mathbf{E} 不再由静止电荷的受力 \mathbf{F}/q 定义, 而是单纯根据静止电荷分布产生平方反比的电场叠加去定义.

源不断做功驱动回路中电流流动的,非保守的,不能引入势来描述的场.虽然不能引入电势来描述非保守力,但可以用**电动势**(electromotive force) 来描述非静电力:

$$\mathcal{E}_\gamma = \int_{A \xrightarrow{\gamma} B} \mathbf{K} \cdot d\mathbf{r} ; \quad \mathcal{E}_\Gamma = \oint_{\Gamma} \mathbf{K} \cdot d\mathbf{r}$$

不像对静电场的积分那样仅仅取决于端点,电动势与积分路径一一对应.回路电动势一般来说不为零,从而相同端点的两个路径电动势也不一定相等:

$$\mathcal{E}_\Gamma \neq 0 ; \quad \mathcal{E}_\gamma \neq \mathcal{E}_x$$

非静电力既然能够作用在电子上,其形式一般就是电磁力,万有引力,或是惯性力.磁场力对应**动生电动势**(motional emf),而由于变化的磁场激发的涡旋电场部分的电场力对应**感生电动势**(transformer emf).电子会受到引力,1967 年 Schiff 指出电子会因为引力的原因聚集到金属的底部.而后 Dessler 提出了严格的考虑晶格在重力场下压缩与不均匀情况下的电子平衡情况.引力是很弱的保守力,对金属内部电荷体系的影响虽然很小但绝不是不可观测.而与引力类似的惯性力则可以人工制造出很大的数值,而且可以使它非保守.试想加速旋转一个线圈,那么在随晶体加速的参考系里电子受到的角向惯性力就和受到一个电场力造成的效果没有本质的区别.

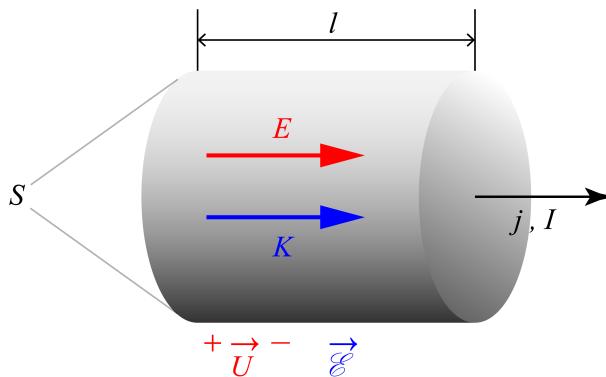


图 20.6: 宏观欧姆定律

还有一些电动势一般无法用非静电力描述.因为在每一个回路中某些部分的确有能量的输入,但不是以作用在某些具体的电荷上的力的形式.而是更加广义的一些力作用在一些流上.例如温差电偶利用两种物质的电子的热扩散性质差异,从高温端吸取热量而驱动电流.化学反应更是源源不断地产生新的物质,利用反应的自发性而驱动物质做定向的移动形成电流.在这些抽象的场合,我们采用某段电路中移动单位电荷,外界向体系注入的能量来定义电动势:

$$\mathcal{E} = \frac{dW}{dQ}$$

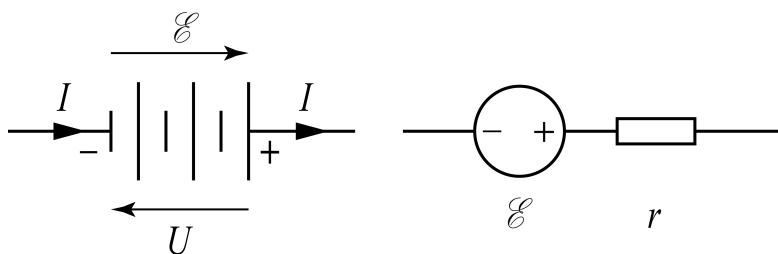


图 20.7: 电池的欧姆定律

最后,从微观到宏观,我们考虑一段长为 l ,面积为 S ,电阻率为 ρ 的导体上的欧姆定律:

$$E + K = \rho j$$

等式两边同时乘以 l ,把电流密度写成 I/S .得到:

$$U + \mathcal{E} = IR ; \quad R = \rho \frac{l}{S}$$

一般叫做部分电路的欧姆定律.而如果考虑**电池**(battery) 的两端,正常工作时一般正极电压更高,故约定电动势与电压取相反的方向,电流则与电动势方向一致,却与电压方向反向:

$$U = \mathcal{E} - Ir$$

为一段含源电路的欧姆定律. 此时我们把电池等效为理想电压源和一个定值电阻 r 的串联. 电阻即为电池的内阻(internal resistance).

20.2 电路与电路方程

具体到直流电路中, 我们一般提供以下理想的电路元件:

名称	电阻	恒压源	恒流源	接地
符号	R	$\begin{array}{c} \oplus \\ \ominus \end{array}$	$\begin{array}{c} \oplus \\ \ominus \end{array}$	$\begin{array}{c} \parallel \\ \perp \end{array}$

图 20.8: 常见电路元件

实际电路总是由这些元件和导线构成, 导线和元件的内阻均被用实际电阻显式表示. 接地符号可以被视作特殊的导线. 一般与真实的“地面”没有任何联系, 电路中一般“地”指的是公共的零电势参考点. 电路中如果仅仅有一个接地符号, 就把接地的点的电势认为是零. 如果出现了多个接地符号, 就用导线把这些点连接到同一个点, 并把这个点电势视作零.

任何一个元件, 或者二端的部分电路, 为了研究的统一性与方便, 我们规定两端的电压的正方向与流过的电流的正方向都取“相关方向”, 沿着这个方向, 元件或部分电路表现的像一个电阻, 或者宽泛一点说, 像一个用电器. 即电流就是沿这个方向流动, 而且电压也沿这个方向降低. 比如在如图20.9所示的左侧大电池为右侧小电池充电的过程中, 右侧电路对上下两端口以向下为电压电流的相关参考方向, 则电压和电流分别为:

$$U = +2.8V, \quad I = +0.2A$$

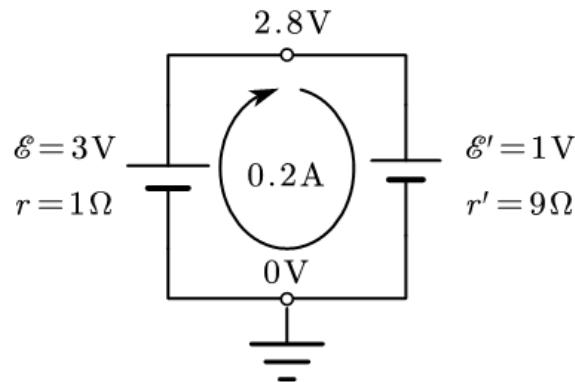


图 20.9: 相关方向

但如果对左侧电路依然取向下为电压电流的相关参考方向, 则电压和电流分别为:

$$U = +2.8V, \quad I = -0.2A$$

可以发现, 如果一个元件或部分电路按照相关方向计算的电压电流同时是正或者负, 这个元件就是类用电器的, 整体效果是消耗电能, 而如果一正一负, 就是类电源的, 整体效果是产生电能.

在具体分析复杂电路时, 顺次串联的各个部分电路相关方向一般取为同向; 讨论单个回路时, 各个部分的相关方向也是按照某种环绕方向确定的; 多个二端电路并联在两点之间也通常取各个支路为统一约定的从一个点到另一个点的相关方向.

对于之前的三种元件, 电阻由于对称性没有天生的电压电流相关方向. 无论如何取相关方向, 其电压 U , 电流 I 和其参数电阻 R 都一定要满足下式, 称作伏安特性(V-A characteristic):

$$U = IR$$

恒压源, 顾名思义, 总是能保持其正极电压比负极电压高 \mathcal{E} . 在正常工作状态时, 电流应当从正极流出, 使得其成为类电源的元件. 取此时电流方向为其相关方向 (具体问题中相关方向也可以与之相反), 则其特性

为:

$$U = -\mathcal{E}, \quad I > 0$$

理想的恒压源同样能够工作在 $I < 0$ 的类用电器区域, 此时同样有 $U = -\mathcal{E}$. 但是对理想电压源短路是没有意义的, 一般恒压源所在支路一定需要串联电阻.

恒流源, 顾名思义, 总是能保持从负极向正极输出电流 \mathcal{I} . 在正常工作状态时, 正极电压应当比负极电压高, 使得其成为类电源的元件. 取此时电流方向为其相关方向 (具体问题中相关方向也可以与之相反), 则其特性为:

$$I = \mathcal{I}, \quad U < 0$$

理想的恒流源同样能够工作在 $U > 0$ 的类用电器区域, 此时同样有 $I = \mathcal{I}$. 但是对理想电压源开路是没有意义的, 一般恒流源所在支路一定需要并联电阻.

所以我们考虑的电路中, 如果出现恒压源或恒流源, 一般以以下带分压内阻或分流内阻的形式存在, 如果考虑整个元件的特性, 就给出了著名的输出特性(output characteristic):

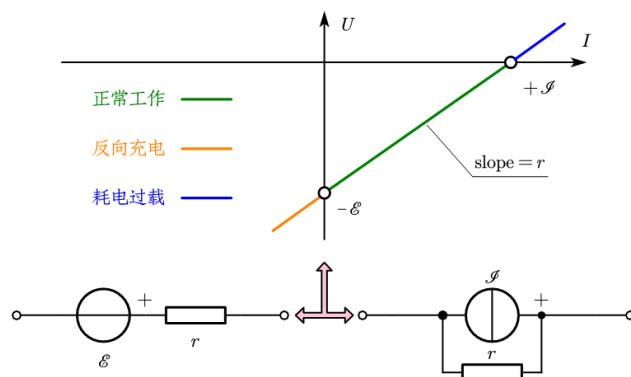


图 20.10: 源的统一输出特性

$$U = -\mathcal{E} + Ir = (I - \mathcal{I})r$$

我们也从中发现了, 如果恒压源与恒流源满足内阻相等都是 r 且 $\mathcal{E} = \mathcal{I}r$, 那么它们的输出特性是完全一致的. 也就从功能上等价且可以相互替代, 这是我们之后小节要介绍的戴维南-诺尔顿定理的一个预兆. 但这主要只有理论价值, 实际情况分压内电阻总是要求很小, 分流内电阻总是要求很大, 这样在较大输出范围内才能凸显其恒压恒流的特性.

我们只有三种元件 (电阻, 带内阻的恒压源, 带内阻的恒流源), 而且已经成功给出了三种元件的特性, 它们实际上就是元件上取相关方向以后电压与电流之间的函数关系, 从而只要知道电流或电压中的一者就可以表示出另一者来. 但是如果涉及到电路的求解, 就需要把不同的元件联系起来. 此时还会涉及到两类方程. 分别称作基尔霍夫电流定律(KCL, Kirchhoff's current law) 与基尔霍夫电压定律(KVL, Kirchhoff's voltage law).

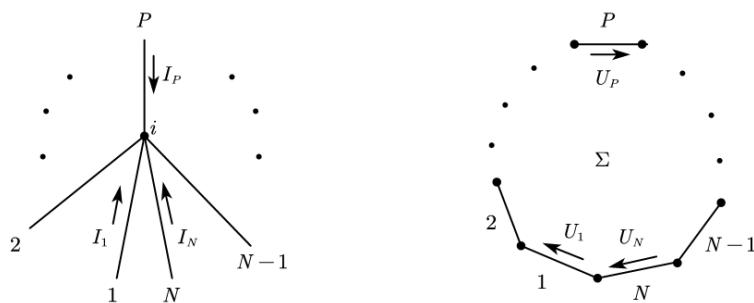


图 20.11: KCL 与 KVL

KCL 的含义是,如果多个元件的一端连在一点.那么取指向或背离公共节点的方向为所有元件的相关方向,必有:

$$\sum_{P \in i} I_P = 0$$

我们约定小写拉丁字母 $i, j, k \dots$ 用来表示点,大写拉丁字母 $P, Q, R \dots$ 用来表示元件所在的支路(边),而从属符号 \in 用来表示结合关系, $P \in i$ 念做“边 P 以 i 为端点”.KCL 因此也称作**节点电流定律**.

KVL 的含义是,如果多个元件的首尾相连构成回路.那么取顺时针或逆时针方向为所有元件的相关方向,必有:

$$\sum_{P \in \Sigma} U_P = 0$$

我们约定大写希腊字母 $\Sigma, \Pi, \Phi \dots$ 用来表示回路,同样用从属符号 \in 用来表示结合关系, $P \in \Sigma$ 念做“边 P 在回路 Σ 上”.KVL 因此也称作**回路电压定律**.

而需要注意的是,KCL 与 KVL 都可以以元件电流或电压作为自变量.比如用电流做为 KVL 方程的自变量就构成了所谓的数电压法.下图所示的电路回路使用数电压法列顺时针方向的 KVL 方程就如右式:

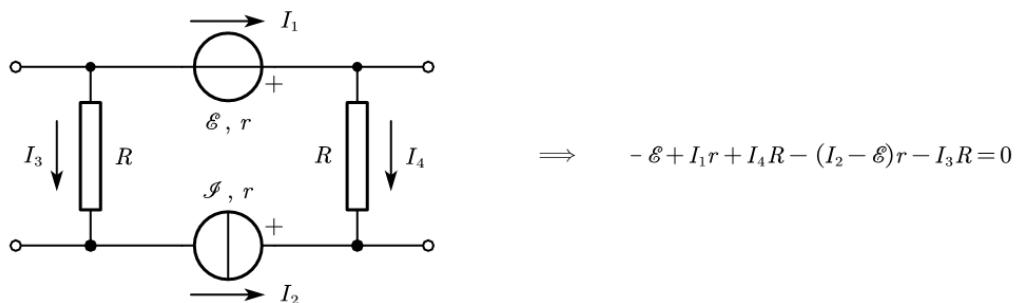


图 20.12: 回路: 数电压法

而节点电流也可以使用电压作为变量.另外还需要注意,利用数电压法我们可以得到 KVL 方程的一种非完整形式.由于我们对于节点总是喜欢定义其电势,那么如果节点 i, j 之间存在一条长的支路链,如下图所示,那么取 $i \rightarrow j$ 方向为参考方向,电势降低就满足:

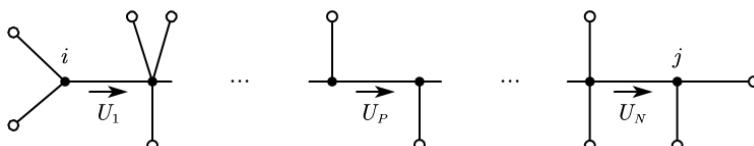


图 20.13: 路径: 数电压法

$$\varphi_i - \varphi_j = \sum_P U_p$$

尤其是对于以后用得上的星型电阻网络问题,以三端为例, KCL 或节点电流定律就可以在此时列为:

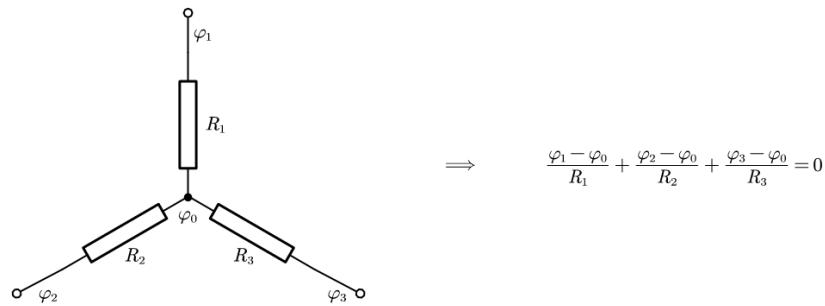


图 20.14: 节点电压法使用 KCL

一个典型的电路问题提法是如何?它又具有哪些元素与性质?解决方案又有哪些?我们下面从基础开始细细分析.

20.3 电路分析基础

20.3.1 电路的整体结构与拓扑学结论

电路分析中的以下三个重要概念的术语与立体几何学中的术语是可等值替换的:

- 支路 (branch)-边 (edge)
- 节点 (node)-顶点 (vertex)
- 回路⁸ (circuit)-面 (face)

它们分别是什么含义?

最基础的概念莫过于支路. 我们暂时采取最保守的做法, 认为每一个元件都占据一个独立的支路, 不同的支路之间可以是串联, 并联, 或者组合在一起的混联, 或者什么也不是 (典型的情况比如五条支路构成的桥式连接).

接下来定义节点的概念: 节点把支路连接在一起. 在定义这个概念时我们首次遇到了拓扑等价 (topological equivalence) 的概念. 众所周知, 如果是用真实导线把真实的元件连接, 那么连接方式不一定与电路图是完全对应的. 事实上就电路图本身也具有一定的变化范围, 比如对连接点的位移与导线的弯折具有等价性, 如下图左. 再比如说如果只允许电路图中出现三导线相交的情况时, 把四个元件的一端相连就有如下图右的三种方式, 当然, 如果在中间红色导线上接上其他元件或是用理想电流表去测量其电流, 那么这三个电路并不等价, 但是如果只考虑下面要阐述的电路问题的标准提法的求解那么三个电路并无区别.

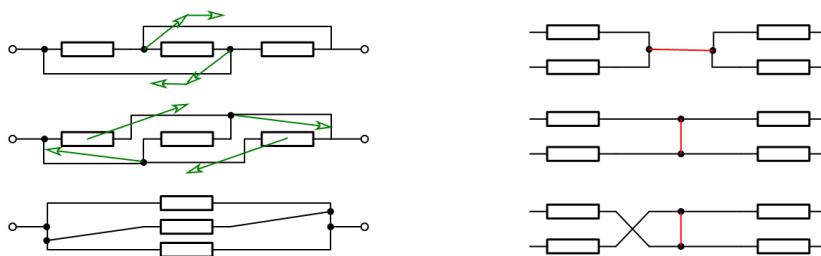


图 20.15: 拓扑等价变形

⁸注意, 真正与面形成对应关系的应该是之后要引入的网孔的概念

既然电路可以被拓扑变形,那么节点的概念也就要被重新被考量了.事实上,节点相当于彼此等势的导线延伸体.导线无论如何连接与延伸,只要是彼此连通的点就被视作属于同一个节点.从这个视角上来看,上面的左图从一开始实际上就只有左右两个节点,右边的三个图中间全部都归为一个节点.

但是,出于几何上的直观,如果一个节点只被两个元件一端共用,那么这两个元件就称作串联(series connection)的关系,两个元件的串联或多个元件的顺次串联的情况下,我们有理由把这一串元件全都视作在一条“强支路”上,原来的支路被合并,原来的中间的节点也就随着取消,剩下的节点至少应该是三条或三条以上的支路汇合的“强节点”.按照这种方式理解的强支路与强节点赋予电路以类似于多面体的拓扑结构.

要使得电路的结构完全类似于多面体,还缺少至关重要的一环: 电路的回路就类似于多面体的面. 电路的回路是由多段支路首尾相连构成的不相交的环. 为了说明清楚回路分析中的一个重要因素: 回路基. 我们还需要引入可平面化图(planar graph)的概念. 我们知道, 图是用来表示事物的连接关系的重要数学工具, 在这个意义上就无所谓图所在空间的维数的说法. 比如以右图表示的著名的三对三图(three for three graph)中, 有三户家庭分别都需要从水厂, 气厂, 电厂输送过来的资源, 那么三户家庭与三个工厂看作节点, 输送管道看作支路, 这就构成了 $K_{3,3}$ 图. 当然任何维度空间中的网络, 多面体的边与顶点自然也构成了一些图. 但是

原则上抽象为一个图以后就失去了任何抽象前的维度信息. 这样就可以定义平面图(plane graph): 它是被画在一个平面上的, 同时支路与支路没有相交的图. 右边画的 $K_{3,3}$ 图就不是一个平面图. 事实上, 它也不是一个可平面化图: 可平面化图指的是虽然这个图可能以某种方式画在平面上支路会相交, 但是总能找到使边不相交的办法. 而著名的库拉托夫斯基定理⁹(Kuratowski's theorem) 指出:

任何不可平面化的图, 将串联支路合并后, 必然包含一个五阶完全图¹⁰或三对三图为子图.

我们下面只研究平面图, 但除了网孔概念以外, 其余性质都是所有电路中的图共有的.

在平面图中网孔(mesh)就是一个比较自然的概念了: 平面图的每一个格子就是一个网孔. 比如下图所示的三维立方体就产生了一个八节点十二支路的图. 它是可以平面化的, 按照右边的方式平面化以后, 就会导致变成 5 个网孔的平面图.

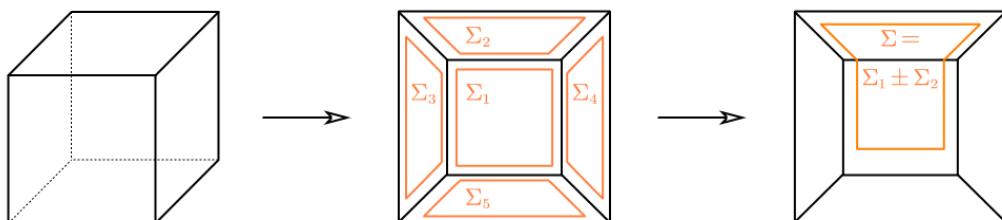


图 20.17: 立方体的平面化与网孔合并为复杂回路

可以发现, 网孔不过是特殊的一组回路. 这些回路内部的平面区域已经不再有其他支路. 如果网孔的个数为 F , 那么整个二维平面就被这些网孔回路分为 $F + 1$ 个区域, 每个网孔内部是一个区域, 而所有网孔外还有一个大的区域.

网孔作为一组特殊回路的特殊之处就在于, 网孔是一组自然的回路基(circuit basis). 所谓回路独立就是指它们对应的回路电压方程独立, 所谓回路基就是指所有可能回路的回路电压方程中选取的一组可以用来

⁹直到 1930 年才被严格表述与证明.

¹⁰即五个点彼此连接

表示¹¹所有其他方程的独立方程对应的独立回路. 可以想见, 回路的取法是远多于网孔的个数的, 比如可以把上图两个相邻的网孔 Σ_1, Σ_2 合并为较大的回路 Σ^{12} . 所以回路基的取法并不是唯一的, 原则上存在很多种独立回路的取法.

图论给出, 对于任何一个电路图, 都存在一个特征的数字: 回路秩¹³(circuit rank), 对于平面图它就是网孔的个数, 对于非平面图可以用之后给出的办法计算. 我们取一组最大数目的互相独立的回路做回路基时, 其个数总是会等于这个回路秩. 利用这个特性, 即使是复杂的非平面电阻网络, 只要我们总是避免选取的回路不独立, 选够回路秩那么多回路时, 就得到了一组回路基.

联系支路, 节点, 网孔(可换为回路基, 下同)三者的最核心的定理, 莫过于欧拉早在 1758 年就发现并证明的定理: 如果一个三维空间中的没有洞的多面体(比如像甜甜圈那样, 但是表面被磨为多个平面的多面体就是有洞的多面体)的顶点数为 V , 边数为 E , 而面数为 F , 那么必然有:

$$E = V + F - 2$$

但是如果变成电路问题, 由于多面体的面数与它所对应的被平面化的图的网孔数(或者图的回路秩, 下同)总是差一. 故对于电路问题, 节点数 V , 支路数 E 和网孔数 F 三者满足的关系式应当为:

$$E = V + F - 1$$

默认以上结果, 就意味着我们终于具备了开始研究电路问题的数学基础. 接下的第一步, 就是严格的表述一个电路问题.

- 已知量:

- 首先, 必须给定的是电路的拓扑, 即已知电路由 V 个节点, E 条支路构成, 还应当能判断每一个节点 i 与每一条支路 P 之间是否有从属关系 $i \in, \notin P$.
- 其次, 每一条支路上的元件种类: 电阻, 恒压源或恒流源. 与对应的参数 $R, \mathcal{E}, \mathcal{I}, r$ 也应当事先给定.

- 待求量:

- 为每一条支路指定好方向以后, 支路上的电流和电压都是待求的. 由于电压电流通过支路元件特性相关联, 一共 E 个未知数.
- 节点上的电势通常也是我们关心的, 此时一般要约定好某节点接地视作零电势, 其他节点的电势均是相对它的电势, 都是待求量, 一共 $V - 1$ 个未知数.
- 平面图形成了 F 个网孔. 图论可以证明, 因为满足 KCL 方程, 合理的某种支路电流分布总存在唯一的某种网孔电流(mesh current)的叠加与之等价. 故也会产生对应的网孔电流待求量, 一共 F 个未知数.

¹¹通过方程的线性组合

¹²从集合论的角度来讲, 这种合并的本质是对称差(symmetric difference), 它被定义为:

$$A \Delta B = A \cup B - A \cap B$$

对于回路的合并这是很自然的: 上式代表保留各自的边, 去除公共边. 但是对于回路合并对称差符号写为 \pm 符号是非常贴切的, 因为容易理解对称差两次必然相互抵消:

$$(A \Delta B) \Delta B = A$$

从而从合并后的网孔去再次对称差合并前的网孔, 又会回到合并的另一个网孔, 像极了加减法:

$$\Sigma_1 \pm \Sigma_2 = \Sigma \Leftrightarrow \Sigma \pm \Sigma_2 = \Sigma_1$$

¹³最早被基尔霍夫(G. Kirchhoff)在 1847 年引入, 当时称作圈数(cyclomatic number).

网孔(回路)电流法是一种用来表示电路中的电流分布的快捷方法,其优点是自动会满足KCL方程且未知数个数相对较少。其做法是:对于每个网孔 Σ 约定一个顺时针或逆时针的参考方向,并给定一个网孔电流 I_Σ 。那么对于每一条支路 P ,再为支路约定参考方向,则其上的电流等于:

$$I_P = \sum_{\Sigma \in P} \pm I_\Sigma$$

再一次地 \in 表示结合关系, $\Sigma \in P$ 念做“ Σ 以 P 为边”。需要注意,如果 Σ 就是取网孔,那么与任何一条支路发生结合关系的网孔最多只能有两个,即上式最多只有两项,而在图的边缘的支路则仅仅属于一个网孔。但是,如果取更广义地取回路基,则与同一条支路发生结合的回路个数可以是任意个。电流的符号需要根据回路电流的参考方向与后面对支路求和时的方向是否相同来决定,同向取正,反向取负。下图对同一个平面电路图,用网孔电流法和回路电流法分别表示两条代表性支路上的电流:

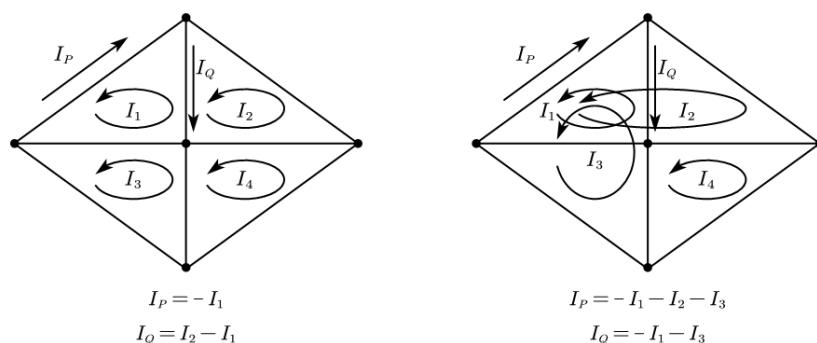


图 20.18: 网孔(回路)电流法

最后,通过已知量求未知量的方法,应当是各类KCL和KVL方程。它们应当如何列,列多少个?这就形成了电路问题的求解套路:

20.3.2 电路问题的求解套路

电路问题的求解有三种套路:支路电流法,节点电势法和网孔(回路)电流法。无论采用哪一种方法,无非是对之前梳理的三类待求量中,选取一类集中加以解决:解决的方案无非是设它们为未知数,并找到与未知数的个数恰好相等的,包含且仅包含这些未知数的KCL或KVL方程,这必然得到线性方程组。解出它们并根据它们得到另外两类待求量。下面逐一论述之。

1. 支路电流法

如果为每一条支路约定好参考方向,并设其上电流为 I_P ,则一共有 E 个未知数。

我们需要列出几乎全套的KCL,对每一个节点*i*,这个方程写作:

$$\sum_{P \in i} \pm I_P = 0$$

正负号约定是:如果支路电流参考方向为指向该节点,则取正,若背离该节点,则取负。但是所有的KCL放在一起并不独立,这是因为所有的KCL方程求和后,每一条支路的 I_P 恰好在两端点处的KCL中出现两次,一正一负,故最后得到零等于零。从而需要去掉一个任意节点*j*,即可以列出 $V - 1$ 个方程:

$$\forall i \neq j, \quad \sum_{P \in i} \pm I_P = 0$$

对于每一个网孔还要列一个KVL方程,为了方便与统一见,我们把所有带内阻的恒流源都变成等效的带内阻恒压源,而电阻元件也视作 $\mathcal{E} = 0$ 的特殊情况,下同。这样KVL方程中的电压就可以用电流表

示,一共是 F 个方程:

$$\forall \Sigma, \sum_{P \in \Sigma} \pm U_P = 0$$

$$U_P = -\mathcal{E}_P + I_P R_P$$

正负号约定是,如果约定回路的环绕方向与支路参考方向一致则取正,相反则取负.

整理,对以上 KVL 方程把未知数置于左侧,常数(非齐次项)置于右侧,得到:

$$\forall i \neq j, \sum_{P \in i} \pm I_P = 0$$

$$\forall \Sigma, \sum_{P \in \Sigma} \pm R_P I_P = \sum_{P \in \Sigma} \pm \mathcal{E}_P$$

一共是 $V - 1 + F$ 个方程,恰好等于未知数的个数 E ,因为:

$$E = V + F - 1$$

2. 节点电势法

如果为某个节点 j 接地,即约定其电势为 $\varphi_j = 0$,对其他节点 $i \neq j$ 设电势 φ_i ,则一共有 $V - 1$ 个未知数.此时 KVL 方程已经给不出有意义的方程,只能列 KCL 方程.任何一条支路 P ,如果两端点为 i, k ,那么如果约定 i 到 k 方向为其参考方向,有:

$$U_P = \varphi_i - \varphi_k = -\mathcal{E}_P + I_P R_P \Rightarrow I_P = \frac{\varphi_i}{R_P} - \frac{\varphi_k}{R_P} + \frac{\mathcal{E}_P}{R_P}$$

那么根据 KCL 方程,所有流入 k 节点的电流应当和为零:

$$\left(\sum_{P \in k} \frac{1}{R_P} \right) \varphi_k - \sum_{i \in P \in k} \left(\frac{1}{R_P} \varphi_i \right) = \sum_{P \in k} \frac{\mathcal{E}_P}{R_P}$$

其中 $i \in P \in k$ 念做 $i, j P$.同样的,这 V 个方程并不独立,应当再去掉一个 k 的方程,由于 $\varphi_j = 0$ 不必解,不妨就去掉这一个,并使左侧不出现 φ_j :

$$\forall k \neq j, \left(\sum_{P \in k} \frac{1}{R_P} \right) \varphi_k - \sum_{\substack{i \in P \in k \\ i \neq j}} \left(\frac{1}{R_P} \varphi_i \right) = \sum_{P \in k} \frac{\mathcal{E}_P}{R_P}$$

这样方程数也是 $V - 1$ 个,等于未知数个数 V_1 .

3. 网孔电流法

如果对每个网孔,约定参考方向后设每个网孔电流为 I_Σ ,则一共有 F 个未知数.此时 KCL 方程已经给不出有意义的方程,只能列 KVL 方程.任何一条支路 $P \in \Sigma$,如果按照所在网孔 Σ 的参考方向,有:

$$U_P = -\mathcal{E}_P + I_P R_P = -\mathcal{E}_P + R_P \sum_{\Pi \in P} \pm I_\Pi$$

求和中恰好有一项就是原来的 I_Σ ,它取正号,而其他的回路 Π 应当根据与回路 Σ 在公共边 P 上的方向是否一致来决定正负号,同向为正,反向为负.将上式求和得到 KVL:

$$\forall \Sigma, \left(\sum_{P \in \Sigma} R_P \right) I_\Sigma + \sum_{P \in \Sigma} \left(R_P \sum_{\Pi \in P \in \Sigma} \pm I_\Pi \right) = \sum_{P \in \Sigma} \mathcal{E}_P$$

其中 $\Pi \in P \in \Sigma$ 念做“ Π 和 Σ 共用边 P ”.这样方程数也是 F 个,等于未知数个数 F .

一方面,三种方法的未知数与方程个数分别是 $E, V - 1, F$,而 $E = V - 1 + F$,所以显然后两种方法的未知数和方程数都是比第一种少的,在实际应用过程中更加省事,原则上我们应当尽可能地使用后两种方法来解决实际问题.但是另一方面,第一种方法得到的方程形式上更简洁而易于分析,所以理论推导的时候也青睐基础的支路电流法.

20.4 电路分析方法

- 叠加原理: 用上一节介绍的任何一种方法求解方程组时, 右边的非齐次项都可以进行拆分与叠加. 直接推论是, 如果电路只含电压源, 那么保留各个单个电压源, 其他电压源置零, 解出的各支路电流的和即每个电压源都存在时的电流.
- 任何二端不含源电阻网络都可以等效为一个线性电阻.
- 戴维南-诺尔顿原理: 任何二端含源网络如果把各电压源电流源置零之后形成的不含源电阻网络的等效电阻网络电阻为 r , 其开路电压 U , 其短路电流 I . 那么这个电路等效于一个电压源或电流源, 内阻即 r , 电压或电流即 U, I .
- 特勒根原理: 单一电路中若各个支路选取相关方向, 则:

$$\sum_P U_P I_P = 0$$

如果两个电路的图一样, 即具有相同的节点, 支路, 网孔结构, 但是元件可以不一样. 那么选取各个支路为相关方向, 也有:

$$\sum_P U_{1P} I_{2P} = 0 = \sum_P U_{2P} I_{1P}$$

第二十一章 静磁学

21.1 电流与磁场

21.1.1 磁场地位与电流分布

人们对磁的认识一共经历了几个阶段:

1. 起: 远古-**奥斯特**(H. C. Ørsted). 磁现象被认为与**磁体**(magnet) 相关. 它被设想为由**磁荷**(magnetic charge) 造成.
2. 承: 奥斯特-**麦克斯韦**(J. C. Maxwell). 奥斯特发现了电流的磁效应, **法拉第**(M. Faraday) 建立了“场”的观念, 可以定量描述电流产生的磁场, 法拉第与**亨利**(J. Henry) 等人发现了电磁感应, **安培**(A.-M. Ampère) 的**分子电流假说**解释了物质磁性的本质也是电流产生磁场.
3. 转: 麦克斯韦-**狄拉克**(P. Dirac). 麦克斯韦和**洛伦兹**(H. A. Lorentz) 给出了电磁场的普遍理论, 即**电动力学**(electrodynamics)¹, 洛伦兹和**爱因斯坦**(A. Einstein) 等人将电动力学纳入相对论框架. **赫兹**(H. R. Hertz) 的实验有效地支持了电磁场论. 洛伦兹等人的经典电子论是微观粒子与电磁场之间相互作用的第一个完备理论, 微观粒子被陆续发现, 电子的**自旋**(spin) 与**磁矩**(magnetic moment) 被发现与研究.
4. 合: 狄拉克-现今. 狄拉克主张把电子量子化为场, 提出了**量子电动力学**(quantum electrodynamics), 预言了反物质的存在, **安德森**(C. D. Anderson) 实验室发现正电子. 狄拉克提出了量子化的**磁单极子**(magnetic monopole) 模型, 但至今未找到磁单极子.

现在, 我们对**磁场**(magnetic field) 概念的理解就可以用下图来概括:

如何证明磁场是电场的相对论效应? 设想原参考系中只有电场, 那么无论原参考系中的速度 v_x, v_y, v_z 如何, 受力都不取决于速度大小, 它是:

$$F_x = qE_x$$

$$F_y = qE_y$$

$$F_z = qE_z$$

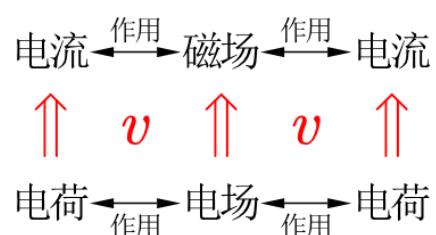


图 21.1: 磁场与电荷的关系

根据相对论的力变换公式:

$$F'_x = F_x - \frac{u}{c^2} \cdot \frac{v_y F_y + v_z F_z}{1 - \frac{u v_x}{c^2}} , \quad F'_y = \frac{F_y}{\gamma(1 - \frac{u v_x}{c^2})} , \quad F'_z = \frac{F_z}{\gamma(1 - \frac{u v_x}{c^2})}$$

代入原式, 并假设原参考系速度和变换速度相对 c 都是小量, 故新参考系速度可以使用经典的速度叠加:

$$v_x = v'_x + u , \quad v_y = v'_y , \quad v_z = v'_z$$

¹这个词语却是安培发明的.

近似到与 v'_x, v'_y, v'_z 一阶相关的领头项:

$$F'_x = qE_x - qv'_y \cdot \frac{uE_y}{c^2} - qv'_z \cdot \frac{uE_z}{c^2}$$

$$F'_y = qE_y + qv'_x \cdot \frac{uE_y}{c^2}$$

$$F'_z = qE_z + qv'_x \cdot \frac{uE_z}{c^2}$$

我们如果假设以下**洛伦兹力**(Lorentz force) 公式的成立性:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

它是磁场力分量 $\mathbf{F}_m = q\mathbf{v} \times \mathbf{B}$ 与电场力分量 $\mathbf{F}_e = q\mathbf{E}$ 的和. 则受力公式简化为:

$$F'_x = qE'_x + qv'_y B'_z - qv'_z B'_y$$

$$F'_y = qE'_y + qv'_z B'_x - qv'_x B'_z$$

$$F'_z = qE'_z + qv'_x B'_y - qv'_y B'_x$$

对比即知道:

$$B'_x = 0 \quad , \quad B'_y = \frac{uE_z}{c^2} \quad , \quad B'_z = -\frac{uE_x}{c^2}$$

这实际上:

$$\mathbf{B} = -\frac{\mathbf{u} \times \mathbf{E}}{c^2}$$

如果原参考系是静止点电荷产生的电场, 现在换参考系以后, 点电荷做速度为 $-\mathbf{u}$ 的匀速直线运动而产生以上磁场, 那么就能推理出, 如果点电荷以速度 \mathbf{v} 做匀速直线运动, 那么它产生的磁场必然为:

$$\mathbf{B} = \frac{1}{4\pi\epsilon_0 c^2} \frac{q\mathbf{v} \times \mathbf{e}_r}{r^2}$$

我们从中总结出:

- 只有运动的电荷才会受到磁场力.
- 只有运动的电荷才产生磁场.
- 但是运动与静止具有相对性, 所以电场和磁场的区别就像静止与运动的区别那样是相对的, 合称**电磁场**(electromagneto field).

我们还注意到来自理论推导过程中产生的重要公式:

$$\mathbf{B} = \frac{1}{4\pi\epsilon_0 c^2} \frac{q\mathbf{v} \times \mathbf{e}_r}{r^2}$$

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B}$$

由于推导的近似性这两个式子似乎没有太多说服力. 但是历史上这两个式子是作为实验规律先被总结, 再有麦克斯韦的普遍电磁理论, 最后才是洛伦兹和爱因斯坦对其背后的相对论基础的研究.

21.1.2 毕奥-萨伐尔定律

强调一下本章研究的范围是重要的。我们暂时只研究静磁场(static magnetic field)，其场源为电流，且其分布不随时间，从而改变的磁场也不随时间改变。或者更简单的说，我们要研究的体系就是上一章建立的稳恒电流体系，它由不变的电荷产生不变的电场，不变的电场驱动不变的电流，现在增加了不变的电流产生不变的磁场这一要素。上一章的讨论告诉我们这就是要求电流分布是无散的：

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

如何把微观电荷的运动同宏观的电流 j, I 建立联系？引言中我们指出来过，微观运动电荷产生的磁场正比于量：

$$\mathbf{C} = q\mathbf{v}$$

这个量被称作电流元(current element)。事实上作为微观电荷 q 通常是基本电荷 e 的量级，从而上述量是一个微元，即使取物质中一个很小的体积 dV ，其内部包含的所有电荷个数 dN 依然未达到宏观量级时，总的电流元依然是个小量，记做 $d\mathbf{C}$ 。我们知道电流密度的定义为：

$$\mathbf{j} = \rho\mathbf{v} = nq\mathbf{v}$$

那么体积产生的总的电流元：

$$d\mathbf{C} = dNq\mathbf{v} = nqvdV = \mathbf{j}dV = \rho dV \cdot \mathbf{v}$$

称为体电流元(volume current element)。同理，我们可以写出面电流元(surface current element)和线电流元(curve current element)：

$$d\mathbf{C} = i dA = \sigma v dA \quad , \quad d\mathbf{C} = I dl = \lambda v dl$$

注意上面线电流元的写法，由于我们不认为 I 是矢量，故把电流元的矢量性由线元 dl 来承载。这么做的好处之后就能体会到。我们发现电流元与电荷元形成了对应关系：

	微观单元	体元	面元	线元
磁	$q\mathbf{v}$	$\mathbf{j}dV$	$i dA$	$I dl$
电	q	ρdV	σdA	λdl

图 21.2: 电荷元与电流元对应

毕奥-萨伐尔定律(Biot-Savart law)是一个实验定律，它用于描述真空中两个孤立的电流元之间的相互作用力：

$$d\mathbf{F}_{12} = \frac{\mu_0}{4\pi} \frac{I_2 dl_2 \times (I_1 dl_1 \times \mathbf{e}_{12})}{r^2}$$

其中系数 $k_m = \frac{\mu_0}{4\pi} \approx 1 \times 10^{-7} \text{H/m}$ ² 为磁力常数，而 μ_0 称作真空磁导率(magnetic permeability of vacuum)。事实上：

$$k_e/k_m = c^2 \quad \text{或者} \quad c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$

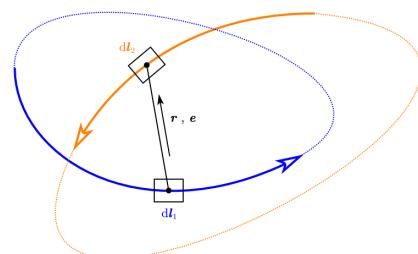


图 21.3: 电流体系间的相互作用力

²注意，2019年5月20日之后，新的国际单位制标准下其值为 $1.000\ 000\ 000\ 82(20) \times 10^{-7} \text{H/m}$ ，不再有严格的 $\mu_0 := 4\pi \times 10^{-7} \text{H/m}$ 。

c 是真空中的光速. 这也揭示了电磁理论统一性背后的相对论本质. 我们这里还用到了电感的国际单位亨利, 而磁场的国际单位特斯拉(Tesla) 与它的关系为:

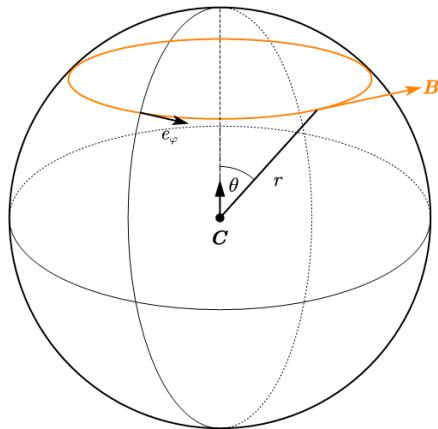
$$1\text{H} = 1\text{T} \cdot \text{m}^2/\text{A}$$

另一个常用的单位来自 CGS 单位制, 即高斯(Gauss), 以纪念这位伟大的数学家在物理学上同样瞩目的贡献. 高斯在历史上被广泛使用, 直至今天也如此, 例如地表地磁场大概就在 $0.25 - 0.60\text{G}$, 它与特斯拉的换算关系为:

$$1\text{T} = 10000\text{G}$$

根据场论的精神, 我们将上式拆解为电流产生场的定律和电流在场中受力的公式:

$$\text{d}\mathbf{F}_{12} = I_2 \text{d}\mathbf{l}_2 \times \text{d}\mathbf{B}_{12}, \quad \text{d}\mathbf{B}_{12} = \frac{\mu_0}{4\pi} \frac{I_1 \text{d}\mathbf{l}_1 \times \mathbf{e}_{12}}{r^2}$$



如前所述, 毕奥-萨伐尔定律可以在相对论背景下由库仑定律推出, 故它继承了库仑定律的平方反比和叠加原理的特性. 但是它突出的特点是具有横向力(transverse force) 的特点: 力的方向与受力物体电流元垂直, 也与另一个只依赖于场源的磁场矢量垂直, 但这个矢量又与场源电流元的方向垂直. 两次叉乘这让力的计算变得比较复杂. 让我们以场源电流元 C 为中心, 其方向建立极坐标系, 那么磁场的大小与方向就可以从公式得到:

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{C \sin \theta}{r^2} \mathbf{e}_\varphi$$

磁场符合叠加原理, 例如一个载流线圈, 源点坐标 \mathbf{r}' , 场点坐标 \mathbf{r} , 记 $\mathbf{R} = \mathbf{r} - \mathbf{r}'$, $\text{d}\mathbf{l} = \text{d}\mathbf{r}'$, 则其产生的磁场公式为:

$$\mathbf{B} = \oint \frac{\mu_0}{4\pi} \frac{I \text{d}\mathbf{l} \times \mathbf{e}_R}{R^2}$$

图 21.4: 电流元产生的磁场

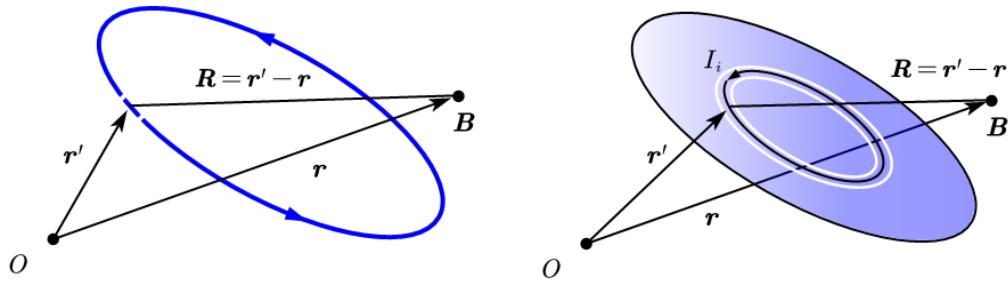


图 21.5: 连续体系产生磁场

载流线圈具有非常重要的地位, 任何体电流分布, 只要是有界区域内的稳恒电流, 电流线就必然构成闭合的回路. 那么就一定可以分解为电流圈. 如果用求和的极限表示积分但是暴力地省去极限符号, 此时磁场计算方法为:

$$\mathbf{B} = \int_{V'} \frac{\mu_0 j \times \mathbf{e}_R}{R^2} \text{d}V = \sum_i \oint \frac{\mu_0 I_i \text{d}\mathbf{l}_i \times \mathbf{e}_R}{R^2}$$

而一个线圈或一个体电流分布在外磁场中的受力为:

$$\mathbf{F} = \oint I \text{d}\mathbf{l} \times \mathbf{B} \Rightarrow \mathbf{F} = \sum_i I_i \text{d}\mathbf{l}_i \times \mathbf{B}$$

在稳恒电流情况下, 讨论一个电流元产生的磁场是没有意义的, 因为空间的磁场必然是整个体系产生的磁场的叠加, 从而单个电流元的磁场并不具有可观测的意义, 除非电流可以独立的被操控以改变. 而讨论一个电流体系各个部分受到的作用力时, 尤其是一个电流线圈受到的作用力时, 区分内力和外力就变得尤其重要. 我们发现, 两个电流元之间的相互作用力之和:

$$d\mathbf{F}_{12} + d\mathbf{F}_{21} = \frac{\mu_0}{4\pi} \frac{I_2 d\mathbf{l}_2 \times (I_1 d\mathbf{l}_1 \times \mathbf{e}_{12})}{r^2} + \frac{\mu_0}{4\pi} \frac{I_1 d\mathbf{l}_1 \times (I_2 d\mathbf{l}_2 \times \mathbf{e}_{21})}{r^2}$$

根据三重矢积公式 $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} + (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$, $\mathbf{e}_{12} = -\mathbf{e}_{21}$. 得到:

$$d\mathbf{F}_{12} + d\mathbf{F}_{21} = -\frac{\mu_0}{4\pi} \frac{(I_1 d\mathbf{l}_1 \times I_2 d\mathbf{l}_2) \times \mathbf{e}_{12}}{r^2}$$

一般来说不总为零, 可见朴素的牛顿第三定律在静磁场情况下失效了. 我们不得不把 $d\mathbf{F}_{12}$, $d\mathbf{F}_{21}$ 不再视作相互作用力, 而是需要通过场作为媒介. 事实上两个运动电荷之间的磁场力是在作为电场力的相对论修正, 从相对论的根本原理上, 相互作用传递速度的有限性导致力的传播被推迟, 原则上也不可能符合牛顿第三定律.

但是如果我们计算两个线圈之间的相互作用力之和:

$$\begin{aligned} \delta\mathbf{F} &= \oint_1 \oint_2 d\mathbf{F}_{12} + \oint_2 \oint_1 d\mathbf{F}_{21} \\ &= \frac{\mu_0 I_1 I_2}{4\pi} \left(\oint_1 \oint_2 \frac{(\mathbf{r}_2 - \mathbf{r}_1) \cdot d(\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} d\mathbf{r}_1 + \oint_2 \oint_1 \frac{(\mathbf{r}_1 - \mathbf{r}_2) \cdot d(\mathbf{r}_1 - \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|^3} d\mathbf{r}_2 \right) \end{aligned}$$

但是:

$$\begin{aligned} \nabla_2 \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} &= -\frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|^3} \quad \Rightarrow \quad \oint_2 \frac{(\mathbf{r}_2 - \mathbf{r}_1) \cdot d(\mathbf{r}_2 - \mathbf{r}_1)}{|\mathbf{r}_2 - \mathbf{r}_1|^3} = 0 \\ \nabla_1 \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} &= -\frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \quad \Rightarrow \quad \oint_1 \frac{(\mathbf{r}_1 - \mathbf{r}_2) \cdot d(\mathbf{r}_1 - \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|^3} = 0 \end{aligned}$$

从而得到:

$$\delta\mathbf{F} = 0$$

当两个环重合时, 这也代表“自己给自己的力”是零. 即使是磁场给线圈的力, 也因为是线圈自己产生的磁场而无法对自己产生力, 这个结果其实也是空间平移对称性(而不是牛顿第三定律)的必然结果. 而空间旋转对称性则给出, 任意选取 O 点计算各个电流元受力对 O 的力矩, 总和必然为零. 或者说两个线圈之间的作用力与作用力矩之间必然符合牛顿第三定律. 其证明在此不再赘述. 这样我们就证明了此前给出的理论公式与牛顿第三定律之间是自洽的, 无散的电流体系下牛顿第三定律依然成立, 尽管表面上电流元的毕奥萨伐尔定律与牛顿第三定律有冲突.

更有甚者, 我们指出, 根据麦克斯韦方程的协变性, 即使是微观电荷做相对论性运动, 只要是稳恒电流体系, 那么毕奥萨伐尔定律可以作为一个推论而具有普适性. 而对于非稳恒体系, 后面我们将会证明, 磁场此时并不能简单地由电流确定而依赖于电场的变化率, 甚至它成为了自由存在的物质形式而摆脱了电荷造成任何约束.

21.2 两个定律与矢势

21.2.1 磁场的环路定律

静磁场的环路定律(circuital law), 又称安培定律(Ampère's law)指出, 对于一个可定向曲面 A 和它的闭合右手边界 ∂A , 有:

$$\oint_{\partial A} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_A \mathbf{j} \cdot d\mathbf{A}$$

我们用毕奥-萨伐尔定律来证明环路定律。首先，上式中的 A 和其边界上的场点位矢取为 r ，则 $dl = dr$ ，而场源电流分解为电流环 I_i 。将电流环上坐标选为 r' 。那么以上积分可按照叠加原理分解为：

$$\Gamma = \sum_i \Gamma_i = \sum_i \oint_{\partial A} \mathbf{B}_i \cdot dl$$

而每一项环量就可以用场点和源点的两次环积分表示，将 e_R 记做 e ：

$$\begin{aligned} \Gamma_i &= \oint_r \mathbf{B}_i \cdot dr \\ &= \oint_r \left(\oint_{r'} \frac{\mu_0 I_i dr' \times e_R}{4\pi R^2} \right) \cdot dr \\ &= \frac{\mu_0 I_i}{4\pi} \oint_r \oint_{r'} \frac{(dr \times dr') \cdot e_R}{R^2} \end{aligned}$$

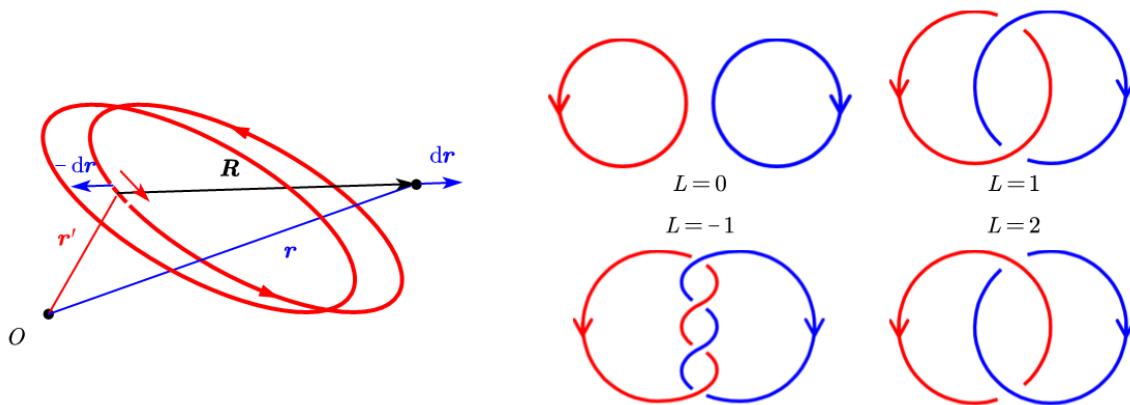


图 21.6: 环绕数

如上图所示，对上面的积分做如下诠释：当从场点 r 去看线圈时，线圈对场点产生了一个立体角 Ω 。当观察点移动 dr 时，考察立体角的改变 $d\Omega$ 。可以等效地把线圈反向移动 $-dr$ 。图中所示情况立体角将减小，算法为：

$$-d\Omega = \int \frac{dA \cdot (-e_R)}{R^2} = \oint_{r'} \frac{(-dr \times dr') \cdot (-e_R)}{R^2}$$

下面考虑以源电流圈（21.6中红圈）为边界的那个面，场点形成的安培回路（21.6中蓝圈）与这个面的交点个数为 $n_1 + n_2$ ，其中 n_1 代表蓝圈正向穿过面时方向与红圈形成右手规则的交点个数， n_2 代表蓝圈正向穿过面时方向与红圈形成左手规则的交点个数。于是再定义环绕数(linking number)：

$$\ell = n_1 - n_2$$

这是一个拓扑学不变量，它只取决于两个回路的整理性质，与面的取法无关，而且具有对称性（红对蓝与蓝对红的环绕数一致）。

从而我们注意到，如果环绕数恰好是一，那么从面上出发， Ω 从一个接近 2π 的大小一直在减小直到变为零，然后变成负的，直到回到面上直到变成接近 -2π ，虽然初末位置立体角不一定都等于 2π ，但是一定有：

$$\oint_r d\Omega = -4\pi$$

相似的，如果考虑一个环绕数为 L 的安培回路，就有：

$$\oint_r d\Omega = -4\pi\ell$$

将此前的立体角微分代入, 得到:

$$\oint_{\mathbf{r}} \oint_{\mathbf{r}'} \frac{(\mathbf{dr} \times \mathbf{dr}') \cdot \mathbf{e}_R}{R^2} = 4\pi\ell$$

最后代入此前的磁场环量, 得到:

$$\Gamma_i = \mu_0 \ell_i I_i$$

最后我们总结出:

$$\oint_{\partial A} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \sum_i \ell_i I_i$$

现在做极限化处理, 注意到求和式实际上就是统计了所有以右手规则为正向通过面的电流, 统一用电流密度表示就变成了自带正负的点乘的积分:

$$\oint_{\partial A} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_A \mathbf{j} \cdot d\mathbf{A}$$

此即环路定律, 由数学上的开尔文-斯托克斯定理, 我们可以把上式写出局域的微分形式:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}$$

21.2.2 矢势与磁场的高斯定律

对电流元产生的磁场 $d\mathbf{B} = \frac{\mu_0}{4\pi} \frac{\mathbf{j} dV \times \mathbf{e}_r}{r^2}$ 引入以下矢量场是方便的:

$$\mathbf{A} = \frac{\mu_0}{4\pi} \frac{\mathbf{j} dV}{r}$$

这是因为如果计算球坐标下的旋度, 考虑到 \mathbf{A} 分解到 r, θ 方向分别为:

$$A_r = \frac{\mu_0}{4\pi} \frac{j dV}{r} \cos \theta, \quad A_\theta = -\frac{\mu_0}{4\pi} \frac{j dV}{r} \sin \theta$$

那么考虑一个回路³ S : $(r, \theta) - (r + dr, \theta) - (r + dr, \theta + d\theta) - (r, \theta + d\theta) - (r, \theta)$, 利用旋度的定义式:

$$\mathbf{n} \cdot \nabla \times \mathbf{A} = \lim_{S \rightarrow 0} \oint_{\partial S} \mathbf{A} \cdot d\mathbf{l} / S$$

得到其旋度:

$$\nabla \times \mathbf{A} = \frac{1}{r} \left(\frac{\partial r A_\theta}{\partial r} - \frac{\partial A_r}{\partial \theta} \right) \mathbf{e}_\varphi$$

代入便得到以下关系式:

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{j dV}{r^2} \sin \theta \mathbf{e}_\varphi = \nabla \times \mathbf{A}$$

\mathbf{A} 这个矢量称作**磁矢势**(magnetic vector potential). 由于叠加原理, 一个电流分布体系的磁矢势和磁场应当为:

$$\mathbf{A} = \int_V \frac{\mu_0}{4\pi} \frac{\mathbf{j} dV}{r}, \quad \mathbf{B} = \int_V \frac{\mu_0}{4\pi} \frac{\mathbf{j} dV \times \mathbf{e}_r}{r^2}$$

同样地将满足:

$$\mathbf{B} = \nabla \times \mathbf{A}$$

以上旋度关系写成积分形式意味着一种计算磁通量的特殊方法:

$$\Phi = \oint_{\partial S} \mathbf{A} \cdot d\mathbf{l} = \int_S \mathbf{B} \cdot d\mathbf{S}$$

³为了避免与磁矢势符号重复, 暂时用 S 表示面元.

对于无边界的闭合曲面, 上式直接证明了磁场**高斯定律**(Gauss's law):

$$\begin{aligned}\partial\partial V = \emptyset \Rightarrow \oint_{\partial V} \mathbf{B} \cdot d\mathbf{S} &= \oint_{\partial\partial V} \mathbf{A} \cdot dl = 0 \\ \Rightarrow \nabla \cdot \mathbf{B} &= 0\end{aligned}$$

我们如果计算此前电流元产生的磁矢势的散度:

$$\nabla \cdot \mathbf{A} = \nabla \cdot \left(\frac{\mu_0 j dV}{4\pi r} \mathbf{e}_z \right) = \frac{\mu_0 j dV}{4\pi} \frac{\partial}{\partial z} \left(\frac{1}{r} \right) = \frac{\mu_0 j dV \cos \theta}{4\pi r^2} = \frac{\mu_0 j dV \cdot \mathbf{e}_r}{4\pi r^2}$$

又发现对电流元得到非零的结果. 但是如果考虑线圈:

$$\nabla \cdot \mathbf{A} = \oint \frac{\mu_0 I dl \cdot \mathbf{e}_R}{4\pi R^2} = \frac{\mu_0 I}{4\pi} \oint \frac{\mathbf{e}_R}{R^2} \cdot dr' = \frac{\mu_0 I}{4\pi} \oint \nabla_{r'} \left(\frac{1}{R} \right) \cdot d\mathbf{R} = \frac{\mu_0 I}{4\pi} \oint d \left(\frac{1}{R} \right) = 0$$

从而对于任意稳恒电流体系, 通过此前积分定义的磁矢势也是无散的. 最后将磁矢势与磁场关系和磁场与电流关系联立, 注意到:

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$$

即三重矢积公式. 我们把梯度算符 ∇ 直接看作了一个矢量. 上式右侧第一项由于磁矢势无散故为零, 从而得到:

$$\nabla^2 \mathbf{A} = -\mu_0 \mathbf{A}$$

电场中的式子 $\nabla^2 \varphi = -\rho/\varepsilon_0$ 与它相呼应.

可以设想存在可以像点电荷产生电场那样的方式产生磁场的物质, 称作**磁单极子**(magnetic monopole), 从而可以产生不为零的净**磁荷**(magnetic charge), 事实证明这样的物质至今都未曾找到, 从而相伴的磁场形式也仅仅存在与理论中. 但是, 总磁荷为零的体系: **磁偶极子**(magnetic dipole), 不仅仅是纯粹的理论模型, 后面可以发现线圈产生的磁场与磁偶极子是相似的.

我们记得电势可以被视作电荷在电场中运动的势能有关的物理量, 即 $E_p = q\varphi$. 那么电荷在磁场中运动时, 磁矢势代表什么物理概念呢? 如果空间中只有纯的磁场, 那么显然磁场力不做功:

$$P = \frac{dW}{dt} = \mathbf{F}_m \cdot \mathbf{v} = (q\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v} = 0$$

所以电荷一定是动能不变的. 但是即使如此, 我们也不认为磁场力是一个保守力, 这是因为保守力的定义应该是: 能够用势能的梯度来表示的作用力. 我们写出一个项:

$$\boldsymbol{\pi} = q\mathbf{A}$$

这个项不再具有能量的量纲而是动量的量纲. 我们知道电场除磁场的量纲为速度, 而两者与对应的矢势标势都差一个长度量纲, 故 $q\mathbf{A}$ 量纲等于用 $q\varphi$ 的能量量纲除速度, 即动量量纲. 这个动量称为**势动量**(potential momentum). 而带电粒子的旧动量 $\mathbf{p} = m\mathbf{v}$ 就称作**机械动量**(kinetic momentum), 它是运动造成的, 而势动量是位置 \mathbf{r} 造成的 $\boldsymbol{\pi}(\mathbf{r})$, 是与磁场相互作用引起的动量. 两者之和:

$$\mathbf{P} = \mathbf{p} + \boldsymbol{\pi}$$

称作**正则动量**(canonical momentum). 不像在电场中的电荷运动机械能(动能+势能)守恒, 这个和也是一般不守恒的:

$$\begin{aligned}\dot{\mathbf{P}} &= \dot{\mathbf{p}} + \dot{\boldsymbol{\pi}} \\ &= \mathbf{F} + q\dot{\mathbf{A}} \\ &= q\mathbf{v} \times (\nabla \times \mathbf{A}) + q\mathbf{v} \cdot \nabla \mathbf{A} \\ &= -\nabla(-q\mathbf{v} \cdot \mathbf{A})\end{aligned}$$

最后一步把 ∇ 视作矢量使用了三重矢积公式. 注意到瞬时的 v 不被视作变量而视作常数而从偏导数符号内外移动不引起改变. 我们发现瞬时的正则动量导数似乎恰好等于某个势能的负梯度:

$$V = -q\mathbf{v} \cdot \mathbf{A}$$

这被称作**广义势**(generalized potential). 它含有速度, 从而每一个时刻有着完全不同的形式. 但不管怎样, 粒子在磁场中的运动的确是写成了一个动量导数等于势的梯度的形式(拉格朗日方程形式). 即使这样, 我们依然不认为磁场力是一个保守力.

我们来看一个具体的例子. 假设空间中的电流分布产生了一个随 y 方向变化的 x 方向的磁矢势分布:

$$\mathbf{A} = (f(y), 0, 0) = f(y)\mathbf{e}_x$$

那么可以验证这个式子的确符合:

$$\nabla \cdot \mathbf{A} = \frac{\partial f(y)}{\partial x} = 0$$

而产生它的电流和对应的磁场就是:

$$\mathbf{j} = -\nabla^2 \mathbf{A}/\mu_0 = -\frac{1}{\mu_0} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) f(y)\mathbf{e}_x = -\frac{f''(y)}{\mu_0}\mathbf{e}_x$$

$$\mathbf{B} = \nabla \times \mathbf{A} = -\frac{\partial f(y)}{\partial y}\mathbf{e}_z = -f'(y)\mathbf{e}_z$$

这表示一个垂直 $x-y$ 平面的磁场. 那么粒子在某一瞬时的广义势:

$$V = -q\mathbf{v} \cdot \mathbf{A} = -qv_x f(y)$$

其瞬时梯度就是:

$$-\nabla V = -qv_x f'(y)\mathbf{e}_y$$

而粒子的正则动量是:

$$\mathbf{P} = (mv_x + qf(y))\mathbf{e}_x + mv_y\mathbf{e}_x$$

根据此前的论述:

$$\dot{\mathbf{P}} = -\nabla V$$

从而我们发现, 在 x 方向粒子的**广义动量守恒**(conervation of generalized momentum):

$$mv_x + qf(y) = \text{Const.}$$

而在 y 方向有一动力学方程:

$$ma_y = -qv_x f'(y)$$

读者可以用牛顿力学同样的论证以上两个式子的成立性.

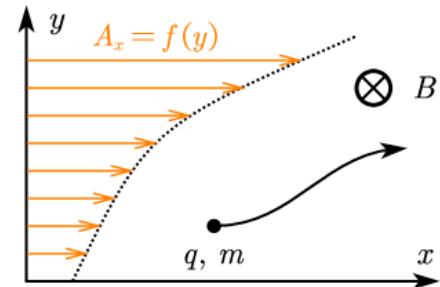


图 21.7: 正则动量守恒

21.3 电流体系

研究更进一步的磁场与电流相互作用的模型前, 熟悉一些常见的电流体系的结论是大有裨益的, 这些包括:

21.3.1 磁偶极子

同样类似于对静电学中一个分布在原点附近的电流体系在远场产生的电势进行展开的思想, 我们此时运用到静磁学情形: 计算一个分布在原点附近的电流体系在远场产生的磁矢势:

$$\mathbf{A} = \int_{V'} \frac{\mu_0}{4\pi} \frac{j dV'}{|\mathbf{r} - \mathbf{r}'|}$$

利用分解为线圈可以简化以上计算:

$$\mathbf{A} = \sum_i \oint \frac{\mu_0}{4\pi} \frac{I_i d\mathbf{l}_i}{|\mathbf{r} - \mathbf{r}_i|}$$

依然利用静电学中的展开公式:

$$\begin{aligned} \frac{1}{|\mathbf{r} - \mathbf{r}_i|} &= [(r - r_i)^2]^{-\frac{1}{2}} = (r^2 - 2\mathbf{r}_i \cdot \mathbf{r} + \mathbf{r}_i^2)^{-\frac{1}{2}} \\ &\approx (r^2 - 2\mathbf{r}_i \cdot \mathbf{r})^{-\frac{1}{2}} = \frac{1}{r} \left(1 - 2\frac{\mathbf{r}_i \cdot \mathbf{r}}{r^2}\right)^{-\frac{1}{2}} \\ &\approx \frac{1}{r} \left(1 + \frac{\mathbf{r}_i \cdot \mathbf{r}}{r^2}\right) \\ &= \frac{1}{r} + \frac{\mathbf{e}_r}{r^2} \cdot \mathbf{r}_i \end{aligned}$$

代入上式, 注意到 $d\mathbf{l}_i = d\mathbf{r}_i$ 得到磁矢势展开的前两项:

$$\mathbf{A} \approx \frac{\mu_0}{4\pi r} \sum_i I_i \oint d\mathbf{r}_i + \frac{\mu_0}{4\pi r^2} \sum_i I_i \oint (\mathbf{e}_r \cdot \mathbf{r}_i) d\mathbf{r}_i$$

第一项由于对回路积分必然为零. 第二个积分中需要注意 \mathbf{e}_r 照理应当是可以提取到积分号外面的常矢量, 但是这样积分内部的 $\mathbf{r}_i d\mathbf{r}_i$ 构成了一个不直观的张量. 我们采用一个更加简单的做法: 常矢量点乘法(method of dot product by a constant vector). 任意取一个常矢量 \mathbf{v} 来点乘积分结果:

$$\mathbf{v} \cdot \oint (\mathbf{e}_r \cdot \mathbf{r}_i) d\mathbf{r}_i = \oint (\mathbf{e}_r \cdot \mathbf{r}_i) (\mathbf{v} \cdot d\mathbf{r}_i) = \mathbf{e}_r \cdot \oint (\mathbf{v} \cdot d\mathbf{r}_i) \mathbf{r}_i$$

这样倒是可以实现 \mathbf{e}_r 移到积分外, 但代价是另一个类似的 \mathbf{v} 又到了积分内. 其实无济于事. 但是注意到中间的积分其实完全可以使用开尔文-斯托克斯定理:

$$\oint (\mathbf{e}_r \cdot \mathbf{r}_i) (\mathbf{v} \cdot d\mathbf{r}_i) = \oint [(\mathbf{e}_r \cdot \mathbf{r}_i) \mathbf{v}] \cdot d\mathbf{r}_i = \int_{S_i} \nabla \times [(\mathbf{e}_r \cdot \mathbf{r}_i) \mathbf{v}] \cdot d\mathbf{S}_i$$

根据叉乘的定义:

$$\nabla \times [(\mathbf{e}_r \cdot \mathbf{r}_i) \mathbf{v}] = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ (\mathbf{e}_r \cdot \mathbf{r}_i) v_x & (\mathbf{e}_r \cdot \mathbf{r}_i) v_y & (\mathbf{e}_r \cdot \mathbf{r}_i) v_z \\ \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \end{vmatrix}$$

其中 $\mathbf{e}_r = (e_{rx}, e_{ry}, e_{rz})$ 和 (v_x, v_y, v_z) 都是常的. 故求偏导只发生在点乘式 $\mathbf{e}_r \cdot \mathbf{r}_i$ 中的一项上给出不为零的系数. 即:

$$\nabla \times [(\mathbf{e}_r \cdot \mathbf{r}_i) \mathbf{v}] = \begin{vmatrix} e_{rx} & e_{ry} & e_{rz} \\ v_x & v_y & v_z \\ \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \end{vmatrix} = \mathbf{e}_r \times \mathbf{v}$$

最终得到:

$$\mathbf{v} \cdot \oint (\mathbf{e}_r \cdot \mathbf{r}_i) d\mathbf{r}_i = \int_{S_i} (\mathbf{e}_r \times \mathbf{v}) \cdot d\mathbf{S}_i = \mathbf{v} \cdot \left[\left(\int_{S_i} d\mathbf{S}_i \right) \times \mathbf{e}_r \right] = \mathbf{v} \cdot (\mathbf{S}_i \times \mathbf{e}_r)$$

由于工具 \mathbf{v} 的任意性, 现在就可以看出以下关系:

$$\oint (\mathbf{e}_r \cdot \mathbf{r}_i) d\mathbf{r}_i = \mathbf{S}_i \times \mathbf{e}_r$$

再代入此前的磁矢势, 得到最终结果:

$$\mathbf{A} \approx \frac{\mu_0}{4\pi r^2} \left(\sum_i I_i \mathbf{S}_i \right) \times \mathbf{e}_r$$

在讨论最终结果前我们对上式出现的一种现象需要做进一步的解释. 如图21.8所示, 一个线圈对应的面矢量 \mathbf{S} 的标准定义实际上对应着以下线圈上的路径积分:

$$\mathbf{S} = \frac{1}{2} \oint \mathbf{r} \times d\mathbf{r}$$

这样定义也绝非完美, 我们自然觉得, 一个线圈对应的面矢量的确定应该是与坐标原点选取无关的, 但上式在表面上有关. 如果原点位移了 \mathbf{R} , 两个积分的差为:

$$\begin{aligned} \mathbf{S}' &= \frac{1}{2} \oint (\mathbf{r} - \mathbf{R}) \times d(\mathbf{r} - \mathbf{R}) \\ &= \frac{1}{2} \oint (\mathbf{r} - \mathbf{R}) \times d\mathbf{r} \\ &= \mathbf{S} - \frac{1}{2} \mathbf{R} \times \oint d\mathbf{r} \\ &= \mathbf{S} \end{aligned}$$

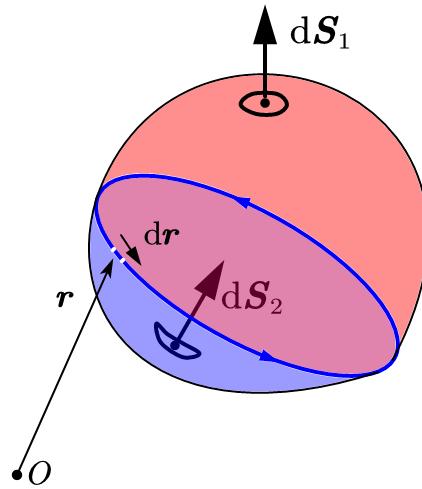


图 21.8: 线圈与面矢量

故以上定义的确不依赖于原点的选取. 但是我们要问, 为什么上式就表示以线圈为边界的面矢量呢? 而这样定义又有何好处? 事实上, 如过按照通常的方式来定义一个边界所确定的面的面矢量, 那么就面临着一个奇怪的问题: 不同的面可能共用了同一个边界, 甚至会出现莫比乌斯带那样, 一个简单的闭合曲线却以一个反常的不可定向的面的边界的形式出现了, 我们甚至可以在一个圆环作为边界的情况下构造出莫比乌斯带型的不可定向曲面来, 将莫比乌斯带的边界扯成一个圆环即可. 为了避免这种情况, 我们最好是用曲线积分来定义面矢量, 这样对于复杂曲线(比如自己打结的曲线, 比如螺旋线)也不必受难以确定以其为边界的曲面之苦. 事实上, 如果对于21.8那样的曲线较为简单的情况, 选取以其为边界的可定向曲面, 依然把右手定则确定的微元面矢量积分定义为整个面的面矢量, 那么对于21.8中两个不同的面:

$$\mathbf{S}_1 = \int_{S_1} d\mathbf{S}_1, \quad \mathbf{S}_2 = \int_{S_1} d\mathbf{S}_2$$

我们自然也要问, 以上两个积分是否会相等? 事实上如果把两个积分相减, 就得到了在一个体的闭合表面上的面元积分:

$$\mathbf{S}_1 - \mathbf{S}_2 = \oint_{\partial V} d\mathbf{S}$$



图 21.9: 莫比乌斯带

要想得到上式的值, 我们又再一次使用常矢量点乘法:

$$\mathbf{v} \cdot (\mathbf{S}_1 - \mathbf{S}_2) = \oint_{\partial V} \mathbf{v} \cdot d\mathbf{S}$$

很显然, 这表示一个常矢量场在闭合面上的通量, 如匀强电场通过一个闭合面的电通量代数和, 从而一定是零. 由 \mathbf{v} 的任意性, 总有:

$$\mathbf{S}_1 - \mathbf{S}_2 = 0 \quad , \quad \mathbf{S}_1 = \mathbf{S}_2$$

从而这个定义也是不依赖于曲面的选取的, 而为了把之前标准定义的曲线积分和对面元的积分联系起来, 我们再使用常矢量点乘法:

$$\mathbf{v} \cdot \frac{1}{2} \oint_{\partial S} \mathbf{r} \times d\mathbf{r} = \frac{1}{2} \oint_{\partial S} (\mathbf{v} \times \mathbf{r}) \cdot d\mathbf{r} = \frac{1}{2} \int_S \nabla \times (\mathbf{v} \times \mathbf{r}) \cdot d\mathbf{S}$$

利用 \mathbf{v} 为常矢量时可以在 ∇ 前后移动的特性, 得到:

$$\nabla \times (\mathbf{v} \times \mathbf{r}) = (\nabla \cdot \mathbf{r})\mathbf{v} - (\mathbf{v} \cdot \nabla)\mathbf{r} = 3\mathbf{v} - \mathbf{v} = 2\mathbf{v}$$

代入上式:

$$\mathbf{v} \cdot \frac{1}{2} \oint_{\partial S} \mathbf{r} \times d\mathbf{r} = \frac{1}{2} \int_S 2\mathbf{v} \cdot d\mathbf{S} = \mathbf{v} \cdot \int_S d\mathbf{S}$$

从而就证明的确这个曲线积分就是面矢量:

$$\frac{1}{2} \oint_{\partial S} \mathbf{r} \times d\mathbf{r} = \int_S d\mathbf{S} = \mathbf{S}$$

关于用曲线积分得到总的面矢量 \mathbf{S} 还有一种几何理解方式, 再仔细观察图21.8就会发现, 项:

$$\frac{1}{2} \mathbf{r} \times d\mathbf{r}$$

实际上就是原点 O 出发向曲线上的一段 $d\mathbf{r}$ 连线形成的三角形面元, 它恰好是从 O 出发向曲线上的所有点引射线形成的一个锥面上的一个面元, 故积分就得到整个锥面的总面矢量且与曲线环绕方向恰好符合右手规则, 它也是一种特殊的以曲线为边界的面的总面矢量.

终于我们回到此前电流体系形成的磁矢势问题, 现在我们就定义一个载流 I_i 的线圈形成的磁矩(magnetic moment) 为:

$$\mathbf{m}_i = I_i \mathbf{S}_i$$

而这个电流体系形成的总磁矩就是:

$$\mathbf{m} = \sum_i \mathbf{m}_i = \sum_i I_i \mathbf{S}_i$$

那么在远场的磁矢势的领头项就是由磁矩产生的平方反比项:

$$\mathbf{A} = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{e}_r}{r^2}$$

这个式子其实非常类似电流元 C 产生的磁场公式:

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{\mathbf{C} \times \mathbf{e}_r}{r^2}$$

但是要计算磁矩产生的磁场, 应对以上形式的磁矢势来求旋度, 计算可得:

$$B_r = \frac{\mu_0}{4\pi} \frac{2m \cos \theta}{r^3} \quad , \quad B_\theta = \frac{\mu_0}{4\pi} \frac{m \sin \theta}{r^3}$$

也许令人惊讶的是, 其关于 r, θ 变化的形式居然与电偶极子产生的电场的公式一致⁴.

磁矩产生的磁场具有非常重要的地位, 第一是因为任何情况下这都会是一个电流体系激发的磁场的领头项, 因为磁单极子不存在, 类似点电荷产生平方反比的电场那样的磁荷项, 由于磁高斯定律, 永远是不可能的. 磁矩的计算方式上面已经给出, 但是我们依然要指出, 将求和再还原为对提电流的积分, 可以得到磁矩的体积分算法:

$$\mathbf{m} = \frac{1}{2} \int_V \mathbf{r} \times \mathbf{j} dV$$

而最简单的磁矩模型就是一个平面圆环, 其电流为 I , 半径为 R . 那么磁矩就是:

$$m = I\pi R^2$$

而考虑 $I \rightarrow \infty, S \rightarrow 0$ 的模型, 就构成了点磁偶极子(point dipole)的情况. 此时体系产生的磁场将严格符合我们得出的三个计算式:

$$\mathbf{A} = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{e}_r}{r^2}, \quad B_r = \frac{\mu_0}{4\pi} \frac{2m \cos \theta}{r^3}, \quad B_\theta = \frac{\mu_0}{4\pi} \frac{m \sin \theta}{r^3}$$

第二, 任何稳恒电流体系可以向线圈分解, 而任何线圈又可以向磁偶极子来分解. 我们把载有电流 I 的线圈所在的可定向曲面分解为无穷多无穷小的网格, 并让每一个网格都载有同样的方向的电流 I , 那么网格形成的总电流分布实际上就是原线圈的电流分布, 因为有公共边的两个网格在公共边上的电流一定恰好反向而产生的电流互相抵消, 最后只剩下最外面的一圈边界上电流被保留下来. 从而不管是产生的场还是在外场中的受力均可以用于等效原来的线圈.

磁偶极子的磁势能, 受力与力矩:

$$V = -\mathbf{m} \cdot \mathbf{B}, \quad \mathbf{F} = \mathbf{m} \cdot \nabla \mathbf{B}, \quad \mathbf{M} = \mathbf{m} \times \mathbf{B}$$

21.3.2 磁化强度

- 磁化强度(magnetization):

$$\mathbf{M} = n\mathbf{m}$$

- 磁化强度造成的体电流与面电流分布:

$$\mathbf{j}_M = \nabla \times \mathbf{M}, \quad \mathbf{i}_M = \mathbf{M} \times \mathbf{n}$$

21.3.3 若干对称体系的磁场

- 无限长载流直线:

$$\mathbf{A} = -\frac{\mu_0 I}{2\pi} \ln r, \quad \mathbf{B} = \frac{\mu_0 I}{2\pi r} \mathbf{e}_\varphi$$

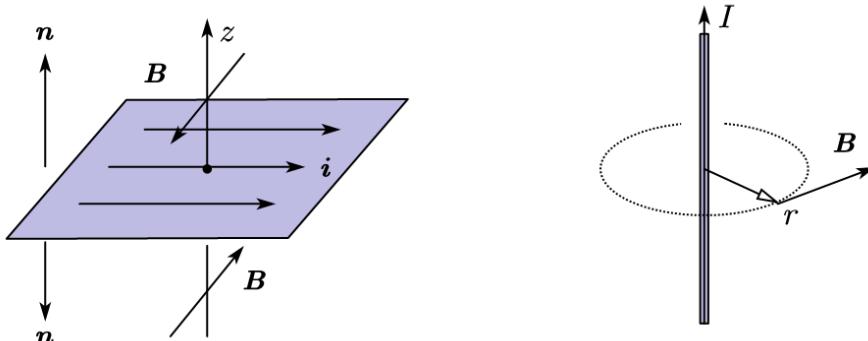


图 21.10: 无限大平面与无限长直线

⁴其实这个形式是拉普拉斯算符, 或者说, 背后的对称性导致的必然结果

- 无限大载流平面:

$$\mathbf{B} = \frac{1}{2}\mu_0 \mathbf{i} \times \mathbf{n}$$

- 圆环: 半径为 R 的标准载流线圈轴线上磁场:

$$\mathbf{B} = \frac{\mu_0 I R}{2(R^2 + z^2)^{\frac{3}{2}}} \mathbf{e}_z$$

该处的径向场强可以由磁高斯定理确定:

$$\frac{1}{2} \frac{dB_r}{dr} + \frac{dB_z}{dz} = 0 \quad \Rightarrow \quad B_r \approx -\frac{3\mu_0 I R z r}{(R^2 + z^2)^{\frac{5}{2}}}$$

- 无限长密绕螺线管: 无论其横截面形状, 总是有:

$$B = \mu_0 i = \mu_0 n I = \frac{\mu_0 N I}{l}$$

21.4 磁介质与磁能

21.4.1 微观角度理解磁化

- 分子的固有磁矩: 电子自旋磁矩 + 电子轨道磁矩.
- 如果分子具有固有磁矩, 那么按照热力学规律发生取向磁化: 磁矩倾向于取与外磁场相同的方向, 宏观上产生顺磁性(paramagnetism). 类似于电介质的极化, 微观上顺磁磁化规律可以写作:

$$\bar{\mu} = \alpha \mathbf{B}$$

- 电子的磁矩与角动量的比称为旋磁比(gyromagnetic ratio), 轨道旋磁比与自旋旋磁比分别为:

$$\gamma_L = -\frac{e}{2m_e} \quad , \quad \gamma_S = -\frac{e}{m_e}$$

角动量量子化理论告诉我们, z 方向上电子轨道与自旋角动量只能为:

$$L_z = \pm n\hbar \quad , \quad S_z = \pm \frac{\hbar}{2}$$

故 z 方向上的磁矩的单位就是著名的玻尔磁子(Bohr magneton):

$$\mu_B = \frac{e\hbar}{2m_e} = 0.00579 \text{ meV/T}$$

事实上由于自旋-轨道耦合, 分子磁矩可以是玻尔磁子的分数倍.

- 原子核也有磁矩, 显然由于质子质量比质子质量大三个量级, 故根据旋磁比算法, 其磁矩也小三个量级. 虽然对分子磁矩和宏观磁化几乎没有贡献, 但是利用在磁场下形成的各个能级间跃迁吸收发射电磁辐射的原理可以造成核磁共振(nuclear magnetic resonance) 现象.
- 如果分子的固有磁矩为零, 那么在外磁场下会产生一个与外磁场反向的磁矩, 宏观上造成抗磁性(diamagnetism). 从微观来看似乎与电介质极化的方向相反, 但是从宏观效果上却与电介质削弱电场的效果相同, 这是电与磁固有的区别导致的.

- 抗磁性的微观起源为电子轨道运动的拉莫尔进动(Larmor precession). 由轨道运动产生的力矩, 角动量, 磁矩三者关系可得:

$$\begin{aligned} \mathbf{M} &= \boldsymbol{\mu} \times \mathbf{B} \quad , \quad \boldsymbol{\mu} = \gamma \mathbf{L} \quad , \quad \frac{d}{dt} \mathbf{L} = \mathbf{M} \\ \Rightarrow \quad \frac{d}{dt} \mathbf{L} &= -\gamma \mathbf{B} \times \mathbf{L} = \frac{e \mathbf{B}}{2m} \times \mathbf{L} \end{aligned}$$

故在运动学上, 电子将以以下角速度发生进动:

$$\boldsymbol{\Omega} = \frac{e}{2m} \mathbf{B}$$

电子原来的轨道运动半径如果按照玻尔半径 a_0 估计, 而角动量量级为 \hbar , 那么现在附加的角速度与原来的角速度的比为:

$$\Omega/\omega \sim \frac{ea_0^2}{2\hbar} B$$

故产生的磁矩与原来的玻尔磁子的比值的量级也与之相当, 得到磁矩的量级为:

$$\bar{\boldsymbol{\mu}} \sim -\frac{e^2 a_0^2}{4m} \mathbf{B}$$

系数便是抗磁性磁化的分子磁化率的量级, 它一般比顺磁性磁化要小一到两个量级.

- 有一些金属, 或者特殊的金属化合物具有独特的铁磁性(ferromagnetism), 微观上它们由介观的包含大量原子的磁畴(magnetic domains) 构成. 每一个磁畴由于单元之间的关联已经超过了热力学的涨落而成为了决定性的因素, 使得所有电子的自旋几乎都严格指向同一个方向, 磁化达到饱和. 而材料的磁化过程其实是不同的磁畴在外磁场下的转向. 它的磁化率将远远大于简单的顺磁和抗磁磁化. 而且体现出非线性和历史相关性.

21.4.2 宏观角度理解磁化

- 出于 \mathbf{B} 和 \mathbf{M} 的旋度分别为总电流密度和磁化电流密度, 引入磁场强度(magnetic field strength) 以区别于以往一贯描述磁场的 \mathbf{B} , 一般可区别称作磁通密度(magnetic flux density):

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M}$$

这样这个矢量的旋度就只取决于外电流:

$$\nabla \times \mathbf{H} = \mathbf{j}_f$$

- 在顺磁或铁磁的柱状介质外绕制密绕螺线管, 那么没有介质时其磁通密度为:

$$B_0 = \mu_0 i$$

那么, 由于此时加入介质时磁化电流不影响 \mathbf{H} 的计算, 故磁场强度为:

$$H = \frac{B_0}{\mu_0} = i$$

但是磁化电流的存在将产生一个与原磁场同向的附加磁场, 使得介质内部的磁场变大, 那么定义此时磁场与原磁场的比值为相对磁导率(relative magnetic permeability):

$$B/B_0 = \mu_r$$

那么把绝对磁导率定义为 $\mu = \mu_r \mu_0$, 就有以下关系式:

$$B = \mu H = \mu i$$

$$M = (\mu_r - 1)H = \chi_m H$$

其中 χ_m 就是宏观的**磁化率**(magnetic susceptibility). 对于抗磁性物质, 其值是小于零的.

21.4.3 磁场能量

磁场体系的能量的计算方法可以证明可以用电流与矢势和用能量密度两种计算方法:

$$I = \frac{1}{2} \int \mathbf{j} \cdot \mathbf{A} dV = \int \frac{1}{2\mu_0} \mathbf{B}^2 dV$$

同样地, 两个体系同时存在时, 体系将产生自能与相互作用能. 如两个电感元件如果之间存在互感, 那么其总能量应当为:

$$E = \frac{1}{2} L_1 I_1^2 + \frac{1}{2} L_2 I_2^2 + M I_1 I_2$$

如果存在介质, 其能量密度需要加上磁化带来的能量, 能量密度变为:

$$w = \frac{1}{2\mu} \mathbf{B}^2 = \frac{1}{2} \mu \mathbf{H}^2 = \frac{1}{2} \mathbf{B} \cdot \mathbf{H}$$

第二十二章 磁生电

22.1 电磁感应

静止电荷产生的静电场，稳恒电流产生的静磁场，这些都已经被我们研究地足够充分了。是时候，我们也该思考以下问题：

1. 如果电荷不是完全静止，静电场是否会有改变？
2. 如果电荷不是完全静止，会不会产生一个磁场？
3. 如果电流不是完全稳恒，静磁场是否会有改变？
4. 如果电流不是完全稳恒，会不会产生一个电场？

其中问题 1, 3 在我们看来已经有答案：非静止的、非稳恒的源自然要产生非静电场和非静磁场，至少产生的场需要相对元产生一定的推迟。问题 2, 4 就值得更进一步地研究。其中问题 2 的答案是可以被推理出来的，它就是下一章麦克斯韦发现的重要现象：不能单独假设只有电流会产生磁场，不然磁场的旋度将不会自洽，从而根据电荷守恒发现只需要假设电荷运动造成的变化的电场也能造成一个磁场。尽管第三点更容易被解释。但是历史地看，人们更早地得到了问题 4 的答案：变化的磁场（源于非稳恒的电流）必然造成一个新的电场，它完全独立于电荷产生的那一类电场，因为这个电场有旋度而无散度。它被称作**涡旋电场**(eddy electric field)。

在对历史上的背景做一个回顾前重新探讨一下此前概念的定义是有必要的。我们有洛伦兹力公式：

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

这个表达式实际上构成了描述电场、磁场强度的量 \mathbf{E} , \mathbf{B} 的唯一操作性定义：只需要在空间中放置试探电荷 q ，并给它一个测试速度 \mathbf{v} 就可以通过受力的结果定义空间该点的电磁场。它的普遍性与协变性只能通过实验来验证，两者都是随时间变化的矢量场，且其值依赖于参考系的选取：

$$\mathbf{E} = \mathbf{E}(\mathbf{r}, t) \quad , \quad \mathbf{B} = \mathbf{B}(\mathbf{r}, t)$$

$$(\mathbf{E}, \mathbf{B}) \xrightarrow{\mathbf{v}} (\mathbf{E}', \mathbf{B}')$$

而暂时取消关于电势与矢势的定义。很难想像如何在电场可能有旋，磁场可能有散的情形下依然承认：

$$\mathbf{E} = -\nabla\varphi$$

$$\mathbf{B} = \nabla \times \mathbf{A}$$

因为我们记得梯度一定无旋，旋度一定无散：

$$\nabla \times \nabla\varphi = 0 \quad , \quad \nabla \cdot \nabla \times \mathbf{A} = 0$$

于是现在问题变成了，要研究通过此前式子定义的二分量电磁场 (\mathbf{E}, \mathbf{B}) 是如何依赖于产生它们的电荷的运动形式的（注意，后面还讲证明存在不由电荷产生的电磁场）。历史背景上最早被发现的性质莫过于法拉第电磁感应定律(Faraday's law of induction)，它的一个现代版本的表述为：

$$\oint_{\partial V} (E_x dt dx + E_y dt dy + E_z dt dz + B_x dy dz + B_y dz dx + B_z dx dy) = 0$$

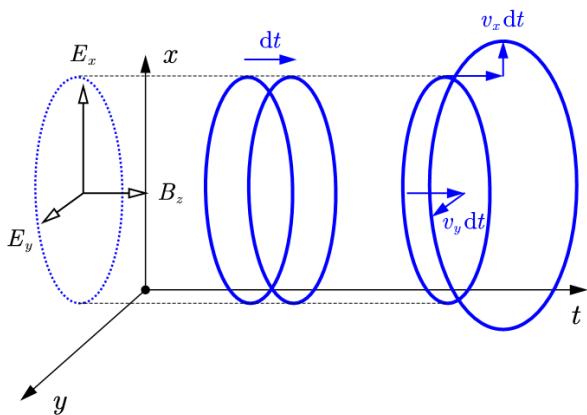


图 22.1: 法拉第电磁感应定律

其中 ∂V 表示任意一个四维时空中一个三维体的边界。而如果取这个体为一个三维空间中静止的二维曲面以一个非常短的时间间隔 dt 为高形成的柱体造成的边界，上式就变成了我们可能更熟悉的式子：

$$\left(\oint_{\partial A} \mathbf{E} \cdot d\mathbf{l} \right) dt + \int_{t+dt, A} \mathbf{B} \cdot d\mathbf{A} - \int_{t, A} \mathbf{B} \cdot d\mathbf{A} = 0$$

$$\Rightarrow \oint_{\partial A} \mathbf{E} \cdot d\mathbf{l} = - \frac{d\Phi_A}{dt}$$

这个式子被当时的诸多物理学家（法拉第，亨利，楞次等）在实验室里测量验证。它无可辩驳地证明了在变化的电磁场下，电场将会成为有旋的场，过去理论的根基无疑发生了动摇。

十分有意思的是，另一个相关的结果也同时被研究了：当我们考虑的不是一个四维空间的母线沿时间方向的柱体，而是如图22.1那样上底下底不一样（但依然垂直于时间轴）的台状体，实际上就是一个在运动的线圈（不限于平移，可以扩大缩小或任意弯曲）。设线圈边缘的速度为 v ，那么在侧面也会造成一个磁场积分的分量，上式就会变成：

$$\Rightarrow \oint_{\partial A} \mathbf{E} \cdot d\mathbf{l} = - \frac{d\Phi_A}{dt} - \oint_{\partial A} (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}$$

$$\Rightarrow \oint_{\partial A} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = - \frac{d\Phi_A}{dt}$$

上式左侧恰好又变成了完整的洛伦兹力（单位电荷受到的）。事实上在实验室它也只能作为一个整体被测量，上式左侧其实就是实验室测得的线圈上的电动势，两项分别构成了感生电动势与动生电动势，右侧则是线圈包含的磁通量对时间的导数。从而一个更复杂但也更浅显易懂的法拉第电磁感应定律表述为：

$$\mathcal{E} = \oint \mathbf{K} \cdot d\mathbf{l} = - \frac{d\Phi}{dt}$$

$$\mathcal{E} = \mathcal{E}_{\text{transformer}} + \mathcal{E}_{\text{motional}}$$

$$\mathcal{E}_{\text{motional}} = \oint \mathbf{E} \cdot d\mathbf{l}$$

$$\mathcal{E}_{\text{transformer}} = \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}$$

但是，法拉第电磁感应定律又不全然是实验定律，背后体现的相对论协变性是不难理解的。我们从本章的论述就会逐渐明白这一点。

22.1.1 动生电动势

动生电动势(motional electromotive force) 并不是什么新奇的现象。在静磁场下我们足以理解动生电动势，它只不过是上一章就研究过的磁场力的另一个效应，此时存在非静电力：

$$\mathbf{K} = \mathbf{v} \times \mathbf{B}$$

考虑它沿导线方向的分量对回路积分就是此前在稳恒电流章节中介绍的电动势:

$$\mathcal{E} = \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}$$

上式经常会被表述为: 导线单位时间切割的磁感线根数. 这样的表述的合理性在于, 注意到上式中的三重标积可以化为:

$$(\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = \mathbf{B} \cdot (d\mathbf{l} \times \mathbf{v})$$

点乘后就是导线元扫过的面元, 故积分的项的确也就表示扫过的面积上的磁通量, 从而不难理解这类电动势自动就符合法拉第电磁感应定律:

$$\mathcal{E} = -\frac{d\Phi}{dt}$$

如果考虑导线中的电荷沿导线方向的另一个速度造成的电流 I , 那么其电流元就会导致一个安培力:

$$d\mathbf{F} = Id\mathbf{l} \times \mathbf{B}$$

一段电路动生电动势做功的功率:

$$dP_{\mathcal{E}} = d\mathcal{E} \cdot I = I(\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}$$

该功率变成了电路内部的能量. 是注意到安培力作为磁场力的另外一个分力也要做功, 功率为:

$$dP_A = d\mathbf{F} \cdot \mathbf{v} = I(d\mathbf{l} \times \mathbf{B}) \cdot \mathbf{v}$$

这个功将转化为导线的机械能. 那么由于磁场力本质上不会做功, 这一点很容易就能够利用三重标积公式在两个功率上得到印证:

$$dP_{\mathcal{E}} + dP_A = 0$$

从而两个功率的大小代表机械能和电能转化的快慢. 对于电动机, 电能转化为机械能, 对于发电机, 机械能转化为电能.

事实上即使对于变化的电磁场, 其中磁场也是永远不做功的. 这一点从磁场的操作定义式就可以看出来, 磁场力作为速度与磁场的叉乘是垂直于速度的. 但要注意, 一般不说洛伦兹力不做功, 因为电场力也要算作是洛伦兹力的一部分, 洛伦兹力没有道理不包括电场力.

当动生电磁感应发生在一个变化的电磁场中运动的线圈上时, 根据 $\mathbf{v} \times \mathbf{B}$ 在线圈上的积分给出了整个感应电动势贡献中的一项, 实际磁通量的导数才对应总的感应电动势, 而动生电动势相当于把磁场视作常数求导得到的值:

$$\mathcal{E}_{\text{motional}} = -\left. \frac{d\Phi}{dt} \right|_{B \text{不变}}$$

22.1.2 感生电动势

动生电动势并没有带来新的物理现象. 但是它的结果却给我们以启示: 磁通量的改变与导线中的感应电动势有关. 那么只需要稍微思考一下另外的参考系中观察者看到的现象就不难推理出另一种感应电动势的可能.

相对性原理容易说明, 如果产生磁场的物质在发生运动以致使其发生变化, 那么即使线圈静止, 也会产生感应电动势. 那么内部电荷受到的非静电力就必然不是磁场力, 因为根据电磁场的定义:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

而现在线圈 $\mathbf{v} = \mathbf{0}$. 我们只能把这个力称作电场力. 这种电场称作感生电场(induced electric field). 有时, 它被理解为变化的磁场产生的, 这其实是因为在线圈看来电动势公式应该¹是:

$$\oint_{\partial A} \mathbf{E} \cdot d\mathbf{l} = -\frac{d\Phi}{dt} = -\int_A \frac{\partial \mathbf{B}}{\partial t} \cdot dA$$

¹在忽略某些可能的相对论效应的意义下.

这个式子的直接推论就是:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

这个式子与电流产生磁场的数学结构相似性:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}$$

暗示了把磁场的改变看作电场的一类源似乎是合适的, 确实从操作性上可以这么做, 甚至原来静磁学中的一切电流产生磁场的结论都可以原封不动地用来求变化的磁场下的感生电场. 但事实上, 我们还有两个更加本质的看法.

一是把感生电场看作静磁场的相对论效应, 这又和上一章静磁场开头时把磁场本身看作电场的相对论效应的做法如出一辙. 原参考系中的速度为 v_x, v_y, v_z , 而具有磁场 B_x, B_y, B_z 时受力为:

$$F_x = q(v_y B_z - v_z B_y)$$

$$F_y = q(v_z B_x - v_x B_z)$$

$$F_z = q(v_x B_y - v_y B_x)$$

这一次倒好, 我们连近似都不做了, 直接认为一个相对原系以 u 沿 x 方向运动的参考系下受力就是原系中的受力:

$$F'_x = F_x, \quad F'_y = F_y, \quad F'_z = F_z$$

用经典的速度叠加:

$$v_x = v'_x + u, \quad v_y = v'_y, \quad v_z = v'_z$$

得到:

$$F'_x = q(v'_y B_z - v'_z B_y)$$

$$F'_y = -quB_z + q(v'_z B_x - v'_x B_z)$$

$$F'_z = quB_y + q(v'_x B_y - v'_y B_x)$$

从而由洛伦兹力公式, 的确在新参考系中产生了电场:

$$E'_x = 0, \quad E'_y = -uB_z, \quad E'_z = uB_y$$

这实际上就是:

$$\mathbf{E} = \mathbf{u} \times \mathbf{B}$$

故如果一个点相对一个稳恒电流体系以 $-\mathbf{v}$ 运动, 即稳恒电流体系相对这个点具有均匀的速度 \mathbf{v} 时, 这个点产生的涡旋电场就是:

$$\mathbf{E} = -\mathbf{v} \times \left(\int \frac{\mu_0 \mathbf{j} \times \mathbf{e}_R}{4\pi R^2} dV \right)$$

而如果电流的各个部分不一致, 由叠加原理也可以得到表达式:

$$\mathbf{E} = \int \frac{\mu_0 (\mathbf{j} \times \mathbf{e}_R) \times \mathbf{v}}{4\pi R^2} dV$$

我们不加证明地指出上式可以通过矢量微分的方法化简成我们马上要得到的形式, 这就是第二种涡旋电场产生的条件的理解方式: 变化的磁场事实上又由变化的电流产生. 最终连接感生电场与其源头: 变化的电流的, 是以下公式:

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} = - \int \frac{\mu_0 \frac{\partial \mathbf{j}}{\partial t}}{4\pi r^2} dV$$

请注意, 我们又非常奇妙地定义了一个磁矢势。这是因为我们依然将磁场体系视作准恒(quasistatic)的。即磁场主导的情况下, 磁场场源运动速度远小于光速以至于产生的电场 $E \ll cB$ 。从而磁矢势就被“近似”定义为与静磁场一样的电流的体积分:

$$\mathbf{A} = \int \frac{\mu_0}{4\pi} \frac{\mathbf{j}}{r^2} dV$$

唯一要注意的是, 由于 $\nabla \cdot \mathbf{A} = 0$ 依赖于电流的稳恒性条件, 以至于它和 $\nabla^2 \mathbf{A} = -\mu_0 \mathbf{j}$ 都从原理上不再准确。将上式对回路进行积分便得到同样的感生电动势公式作为其正确性的一个验证:

$$\mathcal{E} = \oint \mathbf{E} \cdot d\mathbf{l} = -\frac{\partial}{\partial t} \oint \mathbf{A} \cdot d\mathbf{l} = -\frac{d\Phi}{dt}$$

我们趁热也来看看准恒情况下关于电场电势的定义。诚然我们不能为真实的电场定义一个电势 φ 使它满足:

$$-\nabla \varphi = \mathbf{E}$$

因为此时的电场是涡旋的。但是仿照之前磁矢势的“近似”定义, 我们自然也可以自然地“近似”定义电标势为:

$$\varphi = \int \frac{1}{4\pi\epsilon_0} \frac{\rho}{r^2} dV$$

那么, 留下的一个矛盾就变成了:

$$\mathbf{E} \neq -\nabla \varphi$$

这也没啥古怪的, 实际上在准恒条件下“近似”成立²的一个式子为:

$$\mathbf{E} = -\nabla \varphi - \frac{\partial \mathbf{A}}{\partial t}$$

这样就把电场分为了两个部分, 一个是只与每时每刻的净电荷分布有关的“静电场”分量, 一个是由电流的改变造成的“涡旋电场”分量。而电势总是定义在前者上的。

事实上我们此前稳恒电流章节大量使用到了电路中各点的电势, 在涡旋电场遍布电路中的各个支路时, 这个电势实际上也就是我们现在所阐明的电势概念。更特别地, 把 $-\frac{\partial \mathbf{A}}{\partial t}$ 视作非静电力 \mathbf{K} , 而如果电路中电荷除了受到静电力 $-\nabla \varphi$ 和非静电力 \mathbf{K} 外不受其他非静电力, 即有微观欧姆定律 $\mathbf{E} = \rho \mathbf{j}$, 在准恒或稳恒的条件下对支路进行积分, 就得到了部分电路的欧姆定律:

$$IR = \Delta\varphi + \mathcal{E}$$

如果将上式乘以电流, 就得到能量等式:

$$I^2 R = UI + \mathcal{E}I$$

尤其是对回路而言, 电压一圈下来没有改变:

$$I^2 R = \mathcal{E}I$$

那么其焦耳热来源为感生电磁感应的电动势对电流的做功 $\mathcal{E}I$ 。因为是电场所以是可以做功的。但是其实这个感生涡旋电场本身作为磁场变化的产物而能量远低于磁场能量, 所以其实最终能量来自于磁场里的能量。下一章我们将讨论电磁场整体的能量。现在不妨这么想: 涡旋电场对感应线圈上的电流做了功, 那么就必然存在运动的场源电流, 而它们又置身于感应线圈上的电流产生的磁场中, 从而移动它们需要外力做额外的功, 很多情况下正是这个功在补偿从磁场中通过电磁感应漏向电场的能量。

²例如让体系以小量角频率 ω 发生振动, 那么下式在 ω 一阶下精确。

当感生电磁感应发生在一个变化的电磁场中运动的线圈上时, 根据 $\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}$ 在线圈上的积分给出了整个感应电动势贡献中的一项, 实际磁通量的导数才对应总的感应电动势, 而感生电动势相当于把线圈视作不动求导得到的值:

$$\mathcal{E}_{\text{transformer}} = -\left. \frac{d\Phi}{dt} \right|_{A \text{不变}}$$

将感应电动势的动生和感生部分合并, 得到完整的法拉第电磁感应定律:

$$\mathcal{E} = -\left. \frac{d\Phi}{dt} \right|_{B \text{不变}} - \left. \frac{d\Phi}{dt} \right|_{A \text{不变}} = -\frac{d\Phi}{dt}$$

22.2 自感与互感

- 自感定义与自感系数公式:

$$\Psi = LI \quad , \quad L = \frac{\mu N^2 S}{l}$$

- 两线圈互感时的相关公式:

$$E = \frac{1}{2}L_1 I_1^2 + \frac{1}{2}L_2 I_2^2 + MI_1 I_2$$

$$\begin{aligned} \Psi_1 &= L_1 I_1 + MI_2 \quad \Rightarrow \quad U_1 = -L_1 \frac{dI_1}{dt} - M \frac{dI_2}{dt} \\ \Psi_2 &= L_2 I_2 + MI_1 \quad \Rightarrow \quad U_2 = -L_2 \frac{dI_2}{dt} - M \frac{dI_1}{dt} \\ U_1 I_1 + U_2 I_2 &= -\frac{dE}{dt} \end{aligned}$$

第二十三章 麦克斯韦方程组

23.1 麦克斯韦方程组

23.1.1 电生磁

磁荷不存在,为电磁学规律带来了充分的简化.但是目前我们发现的三种现象:静止电荷产生电场,稳恒电流产生磁场,变化电流(通过产生变化的磁场)产生涡旋电场,并不是电磁学客观规律的全部.前两章我们已经充分说明,后两种现象其实都是相对性原理的后果:在一个参考系内不变的电荷分布,必将在另一个参考系看来产生电流,稳恒电流也会因为在另一个参考系而随时间变化,从而在一个参考系中纯粹的电场可以在另一个参考系内引发磁场和涡旋电场.电磁学的最后一项现象,由麦克斯韦(J. C. Maxwell)在1861提出的,位移电流产生磁场的现象,在某种程度上也由相对性原理引发.我们之后就可以发现这一点.

一切的源起,在于麦克斯韦对安培环路定律普遍性的质疑.我们不能认为在变化的荷与场的情况下,磁场的安培环路定律具有普适性.它写作:

$$\oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_S \mathbf{j} \cdot d\mathbf{S}$$

为何质疑?考虑一个电容器不断充电的过程,如图所示.我们可以认为在距离电容较远的某点处,取一个垂直于导线的半径足够小的圆盘 S_1 后,边界环路上的磁场与导线上的电流之间符合安培环路定律的关系.但是注意到如果我们选择一个兜状的面 S_2 时,它的边界可以与圆盘边界完全一致.但是由于它把导线和一个极板完全兜住,故这个面上完全没有电流.从而就不能认为这个面和它的边界符合安培环路定律!至少, S_1 和 S_2 绝对不可能同时符合安培环路定律!这是因为,等号左侧这两个面边界上具有同一个磁场的环路积分,但是等号右侧却是不一样的电流通量: S_1 上为导线电流 I 而 S_2 上为零.两个等式是不可能同时成立的.

但是空间中的磁场又的确是客观存在的,而不是虚无缥缈的辅助物理量,我们依然可以通过对带电粒子受到的洛伦兹力的要求来确定空间中某处的磁场值:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

那么如何寻找这种情况下正确反映磁场计算方法的公式?麦克斯韦做到了这一点,而他的做法其实可以被简单的归纳为三点:

逻辑上要求自洽,数学上发展工具,哲学上遵从吝啬原则(law of parsimony)¹.

首先第一点,安培环路定律从逻辑上已经不能自洽了,所以必须修改它,一个合理的想法是为它增添一项:

$$\oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_S \mathbf{j} \cdot d\mathbf{S} + (?)$$

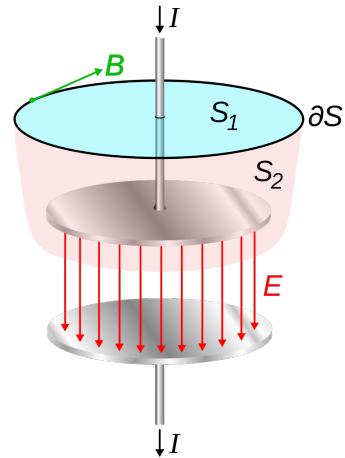


图 23.1: 修正安培环路定律

¹又做奥卡姆剃刀 (Occam's razor): 如无必要,勿增实体.认为基本规律的不同表述之间越简单的描述约能反应基本规律的本质.

增添这一项的目的就是为了能避免如之前电容器充电情况下的矛盾情况的发生. 事实上电容器充电的例子就提供了修改之后的表达式是否正确的试金石.

为了找到这个项, 矢量微分工具就发挥了不可替代的作用. 使用微分形式以上积分等式写作:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + (?)$$

再两边同时取散度, 左边数学上一定变成零, 从而得到添加的项需要符合的条件:

$$\mu_0 \nabla \cdot \mathbf{j} + \nabla \cdot (?) = 0$$

再注意到电荷电流符合连续性方程:

$$\nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = 0$$

从而得到:

$$\nabla \cdot (?) = \mu_0 \frac{\partial \rho}{\partial t}$$

为了实现这一点, 最简单的一种可能性为:

$$(?) = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

这是因为电场的散度就会回到电荷:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0} \Rightarrow \nabla \cdot \left(\mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) = \mu_0 \varepsilon_0 \frac{\partial}{\partial t} (\nabla \cdot \mathbf{E}) = \mu_0 \frac{\partial \rho}{\partial t}$$

但是我们注意到, 使得之前的要求成立的添加项 (?) 不是唯一的, 事实上在刚刚的结果上添加任意一个矢量场的旋度都不改变其散度的值:

$$\nabla \cdot \left(\mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{F} \right) = \mu_0 \frac{\partial \rho}{\partial t} + 0$$

于是最后一个思想就发挥了作用: 如无必要, 勿增实体. 往往简单而自洽的形式就能反映真实的物理规律. 从而麦克斯韦就利用这样的推理得到了修正的安培环路定律:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

同时麦克斯韦提出了以下视角去解释以上式子背后的物理图像: 我们定义位移电流(displacement current):

$$\mathbf{j}_d = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

那么我们的工作可以归纳为一句话: 通过引入虚构位移电流, 导致与真实电流的和形成一个没有散度(即连续)的电流场:

$$\nabla \cdot (\mathbf{j} + \mathbf{j}_d) = 0$$

举例来说, 在之前的电容器充电的场合下, 如果忽略电容器的边缘效应, 那么从上方往下通过的电流为 I 时, 电容器上积累的面电荷密度 σ 的增长为:

$$I = \dot{\sigma} S$$

从而电容内部的场强就会随之增长:

$$E = \frac{\sigma}{\varepsilon_0} \Rightarrow \dot{E} = \frac{I}{\varepsilon_0 S}$$

从而把电场的导数视作位移电流的话, 造成的总电流就是:

$$I_d = j_d S = \varepsilon_0 \dot{E} S = I$$

从而实际上补充位移电流后电流并没有在电容器极板上中断,这样也就不会造成之前电容器上不同面上电流不同而造成的矛盾情况.

如果认为补充这一项后形成的新安培环路定律具有在任意变化的电磁场下的普适性,我们可以发现,这样的一条规律在数学形式上具有与感生电场规律具有某种“对称性”.尤其是当空间中某处没有通过实际电流时,此时两条规律分别为:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \times \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

即,变化的磁场(部分来源于变化的电流分布)会产生电场,变化的电场(部分来源于变化的电荷分布)会产生磁场.

我们曾经解释过第一个式子,它实际上可以解释为动生电动势现象的相对论效应.即在一个参考系中的磁场会在另一个参考系中招致电场:

$$\mathbf{E}' = \mathbf{v} \times \mathbf{B}$$

而一个参考系中的电场在另一个参考系中招致磁场这样的对称行为,实际上我们在之前静磁学章节中建立毕奥-萨伐尔定律时就已经有介绍过,得到的结果为:

$$\mathbf{B}' = -\frac{\mathbf{v} \times \mathbf{E}}{c^2}$$

这就不难看出两者之间的对称性来.而第一个式子,我们通过一个参考系中的动生电动势与另一个参考系中的感生电动势都与法拉第电磁感应定律相联系的思路,就推出了电场的旋度公式:

$$\mathbf{E}' = \mathbf{v} \times \mathbf{B} \Rightarrow \oint_{\partial S} \mathbf{E} \cdot d\mathbf{l} = - \int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \Rightarrow \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

注意到系数 $1/c^2$ 实际上就是 $\mu_0 \epsilon_0$,完全仿照这个过程,我们也能得到磁场的环量也将等于电通量的导数,即对应有:

$$\mathbf{B}' = -\frac{\mathbf{v} \times \mathbf{E}}{c^2} \Rightarrow \oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \epsilon_0 \int_S \frac{\partial \mathbf{E}}{\partial t} \cdot d\mathbf{S} \Rightarrow \nabla \times \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

所以至此我们可以总结:电磁学的所有规律,都可以看做唯一一个规律:即库仑定律,在相对论原理下分化出来的丰富变式.

23.1.2 麦克斯韦方程组

将四个规律:电荷产生电场,电流产生磁场,变化的磁场产生电场,变化的电场产生磁场组合到一起,就得到了麦克斯韦方程(Maxwell equations),实际上指以下方程组²:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

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

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}$$

它完整而全面地描述了场与荷之间的关系.在之前的观点中,我们认为电荷是因,场是结果.这种观点之后将不再适用,这是因为在数学形式上,变化的电场和磁场可以互相产生,之后我们利用这一点能找到完全

²但是其实这四个方程是一个整体,正如电磁场本身应当视作一个整体对待

脱离电荷存在的电磁波解. 而完整的电磁理论实际上只需要再配合电场磁场本身的定义, 即洛伦兹的力公式:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

然而, 同样的一个现象 (电荷与场之间的相互作用) 可以有不同理论来描述. 力学我们就曾经介绍过牛顿经典力学体系和尝试过分析力学的可能性. 静电学和静磁学中, 电场和磁场不仅可以由场强来描述, 我们也研究了由势来描述的可能性. 当时的势与电荷的关系是比较简单的:

$$\varphi = \int \frac{1}{4\pi\epsilon_0} \frac{\rho}{r} dV, \quad \mathbf{A} = \int \frac{\mu_0 \mathbf{j}}{4\pi r} dV$$

那么能不能为变化的电磁场整体采取一个用势来描述的方法? 至少以上两个式子应当被摒弃, 因为他们只描述了电荷直接产生的电场与电流直接产生的磁场. 而我们知道变化的电荷与电流可以交叉地产生磁场和电场. 事实上之前成立的以下两个描述势与场强之间关系的式子也应当考虑摒弃:

$$\mathbf{E} = -\nabla\varphi, \quad \mathbf{B} = \nabla \times \mathbf{A}$$

后一个式子不会招致明显的矛盾, 因为即使在变化电磁场下, 依然有:

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

故无需添加新项即可成为恒等式.

矛盾集中体现在第一条上. 两边同时取旋度就能看出其中的不妥之处:

$$\nabla \times \mathbf{E} = -\nabla \times \nabla\varphi$$

其中左侧根据麦克斯韦方程, 它等于磁场的对时间变化率, 它不需要为零. 但是右侧根据数学规律它必然为零. 这就造成了矛盾.

解决这个问题的核心思想, 其实与麦克斯韦发现位移电流的思想完全一致. 我们添加一个项以使得矛盾消除, 以相同的思路推理, 过程留给读者, 可得这个项为:

$$\mathbf{E} = -\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t}$$

这一点本身又不难理解. 我们知道完整的电磁理论有四个要素, 在最低阶的近似下之前的式子反映了电荷直接产生电场与电流直接产生磁场的要素. 那么我们添加的项实际上反映了变化的电流产生电场的要素: 因为磁矢势由电流产生, 那么它的导数自然反映了变化的电流, 而它构成了电场的第二项.

而变化的电荷产生磁场, 它不会在新的场与势的关系式中体现, 这是因为这样的改变直接同时改变了 \mathbf{B} 和 \mathbf{A} 的值. 前者通过位移电流, 后者之后介绍. 从而使得 $\mathbf{B} = \nabla \times \mathbf{A}$ 依然成立. 从而我们总结出任意情况下描述电磁场的标势和矢势需要符合的关系:

$$\mathbf{E} = -\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}$$

但是, 接下来我们会引入一个惊人的不确定性: 其中的端倪在静电静磁学中其实已经出现苗头: 电势, 具有一个整体常数的不确定性, 这是因为电势的零点选取其实本质上是任意的, 以下变换后产生的标量场依然可以作为电势:

$$\varphi'(\mathbf{r}) = \varphi(\mathbf{r}) + C$$

只有进一步规定无穷远处的电势为零才能回到之前点电荷积分的电势公式:

$$\varphi(\infty) = 0$$

而静磁场的情况则更具有戏剧性: 我们之前通过电流元直接积分产生的磁矢势其实符合以下性质:

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

当然以上两个结果都还要求势不随时间改变, 比如:

$$\frac{\partial \varphi}{\partial t} = 0$$

如果仅仅希望能通过磁矢势的旋度计算磁场 $\nabla \times \mathbf{A} = \mathbf{B}$, 即使进一步要求 $\nabla \cdot \mathbf{A} = 0$, \mathbf{A} 的取法其实也不唯一. 可以证明, 如果找到所谓的三维调和函数, 即在全空间满足以下规律的标量场:

$$\nabla^2 \phi = 0$$

常数就属于调和函数, 除此之外我们还能举出很多的例子:

$$\phi = x, x^2 - y^2, xy, x^2 + y^2 - 2z^2, e^x \sin y, \dots$$

那么任意找一个调和函数梯度形成的矢量场:

$$\mathbf{F} = \nabla \phi$$

那么不难看出这个场就会符合:

$$\nabla \times \mathbf{F} = \mathbf{0}, \quad \nabla \cdot \mathbf{F} = 0$$

从而以下对磁矢势的变换就可以在不改变对磁矢势的要求下, 造成同一个磁场:

$$\mathbf{A}' = \mathbf{A} + \mathbf{F}$$

$$\nabla \cdot \mathbf{A}' = \nabla \cdot \mathbf{A} = 0 \quad \& \quad \nabla \times \mathbf{A}' = \nabla \times \mathbf{A} = \mathbf{B}$$

我们从其中归纳出两个概念. 一是我们要求电势和矢势根据具体问题不同选取不同的条件, 上面过程中出现的各种要求就有:

$$\begin{aligned} \nabla \cdot \mathbf{A} &= 0 \\ \frac{\partial \varphi}{\partial t} &= 0 \\ \varphi(\infty) &= 0 \end{aligned}$$

这些称作**规范条件**(gauge condition). 或者简单称作**规范**.

但是, 即使存在规范条件, 例如静磁场下要求 $\nabla \cdot \mathbf{A} = 0$, 或者进一步放宽干脆没有任何规范条件的限制, 我们发现电势和矢势其实不是位移确定的, 同样的一个电磁场, 在下式成立的条件下:

$$\mathbf{E} = -\nabla \varphi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}$$

势 φ, \mathbf{A} 存在不同的取法, 普遍来说, 存在以下**规范变换**(gauge transformation):

$$\varphi' = \varphi + \phi, \quad \mathbf{A}' = \mathbf{A} + \mathbf{F}$$

它能使得在特定规范下虽然描述场的势改变, 但是它不反映场的真实改变. 即同一个场对应多种相差规范变换的势的取法. 这种现象称作**非物理自由度**(nonphysical dof.). 它深刻反映了标势矢势不能直接反映可测量的物理量的实质, 正似我们在经典力学中一贯强调的势能的道理: 其本身绝对大小没有可测量的意义, 可以测量的是其在过程中的变化, 它才有实际物理意义. 但是, 对于标势矢势的哪一部分具有实际的物理意义还有待之后进行考察. 选取更多的规范可以消除势的不确定性, 这种操作称作**规范固定**(gauge fixing).

这就不难理解在普遍的电磁场方程中, 同样也存在着标势和矢势的不确定性. 更有甚者, 正如电场强度和磁场强度是描述同一个电磁场的两个侧面, 那么标势和矢势也是描述同一个电磁场的两个侧面, 两者的不确定性甚至是互相联系的. 如果没有任何实现假定的规范条件, 我们发现选取任意的随时间变换的标量场做变换:

$$\forall \phi = \phi(\mathbf{r}, t)$$

$$\varphi' = \varphi - \frac{\partial \phi}{\partial t} , \quad \mathbf{A}' = \varphi + \nabla \phi$$

可以发现它不改变通过之前式子计算的电场和磁场的强度. 从而这是一个非常普遍的规范变换.

而对于普遍的电磁理论, 它也是囊括静电学和静磁学的普遍情况的. 可以想见, 在不同的规范条件下就会产生不同的势. 例如对于静电学, 之前我们通过标势来描述电场, 这一点就不是必然的. 如果选取所谓的外尔规范 (Weyl gauge):

$$\varphi = 0$$

那么电场就必须全权由矢势 \mathbf{A} 来描述, 但这一点也毫无难度, 只需要取:

$$\mathbf{A} = - \int \frac{1}{4\pi\epsilon_0} \frac{\rho e_r}{r^3} dV \cdot t$$

这就足以使得:

$$\mathbf{E} = -\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t} = \int \frac{1}{4\pi\epsilon_0} \frac{\rho e_r}{r^3} dV , \quad \mathbf{B} = \nabla \times \mathbf{A} = \mathbf{0}$$

事实上, 外尔规范是在研究磁现象占主导地位时常采用的规范, 此时电磁场整体都单由矢势 \mathbf{A} 来描述, 电场可以形象地视作由矢势变化产生的“感生电场”. 而在电现象占主导地位时, 更常采用库仑规范 (Coulomb gauge):

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

可以证明这样做带来的好处是标势的计算变得很简单: 只需要把电磁场和势之间的关系代入麦克斯韦方程, 我们得到两个恒等式和以下两个联系势和电荷电流的复杂关系式:

$$\nabla^2\varphi + \frac{\partial \nabla \cdot \mathbf{A}}{\partial t} = -\frac{\rho}{\epsilon_0}$$

$$\nabla^2\mathbf{A} - \nabla(\nabla \cdot \mathbf{A}) - \epsilon_0\mu_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} - \epsilon_0\mu_0 \frac{\partial \nabla\varphi}{\partial t} = -\mu_0\mathbf{j}$$

代入规范条件, 首当其冲地我们发现: 第一个式子变成了:

$$\nabla^2\varphi = -\frac{\rho}{\epsilon_0}$$

这和静电学中的公式一致. 从而电势可以完全按照静电学的方式来计算, 它产生一个静电场. 而矢势则用来计算电荷分布变化带来的附加磁场和更高阶的感生电场.

但是, 除了静电或静磁现象占主导地位的两种情况, 我们之后研究的最多的是第三种情况: 那便是辐射占主导地位的情形. 此时无论是电荷密度或是电流密度, 要么其量级远远达不到辐射场对应的大小, 要么就是在一周期内平均值为零从而净效应很弱. 在这一类情况下, 使用的最多的是所谓的洛伦茨规范³ (Lorenz gauge). 写作:

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \varphi}{\partial t} = 0$$

这个神奇的条件使得之前描述势和电荷之间的关系变得非常简单, 背后原因在于可以证明这个条件也是唯一的在相对论变换下不同参考系内形式一致的规范条件. 在这个规范下:

$$\nabla^2\varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = -\frac{\rho}{\epsilon_0}$$

$$\nabla^2\mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0\mathbf{j}$$

这实际上构成了带非齐次项的波动方程. 一般称作达朗贝尔方程 (d'Alembert equations).

由于其最为简洁而普遍的形式. 我们之后的讨论都将以洛伦茨规范为默认规范, 并在以上达朗贝尔方程和关于荷与场强的麦克斯韦方程, 以及用势计算场强的关系式基础上展开讨论.

³这里的洛伦茨 (L. V. Lorenz) 不同于洛伦兹 (H. A. Lorentz), 前者为丹麦物理学家, 后者为荷兰物理学家

23.1.3 电荷在电磁场中的运动

在静磁学中我们简要介绍了磁矢势对于电荷与磁场相互作用的物理意义. 电势为引入场的试探电荷带来相互作用的电势能. 而磁矢势的改变似乎会改变试探电荷的某类动量. 通过之前的讨论, 我们发现电势和矢势其实内部具有非物理的自由度, 从而并不总是为试探电荷带来可以直接观测的物理量, 之前的电势带来电势能的情况, 其实仅仅是特殊情况下成立的结论, 并不具有普适性.

但是电势对应电势能的推导思路是依然可以借鉴的: 我们考虑了洛伦兹力的功, 并把电磁场用势来表示, 依然这么做给出:

$$\begin{aligned}\mathbf{F} \cdot \mathbf{v} &= q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \mathbf{v} \\ &= q\mathbf{E} \cdot \mathbf{v} \\ &= q\left(-\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t}\right) \cdot \mathbf{v}\end{aligned}$$

我们考虑两个关系, 其一为:

$$\frac{d\varphi}{dt} = \frac{\partial\varphi}{\partial t} + \mathbf{v} \cdot \nabla\varphi$$

其背后的道理我们在力学中阐释过: 第一项为电荷运动过程中达到的点具有的电势的全导数, 它由不考虑电荷位移时电势随时间的局部导数和不考虑电势本身随时间改变而仅考虑随电荷位移导致的电势改变的随体导数项合成.

第二个关系为:

$$\frac{\partial}{\partial t}(\mathbf{v} \cdot \mathbf{A}) = \mathbf{v} \cdot \frac{\partial}{\partial t}\mathbf{A}$$

背后的道理为: 当我们考虑 $\mathbf{v} \cdot \mathbf{A}$ 整体作为时空 r, t 的函数时, \mathbf{v} 被我们视作了外来的参变量而不是时空的函数, 从而能够从偏导数符号中提出.

通过这两点, 我们得到了:

$$\mathbf{F} \cdot \mathbf{v} = -q\frac{d}{dt}\varphi + q\frac{\partial}{\partial t}(\varphi - \mathbf{v} \cdot \mathbf{A})$$

而左侧应当是电荷动能 T 的变化率. 那么这就给出了:

$$\frac{d}{dt}(T + q\varphi) = \frac{\partial}{\partial t}q(\varphi - \mathbf{v} \cdot \mathbf{A})$$

事实上, 单纯从洛伦兹力公式出发:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

考虑到等号左侧为粒子动量 \mathbf{p} 的变化率, 我们也可以通过推导 (静磁场那一章我们推导过没有电场的对应结果) 给出以下式子:

$$\frac{d}{dt}(\mathbf{p} + q\mathbf{A}) = -\nabla q(\varphi - \mathbf{v} \cdot \mathbf{A})$$

这两个结论引出了三组物理概念:

一是隶属于能量的概念: T 为动能, $q\varphi$ 现在可称作电磁势能. 而其和 $H = T + q\varphi$ 应当称作广义能量.

二是隶属于动量的概念: \mathbf{p} 为动量, $q\mathbf{A}$ 就是我们新发现的一个物理效应, 事实上磁矢势的物理意义可以理解为与位置有关的存储于场中的粒子的动量, 称作电磁势动量. 而机械动量 \mathbf{p} 与势动量 $q\mathbf{A}$ 的和 $\mathbf{P} = \mathbf{p} + q\mathbf{A}$, 称作广义动量.

事实上大家已经发现, 这种称呼其实就是来源于分析力学中的相应理论架构, 事实上电磁场中的电荷运动这一个实例我们之前分析力学正是以它举过例子, 读者可以对照着理解. 故其实我们还可以把 $H = T + q\varphi$ 称作哈密顿量(hamiltonian), 而 $\mathbf{P} = \mathbf{p} + q\mathbf{A}$ 称作正则动量(cannonical momentum).

最后还有一个概念, 它出现于之前两个等式的右侧, 它也是一种势能函数, 但是并不对应真实能量, 而是出现在电荷运动的拉格朗日量中作为势能函数. 这个项称作广义势:

$$V = q(\varphi - \mathbf{v} \cdot \mathbf{A})$$

其表达式中反常地出现了粒子的速度, 是一种速度势.

以往出现的守恒律与对称性现在就在这两个表达式中体现的淋漓尽致:

$$\frac{dH}{dt} = \frac{\partial V}{\partial t}$$

$$\frac{d\mathbf{P}}{dt} = -\nabla V$$

举例来说: 静电场中带电粒子的运动, 粒子总能量守恒; 静磁场中带电粒子的运动, 粒子干脆动能都不改变. 这都是源于广义势不随时间改变 (时间平移对称性导致能量守恒). 而在静磁场问题中还经常出现某方向动量与某个关于位置的函数的和守恒的情况, 究其本质是因为合理选取规范后, 磁矢势 \mathbf{A} 和电势 φ 都与某个位置坐标无关 (空间平移对称性导致动量守恒).

23.2 平面电磁波

23.2.1 真空中的电磁波

自由空间中的麦克斯韦方程:

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

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

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

$$\nabla \times \mathbf{B} = \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}$$

也可以用势来描述, 此时达朗贝尔方程直接变成波动方程:

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = 0$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = \mathbf{0}$$

注意背后要求洛伦茨规范成立:

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \varphi}{\partial t} = 0$$

十分使得注意的是, 以上所有的方程, 作为齐次方程, 是具有非零解的. 最典型的解为平面波解:

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}$$

$$\mathbf{B} = \mathbf{B}_0 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}$$

$$\mathbf{A} = \mathbf{A}_0 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}$$

$$\varphi = 0$$

$$\frac{E_0}{B_0} = c \quad , \quad \mathbf{E}_0 \perp \mathbf{k} \perp \mathbf{B}_0 \quad , \quad \mathbf{E}_0 \times \mathbf{B}_0 \parallel \mathbf{k}$$

$$\mathbf{E}_0 = -i\omega \mathbf{A}_0 \quad , \quad \mathbf{B}_0 = -ik \times \mathbf{A}_0$$

$$\omega/k = c$$

23.2.2 介质中电磁波的传播

注意在介质中,不能再仿照真空中的电磁波传播公式而写出齐次的麦克斯韦方程.这是因为介质内部完全可以存在电荷和电流,从而计算电场磁场的散度旋度时必然需要考虑到他们.但是我们此前研究过的电介质和磁介质的理论告诉我们,通过引入位移 \mathbf{D} 和磁场强度 \mathbf{H} ,可以有效地规避掉由介质产生的电荷和电流造成的电场散度和磁场旋度,即:

$$\begin{aligned}\nabla \cdot \mathbf{D} &= 0 \\ \nabla \times \mathbf{H} &= \frac{\partial \mathbf{D}}{\partial t}\end{aligned}$$

对于第二式可能需要更多的说明:我们知道如果在静磁学磁介质问题中,如果把总电流密度 \mathbf{j} 分解为磁化电流密度 \mathbf{j}_M 和自由电流密度 \mathbf{j}_f ,那么磁场旋度公式写作:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} = \mu_0 \mathbf{j}_M + \mu_0 \mathbf{j}_f$$

以往的做法是将磁化电流归因于磁化强度的不均匀性 $\mathbf{j}_M = \nabla \times \mathbf{M}$,并进一步定义 $\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}$,这样就给出:

$$\nabla \times \mathbf{H} = \mathbf{j}_f$$

但是现在作为自由介质 $\mathbf{j}_f = 0$.然而又由于是变化的电磁场,故磁场旋度需要增添电场变化带来的项.还有一点需要引起注意:极化强度 \mathbf{P} 的变化也会带来一个极化电流密度 \mathbf{j}_P ,它应当为:

$$\mathbf{j}_P = \frac{\partial \mathbf{P}}{\partial t}$$

对比极化电荷密度 $\rho_P = -\nabla \cdot \mathbf{P}$ 与极化电荷的连续性方程就不难理解这个式子.另外引人注目的一点是,这个电流与极化强度的关系完全类似与位移电流与电场强度的关系.如果做求和就会变成电位移的导数,这正是麦克斯韦当初把电场强度的导数称作“位移电流”的原因.

现在修改之后的磁场旋度写作:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \varepsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{j}_M + \mu_0 \frac{\partial \mathbf{P}}{\partial t} + \varepsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}$$

最后两项十分恰好地凑出了电位移 \mathbf{D} .从而整理就得到了介质中修正的齐次磁场旋度公式.

在普遍的介质中的麦克斯韦方程就写作:

$$\begin{aligned}\nabla \cdot \mathbf{D} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \frac{\partial \mathbf{D}}{\partial t}\end{aligned}$$

在这里介质兼具电与磁的特性,这表现在以下两个用于描述介质的结构方程上:

$$\mathbf{D} = \varepsilon \mathbf{E}$$

$$\mathbf{B} = \mu \mathbf{E}$$

其中 ε, μ 称作介质的电容率和磁导率.它们,超出我们以往的认知的是,往往是复数⁴.这一方面来自于微观极化和磁化机理上的弛豫,我们之后物理光学还将进一步研究这个问题,还有一方面来自于一个非常直接的原因:我们综合了不导电的电介质和磁介质,却忽视了介质本身完全可能导电的情况.对于导电性能较

⁴非各向同性介质中甚至抽象为复张量

差的情形,比如漏电的介质,考虑传导电流 j_c 实际上就是自由电流,那么磁场旋度的方程还需要被进一步修改为:

$$\nabla \times \mathbf{H} = j_c + \frac{\partial \mathbf{D}}{\partial t}$$

如果是考虑单一频率的电磁波,那么对时间求导的操作必然转变为乘 $i\omega$. 如果介质电导率为 σ ,那么可以把等号右侧进行以下变形:

$$j_c + \frac{\partial \mathbf{D}}{\partial t} = \sigma \mathbf{E} + i\omega \varepsilon \mathbf{E} = i\omega \left(\varepsilon - i\frac{\sigma}{\omega} \right) \mathbf{E} = i\omega \mathbf{D}' = \frac{\partial \mathbf{D}'}{\partial t}$$

这样就通过定义复电容率(complex permittivity):

$$\varepsilon' = \varepsilon - i\frac{\sigma}{\omega}$$

实现了把传导电流也吸收进了电位移中,造成了形式上的统一. 代价是电容率和相关的其他所有物理量都需要理解为复数.

这样的做法对于漏电绝缘介质是完全适用的. 但是对于导电性能非常好的导体介质中电磁波的传播,最好是单独建立新的处理方法. 有兴趣的读者可以阅读其他相关书籍. 我们指出,我们统一用复数描述不同情况下介质中电磁波传播的做法,从之后的结果上来看是符合事实的.

把介质的两个结构方程代入介质的齐次麦克斯韦方程,就得到了均匀介质中一般形式的麦克斯韦方程,习惯上总是以 \mathbf{E}, \mathbf{H} 作为变量:

$$\begin{aligned}\nabla \cdot \mathbf{E} &= 0 \\ \nabla \times \mathbf{E} &= -\mu \frac{\partial \mathbf{H}}{\partial t} \\ \nabla \cdot \mathbf{H} &= 0 \\ \nabla \times \mathbf{H} &= \varepsilon \frac{\partial \mathbf{E}}{\partial t}\end{aligned}$$

这两个方程具有高度的对称性,其解也与真空中的解相似,只有两点发生了改变,一是传播速度变为:

$$v = \frac{\omega}{k} = \frac{1}{\sqrt{\varepsilon \mu}}$$

这一点建立在 ε, μ 都是实数的基础上. 如果出现复数的情形,那么上式一般写作:

$$k = \sqrt{\varepsilon \mu} \omega$$

即此时将出现复波矢的情况. 意味着电磁波在介质中的传播伴随着耗散与衰减.

第二点就是电场磁场振幅强度的比例. 真空中这个比为 c ,现在由以下等式描述:

$$\varepsilon E_0^2 = \mu H_0^2$$

对于所谓良导体(good conductor)的情况,电导率 σ 特别大,以至于使得复电容率 ε' 中反倒是与电导率成正比的虚部远大于反应介电的实部.那么电磁波在良导体中的传播就存在严重的衰减. 此时有两个结论:

第一是对电磁场比值的公式,取 $\varepsilon' \rightarrow \infty$,就能发现:

$$E_0 \rightarrow 0$$

即良导体中类似于静电平衡情况下的导体,内部电场是非常微弱的. 磁场相对电场虽然很大,但是也需要迅速沿传播方向衰减.

第二是 k 的公式有以下渐进关系:

$$\frac{2\pi}{\lambda} \sim k \propto \sqrt{\frac{1}{\omega}} \cdot \omega = \sqrt{\omega}$$

即随着频率升高,传播与衰减的特征距离 λ 将越来越小. 这意味着高频交流电路中电流仅仅集中于导线表面薄薄的一层,会减小导线的有效截面积而增大线路损耗. 这个现象称作趋肤效应(skin effect).

23.3 电磁场能量与动量

真空中电磁场的能量密度为:

$$w = \frac{1}{2}\varepsilon_0 E^2 + \frac{1}{2}\mu_0 H^2$$

能流密度, 即著名的坡印亭矢量(Poynting vector) 为:

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}$$

而动量密度与能流密度紧密联系:

$$\mathbf{g} = \mathbf{D} \times \mathbf{B} = \frac{1}{c^2} \mathbf{E} \times \mathbf{H} = \frac{\mathbf{S}}{c^2}$$

23.4 电磁波辐射

23.4.1 电磁辐射概论

麦克斯韦方程实际上描述了两类现象:

一类我们已经通过之前所有章节逐一学习. 表面上看包括静止的电荷产生电场, 稳恒的电流产生磁场, 变化的电荷分布与电流分布产生感应的磁场和电场四类. 但是其实无非都是第一类由库仑定律描述的基本规律在相对论变换下的衍生物. 在这一大类现象中, 场始终都是电荷的附庸, 基本的逻辑是先有荷, 再有场. 空间分布上, 场源电荷对远方的场的支配量级服从平方反比律.

通过本章的推理我们逐渐看到第二类截然不同的图像存在的可能性. 它诞生于同一个麦克斯韦方程, 但是故事不再由电荷主演, 主角阵容中新增了一颗闪亮的明星: 辐射场(radiation field). 平面波实际上就是形形色色辐射场中最单纯的一种. 显然, 它可以独立于电荷而存在. 但是依然, 电荷也愿意主动参与进麦克斯韦方程搭建的舞台中, 与辐射场这样的强势主角又将擦出怎样的火花? 简单来说, 我们将得到以下两个新的现象:

电荷发出辐射, 电荷吸收辐射.

凭借电荷吸收辐射的可能性, 第二类电磁现象也与第一类现象区分开来. 更何况电荷激发辐射场的空间分布也区别于第一类现象, 在辐射情形下, 远方的场是一次反比的. 这一点之后就能发现.

最后我们针对电荷发出辐射与吸收辐射的条件进行推理: 静止的点电荷产生静电场, 匀速运动的点电荷产生一个静电场的相对论变换, 这是毫无疑问的事实. 在这些场合下没有辐射场的产生.

所以, 显然电荷做加速运动是电荷激发辐射场的必要条件.

实际上, 电荷做加速运动其实即使就是电荷激发辐射场的充要条件, 我们之后就能给出电荷辐射场与辐射功率与其加速度的直接联系.

那么如何看待电荷吸收辐射的过程?

事实上这个过程可以看做电荷发射辐射的时间反演. 麦克斯韦方程具有明确的时间反演对称性, 只需要令 $t \rightarrow -t, \mathbf{B} \rightarrow -\mathbf{B}, \mathbf{j} \rightarrow -\mathbf{j}$. 其他量不改变就可以看出这一点. 辐射作为麦克斯韦方程的解, 其时间反演解自然也是合理的. 但需要注意先后发生的两件事总是存在因果的区别, 故时间反演后, 因果关系需要重新解释, 两个物理现象需要理解为不具有相通性⁵的独立现象. 我们可以举一个电荷吸收辐射的例子: 考虑电子的经典模型: 将一个平面电磁波照射到真空中的自由电子上, 那么电子就会在周期性的电场力(速度不大时磁场力可忽略)作用下做简谐振动, 那么根据下一小节的内容我们将知道, 电子会以一定功率向四面八方产生辐射场. 那么根据能量守恒, 电子必然是从原来入射的平面波中吸收了能量, 即平面波在一个周期内对电子做净的正功. 这种先吸收再辐射的行为就称作散射(scattering), 而以上经典的处理方法构成托马斯散射模型(Thomas scattering model). 而散射的总功率, 对应了原来平面波面积为 σ 范围内通过的功率. 这个面

⁵例如, 不能通过洛伦兹变换来颠倒因果.

积称作**散射截面**(scattering cross section). 经典理论计算结果为:

$$\sigma = \frac{8\pi}{3} r_e^2$$

其中 r_e 就是由洛伦兹提出的**经典电子半径**(classical electron radius):

$$r_e = \frac{e^2}{4\pi\epsilon_0 m_e c^2}$$

23.4.2 偶极辐射

通过点电荷产生的电势公式, 我们知道, 数学上以德爾塔函数作为泊松方程的非齐次项的解为一次反比势:

$$\nabla^2 \varphi = -\delta(\mathbf{r}) \Rightarrow \varphi = \frac{1}{4\pi r}$$

数学上容易猜想并证明, 如果将方程替换为达朗贝尔方程, 而如果将非齐次项以 ω 的角频率产生振荡, 那么这个新的方程的解应当变为球面波:

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = -\delta(\mathbf{r}) e^{i\omega t} \Rightarrow \varphi = \frac{e^{i(\omega t \pm kr)}}{4\pi r}, \quad k = \frac{\omega}{c}$$

对于两个可能的解, + 数学上表示内行球面波, - 数学上表示外行球面波. 两个解在时间反演下对称. 目前我们在因果性上是考虑电荷体系的自发辐射, 故考虑负号的外行波解.

现在考虑一个这样的体系: 两个无限小的导体球用长为 l 的导线连接, 第一个球位于 $x = \frac{l}{2}$ 处, 第二个导体球位于 $x = -\frac{l}{2}$ 处. 由于两个球发生了 LC 振荡, 第一个球上电荷量为 $q_+ = q \cos \omega t$, 第二个球上电荷量为 $q_- = -q \cos \omega t$. 那么导线上的电流为 $I = -\omega q \sin \omega t$. 如果换成复数记法, 把实际值看做复数的实部, 则:

$$q_+ = q e^{i\omega t}, \quad q_- = -q e^{i\omega t}, \quad I = i\omega q e^{i\omega t}$$

定义从第二个球指向第一个球的矢量为 \mathbf{l} . 那么此时标势和矢势符合的达朗贝尔方程就写作:

$$\begin{aligned} \nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} &= -\frac{q}{\epsilon_0} \delta\left(\mathbf{r} - \frac{\mathbf{l}}{2}\right) e^{i\omega t} + \frac{q}{\epsilon_0} \delta\left(\mathbf{r} + \frac{\mathbf{l}}{2}\right) e^{i\omega t} \\ \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} &= -i\mu_0 \omega q l \delta(\mathbf{r}) e^{i\omega t} \end{aligned}$$

那么标势和矢势的解就根据之前的数学结论得到:

$$\begin{aligned} \varphi &= e^{i\omega t} \cdot \frac{q}{4\pi\epsilon_0} \frac{e^{-ik|\mathbf{r}-\frac{\mathbf{l}}{2}|}}{|\mathbf{r}-\frac{\mathbf{l}}{2}|} - e^{i\omega t} \cdot \frac{q}{4\pi\epsilon_0} \frac{e^{-ik|\mathbf{r}+\frac{\mathbf{l}}{2}|}}{|\mathbf{r}+\frac{\mathbf{l}}{2}|} \\ \mathbf{A} &= e^{i\omega t} \cdot \frac{i\mu_0 \omega q l}{4\pi} \frac{e^{-ikr}}{r} \end{aligned}$$

这样的解称作**推迟势**(retarded potential) 解. 而对于电势可以进一步在足够远 $r \gg l$ 处进行小量近似, 我们不再考虑两项分母的区别, 这只会带来类似于电偶极子远方的电势那样的平方反比的高阶小量. 而主要关注分子的差:

$$\begin{aligned} e^{-ik|\mathbf{r}-\frac{\mathbf{l}}{2}|} - e^{-ik|\mathbf{r}+\frac{\mathbf{l}}{2}|} &\approx \frac{de^{-ikr}}{dr} \cdot \left(\left| \mathbf{r} - \frac{\mathbf{l}}{2} \right| - \left| \mathbf{r} + \frac{\mathbf{l}}{2} \right| \right) \\ &\approx ik l \cos \theta e^{-ikr} \end{aligned}$$

其中 θ 为 \mathbf{r} 方向与 \mathbf{l} 的夹角. 则在以 \mathbf{l} 方向为极轴的球坐标下, 写出近似后的标势与矢势:

$$\begin{aligned} \varphi &= e^{i(\omega t - kr)} \cdot \frac{ikql \cos \theta}{4\pi\epsilon_0 r} \\ A_r &= e^{i(\omega t - kr)} \cdot \frac{i\mu_0 \omega ql \cos \theta}{4\pi r}, \quad A_\theta = -e^{i(\omega t - kr)} \cdot \frac{i\mu_0 \omega ql \sin \theta}{4\pi r} \end{aligned}$$

最后再由势与场之间的关系:

$$\mathbf{E} = -\nabla\varphi - \frac{\partial\mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}$$

忽略超过一阶反比于半径的项, 就最终得到了电磁场的表达式:

$$\begin{aligned}\mathbf{E} &= E_\theta \mathbf{e}_\theta, \quad E_\theta = -\frac{\omega^2 q l \sin \theta}{4\pi\epsilon_0 c^2 r} e^{i(\omega t - kr)} \\ \mathbf{B} &= B_\varphi \mathbf{e}_\varphi, \quad B_\varphi = -\frac{\omega^2 q l \sin \theta}{4\pi\epsilon_0 c^3 r} e^{i(\omega t - kr)} = \frac{E_\theta}{c}\end{aligned}$$

从而取实部, 计算瞬时能流密度为:

$$S = \frac{\omega^4 q^2 l^2 \sin^2 \theta}{16\pi^2 \epsilon_0 c^3 r^2} \cos^2(\omega t - kr)$$

在一个周期内的平均辐射能流为:

$$\bar{S} = \frac{\omega^4 q^2 l^2 \sin^2 \theta}{32\pi^2 \epsilon_0 c^3 r^2}$$

最后计算全立体角内的总辐射功率:

$$P = \int_0^\pi \bar{S} \cdot 2\pi r^2 \sin \theta d\theta = \frac{\omega^4 q^2 l^2}{12\pi \epsilon_0 c^3}$$

考虑到辐射等价于由两个电荷量恒定, 但在 $\pm l/2$ 范围内做简谐振动, 加速度振幅为 $a = \omega^2 l/2$ 的电荷共同贡献. 那么每一个电荷加速度幅值 a 对应的辐射幅值为:

$$P = \frac{q^2 a^2}{6\pi \epsilon_0 c^3}$$

这就是著名的拉莫尔公式(Larmor formula).

第二十四章 交流电路

24.1 相量表示

24.1.1 拟稳条件与交流元件

- 拟稳条件: 一是忽略传输线上的推迟, 即传输线的分布电容电感电阻电导都被忽略, 如果不忽略同一导线两端电流就会不一致且导线将分去电压. 二是忽略元件内部的推迟. 不忽略会导致元件非理想.
- 拟稳条件对应数学表述:

$$\lambda \gg l \quad , \quad T \gg \frac{l}{c}$$

- 交流电路瞬时值描述: 每一瞬时类似直流电路, 但是电压电流在变化:

$$u = U_0 \cos(\omega t + \varphi_u)$$

$$i = I_0 \cos(\omega t + \varphi_i)$$

直流电路中的图, KVL, KCL 均成立. 但是多了电容电感两种元件, 其特性之后介绍.

- 有效值:

$$U = U_0/\sqrt{2} \quad , \quad I = I_0/\sqrt{2}$$

有效值存在意义是为了直接算平均功率:

$$P = I^2 R = U^2 / R = UI$$

- 交流电路的相量描述:

$$\tilde{U} = U e^{j\varphi_u}$$

$$\tilde{I} = I e^{j\varphi_i}$$

24.1.2 电阻, 电容, 电感特性

- 复阻抗:

$$\tilde{Z} = \frac{\tilde{U}}{\tilde{I}}$$

两种分解方法:

$$\tilde{Z} = |\tilde{Z}| e^{j \arg \tilde{Z}} = Z e^{j\delta}$$

Z 称为阻抗, δ 为余损耗角.

$$\tilde{Z} = R + iX$$

R 称为电阻, X 称为电抗.

- 复导纳:

$$\tilde{Y} = \frac{\tilde{I}}{\tilde{U}}$$

两种分解方法:

$$\tilde{Y} = |\tilde{Y}| e^{j \arg \tilde{Y}} = Y e^{j\delta}$$

Y 称为导纳, δ 为余损耗角.

$$\tilde{Y} = G + iB$$

G 称为电导, B 称为电纳.

- 电阻

$$\tilde{Z} = R$$

- 电感

$$\tilde{Z} = j\omega L$$

$Z_L = \omega L$ 为感抗. 任何阻抗若电抗为正称为感抗.

- 电容

$$\tilde{Z} = -\frac{j}{\omega C}, \quad \tilde{Y} = j\omega C$$

$Z_C = \omega C$ 为容抗. 任何阻抗若电抗为负称为容抗.

- 元件功率分为视在功率, 有功功率, 无功功率.

视在功率:

$$S = UI = I^2 Z$$

有功功率:

$$P = UI \cos \delta = I^2 R = \operatorname{Re}(\tilde{U} \tilde{I}^*) = \frac{1}{2} (\tilde{U} \tilde{I}^* + \tilde{U}^* \tilde{I})$$

无功功率:

$$P' = UI |\sin \delta| = I^2 |X|$$

三者关系:

$$P' = P \tan |\delta|, \quad S^2 = P^2 + P'^2$$

损耗角:

$$\tan \varepsilon = \frac{P}{P'} = \frac{R}{|X|}, \quad \pm \delta = \frac{\pi}{2} - \varepsilon$$

24.2 常见电路

24.2.1 谐振电路

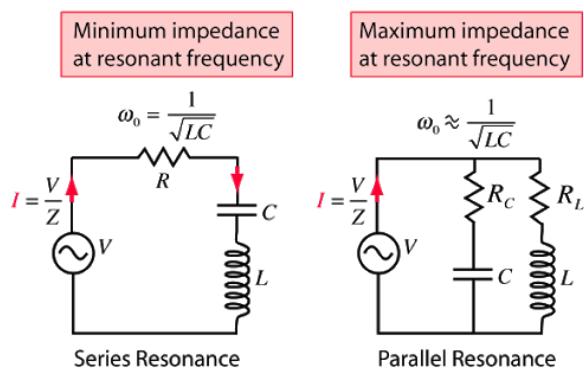


图 24.1: 串联谐振与并联谐振

- 串联谐振:

$$\tilde{Z} = R + j \left(\omega L - \frac{1}{\omega C} \right)$$

谐振条件: 电流作为响应与电压无相位差:

$$\omega = \frac{1}{\sqrt{LC}}$$

此时恰也有最小阻抗, 电流最大.

品质因数 Q :

$$Q = \frac{U_L}{U_R} = \frac{\omega_0 L}{R} = \frac{\omega_0}{\Delta\omega}$$

- 并联谐振: 谐振一般定义为相位谐振点, 即电流作为响应与电压无相位差. 此时阻抗一般接近最大值, 故干路电流反而较小. 支路电路往往远大于干路电流. 振幅达最小值的振幅谐振点一般不与相位谐振点重合.

第二十五章 相对论力学

25.1 相对论运动学

- 光速不变原理: 时空的结构为平直时空: $ds^2 = c^2 dt^2 - dx_i^2$: 物质运动的“舞台”.
- 相对性原理: 物理规律的协变性:

$$\text{Sca.} = \text{Sca.}, \quad \text{Vec.} = \text{Vec.}, \quad \text{Vec.} \cdot \text{Vec.} = \text{Sca.} \text{ etc.}$$

- 四-标量**(four-scalar) 与**四-矢量**(four-vector) 的定义: 类比三维空间标量矢量, 但按闵可夫斯基空间处理.
- 洛伦兹变换的定义: 保度规的变换. 物理上是不同共原点惯性参考者建立的参考系.
- 洛伦兹群:

$$\text{SO}(3, 1) = \langle R_x, R_y, R_z, B_x, B_y, B_z \rangle$$

其中 B_x 为 x 方向的**推促**(boost):

$$\begin{bmatrix} ct \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}' = \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} ct \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

- 经典物理没有协变性, 但电磁学与相对论理论具有协变性. 意为所有物理定律都由协变的对象构成, 这主要包括标量, 矢量, 张量, 旋量四类. 四-矢量尤其常见:

$$\begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{bmatrix}' = \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{bmatrix}$$

出于方便, 四-矢量也被简记为 $A_\mu = (A_0, A_i) = (A_0, \mathbf{A})$. 而 $A_i = (A_1, A_2, A_3)$ 是三维矢量的简记. 四-矢量与四-矢量可以算**伪内积**(pseudo-inner-product):

$$A_\mu B^\mu = A_0 B_0 - A_i B_i$$

伪内积将得到四-标量. 即参考系变换下不变的量.

- 洛伦兹变换下任意四-矢量具有不变量 $A_\mu A^\mu = A_0^2 - \mathbf{A}^2$, 对于类时四-矢量这将大于零, 即 $A_\mu A^\mu = A_{\text{proper}}^2$, 事实上以速度 $\mathbf{v} = \frac{\mathbf{A}}{A_0}c$ 做洛伦兹变换, 可以把空间分量变为零. 这叫做**本征参考系**(proper reference system). 得到本征四-矢量 $A'_\mu = (A_{\text{proper}}, \mathbf{0})$.
- 三个基本图像:
 - 尺缩: $L = L_0/\gamma$.
 - 钟慢: $t = \gamma t_0$.
 - 同时相对性: $\tau = -\beta \frac{x}{c}$.

只要是真的尺子, 真的钟, 就永远不可能错! 没有真的尺子和钟就去构造.

- 相对论质点运动:

具有四-标量: 间隔 $ds = \sqrt{dx_\mu dx^\mu}$, 本征时 $dt_0 = ds/c$, 下四-速度的不变模长 $v_\mu v^\mu = c^2$.

具有四-矢量: 四-位移 $dx_\mu = (cdt, d\mathbf{r})$, 四-速度 $v_\mu = (\gamma c, \gamma \mathbf{v})$, 四-加速度见下节.

- 速度变换公式:

$$\gamma' = \frac{1 - \frac{uv_x}{c^2}}{\sqrt{1 - u^2/c^2}} \gamma$$

$$v'_x = \frac{v_x - u}{1 - \frac{uv_x}{c^2}}$$

$$v'_{y,z} = \frac{\sqrt{1 - u^2/c^2}}{1 - \frac{uv_x}{c^2}} v_{y,z}$$

- 光行差公式: 可以用速度变换推导, 也可以之后用流变换推导:

$$\cos \theta' = \frac{\cos \theta - \beta}{1 - \beta \cos \theta}, \quad \sin \theta' = \sin \theta \cdot \frac{\sqrt{1 - \beta^2}}{1 - \beta \cos \theta}, \quad \tan \theta = \frac{\sin \theta'}{\cos \theta'}$$

- 相对论下讨论刚体? 模型存在缺陷: 隧道佯谬, 转盘佯谬. 故整个刚体运动学都没有必要建立.

25.2 相对论动力学

- 动量-能量构成四矢量的必要性: 守恒律需要被保留且协变.

- 四-动量:

$$p_\mu = \left(\frac{E}{c}, \mathbf{p} \right), p_\mu p^\mu c^2 = E^2 - p^2 c^2 = m^2 c^4, \mathbf{v} = \frac{\mathbf{p} c^2}{E}$$

$$\mathbf{p} = \frac{m \mathbf{v}}{\sqrt{1 - v^2/c^2}}, E = \frac{mc^2}{\sqrt{1 - v^2/c^2}}$$

- 普遍的过程的守恒律:

$$\sum_i p_{i,\mu} = \sum_j p_{j,\mu}$$

- 四-力与四-加速度:

$$a_\mu = \frac{dv_\mu}{dt_0} = \left(\gamma \frac{d\gamma}{dt} c, \gamma \frac{d\gamma \mathbf{v}}{dt} \right), \frac{d\mathbf{v}}{dt} := \mathbf{a}$$

$$F_\mu = \frac{dp_\mu}{dt_0} = \left(\gamma \frac{dE}{cdt}, \gamma \frac{d\mathbf{p}}{dt} \right) := \left(\gamma \frac{W}{c}, \gamma \mathbf{F} \right)$$

- 推论一: $F_\mu = a_\mu$. 但三维矢量之间不会直接符合 Newton-like 的动力学方程.
- 推论而: $F_\mu v^\mu = 0$. 这个三维继续可以得到:

$$\mathbf{F} dt = d\mathbf{p}, \mathbf{F} \cdot d\mathbf{r} = dE$$

- 单方向受力时, 在该方向的推促下力不变.
- 本征加速度与本征力: 它本意指的是四-速度的本征系中的四-加速度与四-力, 不过神奇的是恰好让两者的时间分量变为零 (类空, 与类时的四-速度正交), 空间分量为 $\mathbf{a}_0, \mathbf{F}_0$:

$$\mathbf{F}_0 = m\mathbf{a}_0$$

$$a_\mu a^\mu = -a_0^2$$

- 沿平行于本征加速度与本征力方向施以推促:

$$F = F_0, \quad a = a_0/\gamma^3$$

- 沿垂直于本征加速度与本征力方向施以推促:

$$F = F_0/\gamma, \quad a = a_0/\gamma^2$$

- 快度(rapidity): $\beta = \text{th}\vartheta, \gamma = \text{ch}\vartheta$.
- 自然坐标下的相对论性牛顿定律:

$$F_\tau = \gamma^3 m a_\tau$$

$$F_n = \gamma m a_n$$

- 相对论关于碰撞的新理解:

情形	经典	相对论
弹性	无耗散, $\mathbf{p}' = \mathbf{p}, E' = E$	粒子守恒, $\mathbf{p}' = \mathbf{p}, E' = E$
非弹性	有耗散, $\mathbf{p}' = \mathbf{p}, E$ 不可列	粒子变性, $\mathbf{p}' = \mathbf{p}, E' = E$

只有一个例外: 光子的散射若频率不变, 叫做弹性散射. 如经典的瑞利散射, 米氏散射; 但如果频率变了, 叫做非弹性散射, 如拉曼散射, 康普顿散射.

- 相对论碰撞根据具体情形一般有以下大的分类: 弹性的散射, 非弹性的散射, 粒子反应.
- 质点系的动量-能量也将构成一个四-矢量, 即:

$$(E, \mathbf{p}) = \sum_i (E_i, \mathbf{p}_i)$$

那么这个四-矢量的本征系速度:

$$\mathbf{v} = \frac{\mathbf{p}c^2}{E}$$

称作动量中心系 (center-of-momentum frame), 以代替经典物理的质心系. 简称动心系. 顾名思义, 此系中质点系总动量为零: 而总能量:

$$E_0^2 = E^2 - \mathbf{p}^2 c^2 = m_0^2 c^4$$

m_0 为动心系中的总动质量, 以后称作质点系的等效静质量. 这个式子可以解释一部分的质量起源问题.

- 一个等效静质量 m_0 的质点系若选取相对动心系以 v 运动的参考系, 动量与能量为:

$$\mathbf{p} = \sum_i \mathbf{p}_i = \frac{m_0 \mathbf{v}}{\sqrt{1 - v^2/c^2}}$$

$$E = \sum_i E_i = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}}$$

- 放能反应: $\sum_i m_i > \sum_j m_j$. 则无论取哪个系, 静能转化为动能的量都相等. 定义为反应能 Q :

$$Q = \left(\sum_i m_i - \sum_j m_j \right) c^2$$

- 吸能反应: $\sum_i m_i < \sum_j m_j$. 则具有阈能. **仅在动心系中**, 反应发生的最小动能 (阈能) 的值为:

$$Q = \left(\sum_j m_j - \sum_i m_i \right) c^2$$

25.3 相对论连续物质

- 若有一守恒荷 Q 作为四-标量, 则其空间分布与流动构成一个密度场 ρ, \mathbf{j} :

$$dQ = \rho dV$$

$$dQ = \mathbf{j} \cdot d\mathbf{S} dt$$

- 那么 $(\rho c, \mathbf{j})$ 是一个四-矢量且符合洛伦兹变换, 如 x 方向的推促:

$$\begin{bmatrix} \rho' c \\ j'_x \\ j'_y \\ j'_z \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \rho c \\ j_x \\ j_y \\ j_z \end{bmatrix}$$

- 若有一平面波场, 其必要的一个要素为相位, 它随着时空分布必然写作:

$$\varphi = \omega t - \mathbf{k} \cdot \mathbf{r}$$

ω 称作角频率, \mathbf{k} 为波矢.

- 那么由于相位为四-标量, 且可以写作以下形式, 故 $k_\mu = (\omega/c, \mathbf{k})$ 为四-矢量, 称作四-波矢:

$$\varphi = k_\mu x^\mu$$

如 x 方向的推促:

$$\begin{bmatrix} \omega'/c \\ k'_x \\ k'_y \\ k'_z \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega/c \\ k_x \\ k_y \\ k_z \end{bmatrix}$$

- 可以从上式出发得到光行差公式, 同时也能得到多普勒效应公式. 在前提 $\omega/k = c$ (四-波矢类光) 时:

$$\omega' = \omega \cdot \frac{1 - \beta \cos \theta}{\sqrt{1 - \beta^2}}$$

$$\cos \theta' = \frac{\cos \theta - \beta}{1 - \beta \cos \theta}$$

- 将三维的力密度和功率密度:

$$\mathbf{f} = \frac{d\mathbf{F}}{dV}$$

$$p = \frac{dP}{dV}$$

合并可以得到四-力密度 $f_\mu = (p/c, \mathbf{f})$.

第二十六章 光的干涉

26.1 标量波理论

标量波是指以下标量偏微分 **波动方程**(wave equation) 的解:

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) A(\mathbf{r}, t) = 0$$

分别在直角坐标, 柱坐标, 球极坐标下用所谓的分离变量法可以得到一些正确的基础解¹:

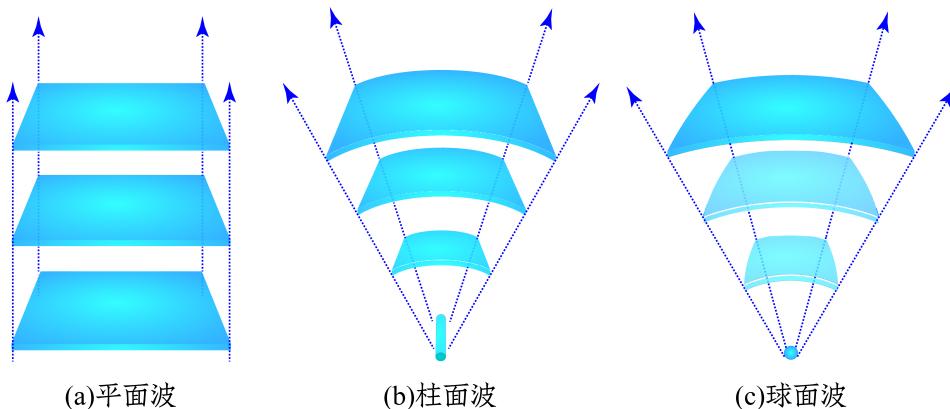


图 26.1: 三种典型的波

$$1. \ A(\mathbf{r}, t) = A \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi)$$

要求 $\frac{\omega}{|\mathbf{k}|} = c$. 此为传播方向为 k 的平面行波.

$$2. \ A(\mathbf{r}, t) = \frac{I}{\sqrt{\rho}} \cos(\omega t - k\rho + \phi)$$

要求 $\frac{\omega}{|k|} = c$. $k > 0$ 时为向外传播的柱面行波, k 小于零则向内传播.

$$3. A(\mathbf{r}, t) = \frac{I}{r} \cos(\omega t - kr + \phi)$$

要求 $\frac{\omega}{|k|} = c \cdot k > 0$ 时为向外传播的球面行波, k 小于零则向内传播.

容易发现,以上波全都具有这样的形式:

$$A(\mathbf{r}, t) = |A|(\mathbf{r}) \cos(\omega t - \varphi(\mathbf{r}))$$

也就是说,每个点的相位都在以共同的 ω 做等速率的增加,代表一种同步的振动.而振动幅度 $|A| > 0$ 是要由于波的传播特点和能量流动的连续性而缓慢变化的.具体传播方向则看相位的梯度:

$$\boldsymbol{k} = +\nabla \varphi$$

¹以下三式都要求 $\omega > 0$ ，数学上看 $\omega < 0$ 也没错，但是利用三角函数的性质可以将 t 系数变号而不带来新的解。

在近似的意义上, $\frac{\omega}{|\mathbf{k}|} = c$ 是没有问题的. 而 c 便是波传播的速度. 在真空与介质 (写作 v) 中出现在光的波动方程中的 $\frac{1}{c^2}$ 中的 c 其实是:

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} , \quad v = \frac{c}{n}, \quad n = \sqrt{\epsilon_r \mu_r} \approx \sqrt{\epsilon_r}$$

约等于符号是因为考虑到了透明光学介质一般磁性质不显著故 $\mu_r \approx 1$.

我们不是很推荐用三角函数来计算波动光学. 要写成复指数的形式, 而实际三角函数理解为它的有效的实部, 虚部是为了方便计算而添加的没有物理效应的. 而且波动光学统一约定, 利用三角函数的形式, 为相位增添一个负号, 使得光在同一刻沿传播方向相位是增加的, 改为:

$$A(\mathbf{r}, t) = |A|(\mathbf{r}) e^{i(\varphi(\mathbf{r}) - \omega t)} = |A|(\mathbf{r}) e^{i\varphi(\mathbf{r})} e^{-i\omega t} = A(\mathbf{r}) e^{-i\omega t}$$

在以后情形下我们都将把这种符号约定称为“光学符号约定”, 而以往的指数上的宗量随时间增加随传播方向减小的符号约定称为“电磁学符号约定”. 以上讨论就引出了复振幅 $A(\mathbf{r})$ 的概念. 而之前找到的各点振幅 $|A|$ 和 φ 现在就被整合在了一起, 作为复振幅的模与幅角:

$$|A|(\mathbf{r}) = |A(\mathbf{r})| , \quad \varphi(\mathbf{r}) = \arg A(\mathbf{r})$$

回过头来再看, 我们在解波动方程时, 实际上从一开始就没有必要认为该标量 $A(\mathbf{r}, t)$ 是一个实数, 在复数域内解, 也能够直接得到完全与上述等效一样的结果. 最后我们发现, 该复数标量场 $A(\mathbf{r}, t)$ 如果能够符合波动方程, 其实部与虚部也是能够分别满足波动方程的:

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) A(\mathbf{r}, t) = 0 \Rightarrow \begin{cases} \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \Re A(\mathbf{r}, t) = 0 \\ \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \Im A(\mathbf{r}, t) = 0 \end{cases}$$

如果再要求这样的 $A(\mathbf{r}, t)$ 具有形式 $A(\mathbf{r}, t) = A(\mathbf{r}) e^{-i\omega t}$, 数学上也就是对变量 t 分离变量, 物理就意味着研究单色光的传播, 那么代入便会发现实际上 $A(\mathbf{r})$ 需要满足的对空间的方程为 **亥姆霍兹方程** (helmholtz equation):

$$(\nabla^2 + k^2) A(\mathbf{r}) = 0 , \quad k = \frac{\omega}{c}$$

这就是整个波动光学理论的核心方程. 在继续阐述之前我们先说说 A 指什么. 其实真空与线性介质 (不导电) 电磁场几乎每个量 (标势, 矢势, 电场, 磁场) 都是符合波动方程的:

$$A : \{\varphi, A_x, A_y, A_z, E_x, E_y, E_z, B_x, B_y, B_z\}$$

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) A(\mathbf{r}, t) = 0 , \quad (\nabla^2 + k^2) A(\mathbf{r}) = 0$$

所以作为复数的标量波理论就构成了光干涉, 衍射等波动光学课题的数学基础. 而且我们率先注意到其中的两点:

一是, 波动方程, 亥姆霍兹方程是线性方程, 也就是说, 如果两个光场都符合方程, 那么两个场直接逐点做标量加法得到的场也会自动符合方程. 这也就是之后干涉, 衍射操作的重要前提. 如果由于两个原因同时产生了两束光波, 那么只要写出各自的场, 加在一起便是整个体系现在的光场. 这就是双光束干涉时我们将会采取的做法.

二是, 平面波与球面波的重要性. 只要是行波, 在任意一点就可以按照平面波来近似. 如果取其 k 方向为 z 方向, 那么局部的光场为:

$$A(\mathbf{r}) = A e^{ikz}$$

但是旁边的点 (x_1, y_1) 的最佳平面波的近似却不一定平行于 z 轴, 它应该为:

$$A(\mathbf{r}) = A e^{i[k_z z + k_x(x - x_1) + k_y(y - y_1)]}$$

这下便有第一个近似条件了, 我们叫做**傍轴条件**(paraxial condition). 它要求光线传播方向与 z 轴夹一个小角度. 此时我们习惯把 x 和 y 方向方向余弦 $\cos \alpha_x, \cos \alpha_y$ 使用的约 90 度的角度 α_x, α_y 的余角记做方向角, 也就是说:

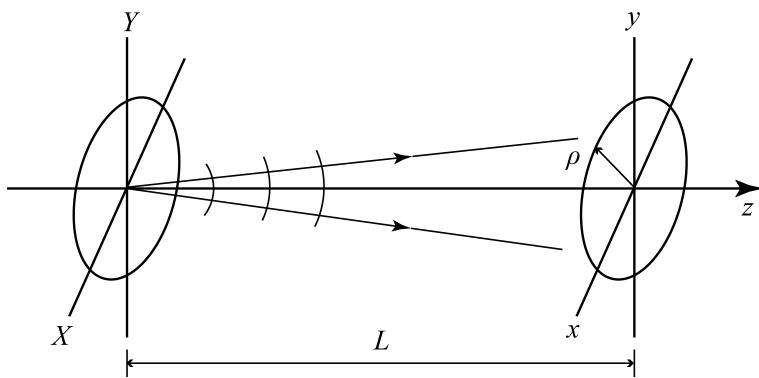


图 26.2: 球面波向平面波近似

$$\sin \theta_x \approx \theta_x = \frac{k_x}{k}, \quad \sin \theta_y \approx \theta_y = \frac{k_y}{k}, \quad k_z \approx \sqrt{1 - \theta_x^2 - \theta_y^2} k \approx k$$

在傍轴近似下, 以上以方向角 (θ_x, θ_y) 传播的平面波被方便地写为:

$$A(\mathbf{r}) = A e^{ikz} e^{i\phi} e^{ik(\theta_x x + \theta_y y)} = A' e^{ik(\theta_x x + \theta_y y)}$$

经常的, 我们只需要关心在像屏 $z = 0$ 上的场, 参考上图. 所以上式我们可以把不变的系数都统一写作 A' .

但是平面波却不总是好的近似, 我们已经发现了, 像上图那样, 如果在物屏 $z = -L$ 上原点 $X = 0, Y = 0$ 处有一点光源发出球面波, 那么光屏上 $(0, 0)$ 处和 (x_1, y_1) 处的近似公式就采取了不同的形式. 所以我们会寻求更好的球面波近似. 但且慢, 是不是任意情况下光场中的任意一点附近的场都能近似为一个球面波呢? 答案显然是否定的. 比如柱面波很显然就不能被近似为球面波, 因为其**波前**(wavefront), 即等相位面沿垂直于传播方向的一个方向弯曲而另一个方向却是平直的. 事实上球面波的光源是点, 柱面波的光源是线. 而介于柱面波与球面波中间的某种不对称波前就连光源都没法找到. 但是球面波的地位仍然是重要的. 这体现在两点上. 一是在真实的光路系统中, 点光源是切实存在且常用的. 此时物方到像方的点到点的消除了像差的理想成像是我们努力追求的方向. 故在这样的系统中光波的确一直都适合被近似为球面波, 而平面波也被当做一种光源在无穷远的特殊状态而处理. 二是, 若把波前上各个点当成是新的次波源向前发射出球面波而在前方相干叠加形成新的光场, 后面在衍射这一章我们数学上可以证明这就是计算光的传播的正确方式. 其中用到的球面波作为一种本质上重要的物理对象, 即**传播子**(propagator), 有着重要的物理地位.

故我们研究从物屏 $z = -L$ 的 (X, Y) 处传播到像屏 $z = 0$ 处 (x, y) 点的球面波的场:

$$\begin{aligned} A(x, y) &= \frac{I}{\sqrt{L^2 + (x - X)^2 + (y - Y)^2}} e^{ik\sqrt{L^2 + (x - X)^2 + (y - Y)^2}} \\ &= \frac{I}{L} e^{ikL} \left[1 + \frac{(x - X)^2}{L^2} + \frac{(y - Y)^2}{L^2} \right]^{-\frac{1}{2}} e^{ikL \left[\left[1 + \frac{(x - X)^2}{L^2} + \frac{(y - Y)^2}{L^2} \right]^{\frac{1}{2}} - 1 \right]} \end{aligned}$$

由于傍轴条件, 我们发现方向角 (u_x, u_y) , $u_x = \frac{x}{L}, u_y = \frac{y}{L}$ 和方向角 (U_x, U_y) , $U_x = \frac{X}{L}, U_y = \frac{Y}{L}$ 都是小量, 而传播到光屏上该点的光的方向角(k 方向)为 $(u_x - U_x, u_y - U_y)$. 仅考虑领头项, 把不随 x, y 变化的系数记为常数, 我们发现:

$$A(x, y) = A e^{\frac{1}{2}kL[(u_x - U_x)^2 + (u_y - U_y)^2]} = A' e^{ik\frac{x^2 + y^2}{2L}} e^{-ik\frac{Xx + Yy}{L}}$$

也就是说光屏上的波由两个相因子决定. 第一个是平方相关于 x, y 的 $\varphi_1 = k \frac{x^2 + y^2}{2L} = \pi \frac{x^2 + y^2}{\lambda L}$, 第二个恰好是平面波的相因子 $\varphi_2 = -k \frac{Xx + Yy}{L} = -k(U_x x + U_y y) = k(\theta_x x + \theta_y y)$. 就是说当点光源相对 $(x, y) = (0, 0)$ 从 $X < 0, Y < 0$ 的方向照过来能够在原点附近获得方向角为正的平面波场.

那么球面波能够被近似为球面波的条件也就不言自明了, 它要求第一个相因子的相位很小不干扰第二个相因子, 也就是 $\varphi_1 \ll \pi$, 也就是远场条件(farfield condition):

$$\rho^2 \ll \lambda L \Leftrightarrow L \gg \frac{\rho^2}{\lambda}$$

其实换一个角度理解这个条件, 如果远场条件被满足, 但是在这样的 ρ 内第二个相因子变化也非常小时, 整个 A 几乎就是常数, 这时候也形成不了有价值的光场, 所以我们如果要求 $\varphi_2 \gg \varphi_1$, 还能发现第三个条件 $X, Y \gg x, y$. 对于这个条件我们做这样的理解, 相位的绝对大小是没有观测效果的. 我们需要的其实是光束干涉时相位的差值. 故这其实是要求物屏的花样尺寸要远大于像屏上干涉花样的观测范围.

在远场条件不被满足的情况下, 考虑第一个相因子对第二个相因子的影响, 会造成干涉图样的扭曲与形变, 几何光学上会造成最基础的像差: 相散和散焦.

在介绍下两节具体的干涉之前, 对干涉的原理做一个简要的介绍是有必要的. 根据之前的介绍, 空间中如果同时存在两个光场:

$$A_1(\mathbf{r}) = |A_1|(\mathbf{r})e^{i\varphi_1(\mathbf{r})} ; A_2(\mathbf{r}) = |A_2|(\mathbf{r})e^{i\varphi_2(\mathbf{r})}$$

我们可以找到其单独存在时的光强, 一般来说, 光强为位置的缓变函数:

$$I_1(\mathbf{r}) = |A_1|^2 = A_1^* A_1 ; I_2(\mathbf{r}) = |A_2|^2 = A_2^* A_2$$

然而光场的叠加是有两种典型的方式的. 最常见的其实是非相干叠加(incoherent superposition). 此时出于下面要介绍的原因, 光强是可以逐点直接相加的:

$$I(\mathbf{r}) = I_1(\mathbf{r}) + I_2(\mathbf{r})$$

将自然界或人造的两束来源不同的光照在一次是无法观察到干涉花样的. 光的干涉条件比机械波干涉条件来的苛刻的多. 在某些细心制备的实验室条件下才能观察到干涉花样, 此时光场是相干叠加(coherent superposition). 在光学之中相干度是十分值得关注的. 我们将在之后的一节中集中讨论. 读者可能自然地认为相干叠加时为复振幅相加:

$$A(\mathbf{r}) = A_1(\mathbf{r}) + A_2(\mathbf{r})$$

但是其实无论非相干叠加还是相干叠加上式都适用, 那么是什么造成了两者的区别呢? 我们把上式模方求光强:

$$\begin{aligned} I(\mathbf{r}) &= A^* A \\ &= (A_1^* + A_2^*)(A_1 + A_2) \\ &= A_1^* A_1 + A_2^* A_2 + A_1^* A_2 + A_2^* A_1 \\ &= I_1 + I_2 + |A_1||A_2|e^{i(\varphi_2 - \varphi_1)} + |A_1||A_2|e^{i(\varphi_1 - \varphi_2)} \\ &= I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\varphi_2 - \varphi_1) \end{aligned}$$

上式中出现的交叉项 $2\sqrt{I_1 I_2} \cos(\varphi_2 - \varphi_1)$ 就叫干涉项(interference term). 它是关于位置的快变函数.乍一看这也未免太快了, 因为只要当空间位置改变一个波长, 相位 φ 就会有 π 的量级的改变, 这样干涉项在这个过程中就会有正负号的改变. 从而产生干涉花样. 在实际的干涉实验中, 合理的仪器设置可以使得这个空间特征长度被放大到人眼或助视仪器可以直接观测的程度. 从而这就是干涉实验的基础.

那么为什么对于非相干叠加情形我们直接去掉了干涉项. 那是因为在之前我们做把光场向标量复振幅简化的过程中去掉了公共的单色光时间演化项 $e^{i\omega t}$. 这并不是永远合理的做法. 真实光场由于后面要介绍的各种原因, 可以认为相位随时间的演化并不是完全线性增加的而是有随时间偏离期望值的方差越来越大的涨落. 从而两束并没有关联的自然光之间不同时刻同一点 \mathbf{r} 的相位差 $\delta = \varphi_2(\mathbf{r}) - \varphi_1(\mathbf{r})$ 是一个在随机涨落的数, 只要涨落够大, $\cos(\varphi_2 - \varphi_1)$ 项的随时间平均就会总是趋近于零. 从而干涉现象也就消失了.

在以下的两小节中, 我们先不用关心相干性的问题, 认为不同光场的叠加总是有“完美的相干性”: 完全稳定的相位差. 我们会在不同干涉装置的介绍中去强调为什么总是能保证“完美的相干性”的成立.

26.2 分波面干涉

以下几种干涉装置一般统一归为分波面干涉(interference by dividing wave-front). 因为它们实现的核心思想都是让本来沿各个方向独立传播的波前的部分改变传播方式而在空间中产生交叠.

- 杨氏双孔干涉: 单色扩展光源-单孔-双孔-光屏.

“完美的相干性”: 诚然, 单色扩展光源可以分解为大量点光源的叠加, 彼此之间是非相干的 (激光等非相干光源除外). 但是, 单孔的存在可以使得我们忽略此前的光场结构而直接把单孔视作单个点光源, 由它发出的光的同一波前上两点具有完全相干性. 实际上, 即使我们对到达单孔的来自扩展光源不同点光源的光进行分解, 那么每一个光波此后通过单孔形成新的球面波, 再通过双孔形成两个具有稳定相位差 (下面会计算) 的相干球面波, 并在光屏上形成干涉条纹, 只要不同点光源形成的条纹位置完全一致, 就等效于从单孔出来的光波整体具有“完美的相干性”.

条纹间距公式:

$$\Delta x = \frac{\lambda L}{d}$$

缺点: 只利用了点光源波面上非常有限的两个小面元. 而以下的都是改善方案, 直接将波面分成两个区域分别进行成像以实现干涉:

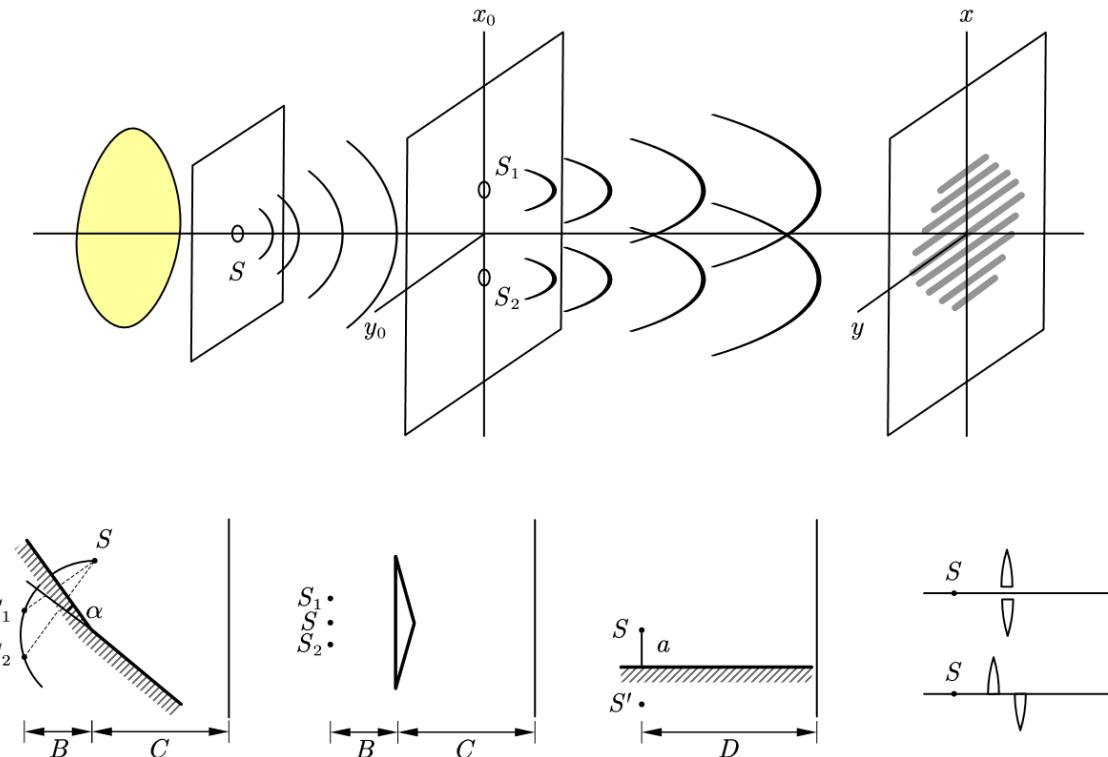


图 26.3: 分波面干涉们

- 菲涅尔双面镜:

$$\Delta x = \frac{\lambda(B + C)}{2\alpha B}$$

- 菲涅尔双棱镜:

$$\Delta x = \frac{\lambda(B + C)}{2(n - 1)\alpha B}$$

棱镜的正反放置有区别.

- 劳埃德镜:

$$\Delta x = \frac{\lambda D}{2a}$$

- 比累对切透镜: 横竖分开有区别.
- 散斑干涉: 如果在带一层玻璃板的镜面上撒粉末, 那么粉末作为散射源, 对于单色光入射情况也可以视作点光源处理. 此时也可以发生分波面干涉: 一部分吸收入射光后形成相应的散射光, 经反射后继续传播. 而一部分入射光先反射再被散射. 相当于在入射光波面上通过粉末和粉末的反射像发生了分波面干涉.

作为干涉, 一般指两个波的叠加. 但是实际情况常常遇到多光束干涉(multiple-beam interference) 的情况. 比如把杨氏实验换成多个孔, 或者把孔换成全同的散射单元. 多光束干涉的进一步推广就是衍射(diffraction), 此时一个空间区域内的整个波前应当被分解为无穷多个面微元, 各产生一个波, 在光屏上叠加, 后面会发现表达式成了一个积分.

26.3 分振幅干涉

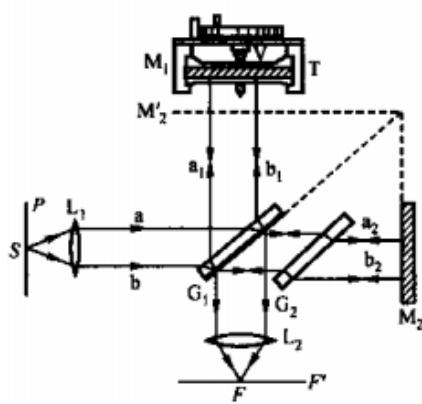


图 26.4: 迈克耳孙干涉仪

以下几种干涉装置一般统一归为分振幅干涉(interference by dividing amplitude). 因为它们实现的核心思想, 相比去把整个波前分成几个部分, 它选择保持了波前的完整性, 而是通过界面效应把其强度分解为透射与反射两个成分.

- 迈克耳孙干涉仪:

G_1 为半透半反镜, G_2 为完全透射的补偿板. M_2, M'_2 对 G_1 为物像关系. 那么在光屏中央为亮斑条件:

$$2d = j\lambda$$

事实上, 等价于薄膜干涉, 可以是等倾或者等厚.

- 薄膜等厚干涉: 光线倾角固定(往往受制于光路), 薄膜不同厚度处形成不同光程差导致干涉条纹.

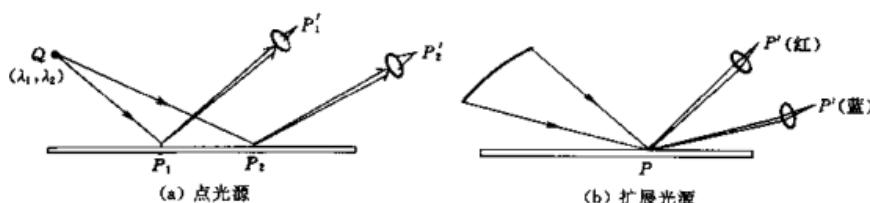


图 26.5: 薄膜等厚干涉

- 薄膜等倾干涉: 薄膜厚度固定, 不同光线倾角处造成不同光程差导致干涉条纹.

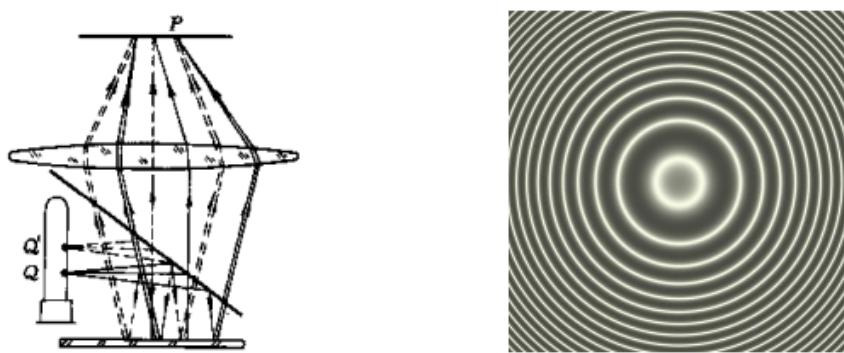


图 26.6: 薄膜等倾干涉

根据有无半波损失, 薄膜等厚等倾干涉的光程差均为:

$$\delta = 2nd \cos \theta \left(+\frac{\lambda}{2} \right)$$

- 法布里-泊罗干涉仪(Fabry-Pérot interferometer).

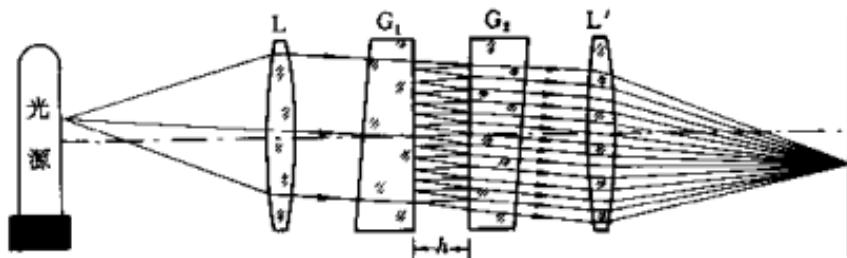


图 26.7: FP 干涉仪

真正的薄膜干涉往往是一个多光束干涉. 以 FP 干涉仪为例, 透射振幅为:

$$A_1 = t^2 A_0 = T A_0 = (1 - R) A_0$$

$$A_{i+1} = r^2 A_i e^{ik\delta} = R e^{ik\delta} \cdot A_i$$

做等比数列求和, 代入薄膜光程差公式, 模方取光强得到:

$$I = \frac{I_0}{1 + \frac{4R}{(1-R)^2} \sin^2(kd \cos \theta)}$$

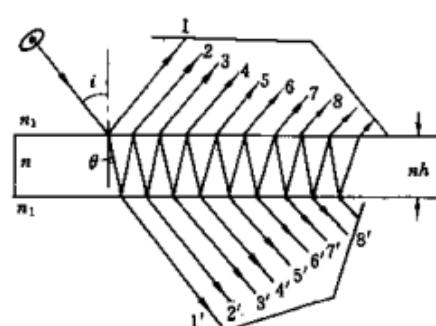


图 26.8: 多光束干涉

26.4 偏振干涉

- 相干度:

$$\gamma = \frac{\langle A_1^* A_2 \rangle}{\sqrt{\langle A_1^* A_1 \rangle \langle A_2^* A_2 \rangle}}$$

一般是个复数, 如果模为零代表完全不相干, 如果模为一代表完全相干, 但是幅角代表两个信号的相位差. 模在零一之间代表部分相干.

- 密度矩阵包含有一个偏振态的全部信息:

$$\mathbf{P} = \begin{bmatrix} \langle A_x^* A_x \rangle & \langle A_x^* A_y \rangle \\ \langle A_y^* A_x \rangle & \langle A_y^* A_y \rangle \end{bmatrix}$$

对角线元素都是实数:两个方向光强. 迹: 总光强.

- 线偏振光:

$$\mathbf{P} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \quad \text{or} \quad \mathbf{P} = I \begin{bmatrix} \cos^2 \theta & \cos \theta \sin \theta \\ \cos \theta \sin \theta & \sin^2 \theta \end{bmatrix}$$

- 圆偏振光, 左旋右旋分别为:

$$\mathbf{P} = \frac{I}{2} \begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix} \quad \text{or} \quad \mathbf{P} = \frac{I}{2} \begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}$$

- 椭圆偏振光, 左旋右旋分别为:

$$\mathbf{P} = \begin{bmatrix} I_x & -i\sqrt{I_x I_y} \\ i\sqrt{I_x I_y} & I_y \end{bmatrix} \quad \text{or} \quad \mathbf{P} = \begin{bmatrix} I_x & i\sqrt{I_x I_y} \\ -i\sqrt{I_x I_y} & I_y \end{bmatrix}$$

- 以上三种光都具有完全相干性. 完全相干的光的任意两个分量, 或者是经过之后的偏振器件处理前后的各分量之间, 都具有模一的相干度. 但是, 两个不同的完全相干光之间又可以是不相干的. 偏振态为完全相干的充要条件为:

$$\det \mathbf{P} = 0$$

- 自然光: 相干度最差的光:

$$\mathbf{P} = \frac{I}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- 一般的光称作部分偏振光. 一般来说密度矩阵写作:

$$\mathbf{P} = \begin{bmatrix} I_x & I_{xy} e^{-i\varphi} \\ I_{xy} e^{i\varphi} & I_y \end{bmatrix}$$

可以证明可以找到“主方向”, 一方面, 这两个方向光强分别达到最大和最小 (取 x 方向最大, y 方向最小), 另一方面取这样的方向后恰好有 $\varphi = \pi/2$. 故部分偏振光在主方向坐标系中永远可以写作:

$$\mathbf{P} = \begin{bmatrix} I_{max} & -iJ \\ iJ & I_{min} \end{bmatrix}$$

即: 视作椭圆偏振光和一个与之不相干的自然光的叠加.

- 偏振片: 筛选某个特定方向的分振动:

$$A_\theta = A_x \cos \theta + A_y \sin \theta$$

- 波片: 命 x, y 方向的透射振动产生一定的相位差. 二分之一波片产生 $\varphi = \pi$ 相位差, 四分之一波片产生 $\varphi = \pi/2$ 相位差, 合理选取时间原点可以认为 y 方向相位不变, 那么:

$$A'_x = e^{i\varphi} A_x \quad , \quad A'_y = A_y$$

26.5 相干性

- 空间相干性: 在光源产生的光场同一时间 (同一波前) 上选取两个不同的空间点, 其信号相干度模不足一, 且随着点距离变大而愈发不相干的情况:

$$A_1 = A(\mathbf{r}_1, t) \quad , \quad A_2 = A(\mathbf{r}_2, t)$$

$$\gamma = \frac{\langle A_1^* A_2 \rangle}{\sqrt{\langle A_1^* A_1 \rangle \langle A_2^* A_2 \rangle}} \quad , \quad |\gamma| < 1$$

一般原因: 光源的扩展性. 对于宽度为 b 的单缝, 相干角度为:

$$\Delta\theta = \frac{\lambda}{b}$$

- 时间相干性: 在光源产生的光场同一位置 (同一波列) 上选取两个不同的时间点, 其信号相干度模不足一, 且随着时间间隔变大而愈发不相干的情况:

$$A_1 = A(\mathbf{r}, t + \Delta t) \quad , \quad A_2 = A(\mathbf{r}, t)$$

$$\gamma = \frac{\langle A_1^* A_2 \rangle}{\sqrt{\langle A_1^* A_1 \rangle \langle A_2^* A_2 \rangle}} \quad , \quad |\gamma| < 1$$

一般原因: 发光的不连续性, 存在波列长度 L . 内在原因为光源的非单色性, 即波长存在不确定度 $\Delta\lambda$. 两者符合:

$$L = \frac{\lambda^2}{\Delta\lambda}$$

第二十七章 光的衍射

27.1 光栅与波带片

- 光栅衍射: 光栅常数, 即空间周期为 d , 缝宽为 a , 总共刻 N 道. 那么:

$$I = I_0 \cdot f_1(\theta) \cdot \left(\frac{\sin N\beta}{\sin \beta} \right)^2, \quad \beta = \frac{\pi d}{\lambda} \sin \theta$$

其中 $f_1(\theta)$ 为之后可以算出来的与单缝衍射有关的慢变 (需要 $a \ll d$) 的单元因子:

$$f_1(\theta) = \left(\frac{\sin \alpha}{\alpha} \right)^2, \quad \alpha = \frac{\pi a}{\lambda} \sin \theta$$

体现光栅的结构的是后面的结构因子:

$$f_2(\theta) = \left(\frac{\sin N\beta}{\sin \beta} \right)^2, \quad \beta = \frac{\pi d}{\lambda} \sin \theta$$

从中可以得到三个信息: 主极大方向 $d \sin \theta_j = j\lambda$, 主极大强度 $I = N^2 I_0$, 主极大宽度 $\delta\theta \sim j\lambda/Nd$

- 圆孔菲涅尔衍射: 波带法与半波带法. 核心在于以下等式:

$$\frac{dS}{r} = \text{const.} \cdot dr$$



图 27.1: 波带法与半波带法

这说明不同位置的波前上的带状面元, 根据其对要计算的点所造成的光程差, 相等的光程差造成完全相等的振幅, 尽管面元的面积不相等. 只需要把这些振幅相等, 相位不等的光矢量进行合成.

合成方法一般有两种, 对于半波带问题把面元分解为半波带显得方便. 而复杂问题需要按照微元分解, 下一节将介绍面元对要计算的点的光矢量贡献还有一个起到调制作用的角度因子, 故合成图像一般为:

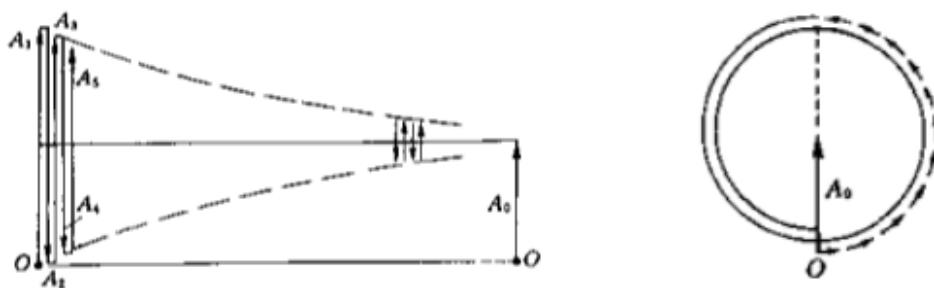


图 27.2: 分组合成与积分合成

- 波带片: 预想将一点光源和一要计算光强的点中间合适距离处的波前按照半波带分解, 再制造一光学器件, 或者遮住奇数半波带而开放偶数半波带, 或者遮住偶数半波带而开放奇数半波带. 就构成了波带片. 如果在一张圆形的片上入射平行光, 而选取距离为 f 的点计算光程, 那么各个半径为:

$$\rho_j = \sqrt{j\lambda f}$$

这样的波带片可以当做一个焦距为 f 的透镜使用. 但是成像并不唯一, 容易证明它具有一组虚焦点和实焦点.

27.2 衍射积分公式

- 基尔霍夫衍射积分公式:

$$A(\mathbf{r}) = \frac{1}{i\lambda} \int A(\mathbf{r}') \cdot F \cdot \frac{e^{ik|\mathbf{r}-\mathbf{r}'|} dS}{|\mathbf{r}-\mathbf{r}'|}$$

F 为角度因子, 若入射波波矢 \mathbf{k}' 方向为 e' , 子波波矢 \mathbf{k} 方向为 e , 面元 dS 方向为 n , 则有:

$$F = \frac{\mathbf{n} \cdot \mathbf{e} + \mathbf{n} \cdot \mathbf{e}'}{2} = \frac{\cos \theta + \cos \theta'}{2}$$

- 夫琅和费衍射公式: 如果在衍射屏后放置焦距为 f 的透镜并在后焦面上观察, 并且忽略角度因子. 把入射场在衍射屏作用后的波前写作 $A(x_0, y_0)$. 那么这个情况具有更简单的公式:

$$A(\theta_x, \theta_y) = \frac{1}{i\lambda f} \iint A(x_0, y_0) e^{-i(k_x x_0 + k_y y_0)} dx_0 dy_0$$

- 矩孔衍射:

$$I = I_0 \cdot \left(\frac{ab}{\lambda f} \right)^2 \cdot \left(\frac{\sin \alpha}{\alpha} \right)^2 \cdot \left(\frac{\sin \beta}{\beta} \right)^2, \quad \alpha = \frac{\pi a}{\lambda} \sin \theta_x, \beta = \frac{\pi b}{\lambda} \sin \theta_y$$

- 圆孔衍射: 直径 d , 艾里斑半角 θ :

$$\theta \approx \frac{1.22\lambda}{d}$$

第二十八章 物理光学

28.1 经典色散理论

蓝天,白云,红太阳.在简单的自然现象中蕴含着光在传播过程中的另外一种典型的现象: **散射**(scattering).而散射却又是一个多么复杂的现象!要完全理解散射,我们得研究光的吸收与发射,因为光与物质中分子的相互作用其实本质上可以看成吸收与发射,这是**分子光学**(molecular optics)的研究范畴.其次,光在介质中的衰减(吸收),不同颜色的光折射率的不同(色散)其实都导致了一类散射现象(瑞利散射).但是还有不同的散射现象则表现出不用的特性,需要从更宏观或更微观的角度去建立新的理论来解释他们.

不同于以往在透明介质中的平面波传播,新的现象需要新的平面波模型.

28.1.1 复波矢与复折射率

我们在波动方程的解中引入复数是为了更好地描述振动形式的解.但这却忽视了另一种解的存在性.就是如果波矢 \mathbf{k} 本身亦为复数,由于同时又是一个矢量,可以表示为 $\mathbf{k} = \alpha + i\beta$.代入场的空间部分:

$$A(\mathbf{r}) = A_0 e^{i\mathbf{k} \cdot \mathbf{r}} = A_0 e^{-\beta \cdot \mathbf{r}} e^{i\alpha \cdot \mathbf{r}}$$

这就代表了一个沿 α 方向传播,沿 β 方向衰减的平面波.而 α 为波矢的实部, β 为波矢的虚部:

$$\alpha = \Re(k) ; \beta = \Im(k)$$

这种波可不可以真空中传播?出人意料的是这居然是可能的.因为如果代入真空中的亥姆霍兹方程:

$$\nabla^2 A + \frac{\omega^2}{c^2} A = 0$$

便会觉得这只需要要求:

$$\mathbf{k}^2 = (\alpha + i\beta) \cdot (\alpha + i\beta) = \frac{\omega^2}{c^2} \Rightarrow \alpha^2 - \beta^2 = \frac{\omega^2}{c^2}, \alpha \cdot \beta = 0$$

这样的波我们并不陌生,早在几何光学我们介绍光的全反射时便研究过另一侧界面中的隐失波场便是以上形式.更典型地,在 $a \times b$ 的矩形波导中传播的波满足:

$$(\frac{2\pi}{\lambda_x})^2 + (\frac{2\pi}{\lambda_y})^2 + (\frac{2\pi}{\lambda_z})^2 = \frac{\omega^2}{c^2}$$
$$\lambda_x = \frac{2b}{m}, \lambda_y = \frac{2a}{n}, m, n = 1, 2 \dots$$

此时在 x, y 方向的振动是驻波, z 方向传播的波则是行波. x, y 方向要形成驻波是因为在边界的导体上要符合相同的边界条件,所以或都是波腹,或都是波节.亦可以用同一个

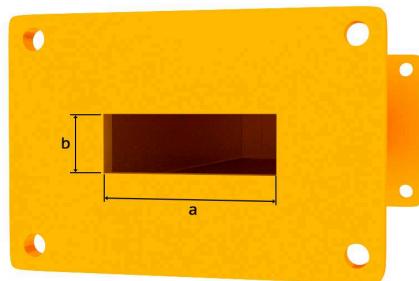


图 28.1: 矩形波导

平面波在侧面上的多次反射的叠加来理解这种兼具驻波和行波特质的场. 我们发现模式为 (m, n) 的波, 会有一个截止角频率 ω_c , 如果角频率低于此值使得波无法向前传播:

$$\omega_c = c \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

那如果在波导的起点的天线给出信号角频率 $\omega < \omega_c$ 会发生什么情况? 此时就是 z 方向波矢 k_z 变为虚数, 意味着波沿 z 方向不是传播而是衰减:

$$\mathbf{k} = \boldsymbol{\alpha} + i\boldsymbol{\beta}$$

$$\boldsymbol{\alpha} = \left(\frac{m\pi}{a}, \frac{n\pi}{b}, 0 \right) , \quad \boldsymbol{\beta} = \left(0, 0, \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 - \frac{\omega^2}{c^2}} \right)$$

在真空中尚可以产生衰减的波¹, 更何况在各种介质中. 事实上, 描述光在介质中随着传播的距离而衰减的现象早有一经验公式, 即**比尔-朗博-布葛律**(Beer-Lambert-Bouguer law):

$$\frac{I(l)}{I(0)} = e^{-(\sum_i n_i \sigma_i)l}$$

其中 $I(0)$ 为在传播 l 前的光强, $I(l)$ 为在传播 l 后的光强. n_i 是构成介质的第 i 种分子的数密度, σ_i 为描述其吸收本领的**衰减截面**(attenuation cross section). 例如对大气来说主要的衰减因素就来自气溶胶, 干净的海水则能提供 200m 左右的透光带, 如果有的话, 浮游植物含有的光合色素是造成衰减的主要原因. 而对于强衰减性质的介质, 我们可以唯象地认为波在介质中的传播为:

$$|A(l)| = |A(0)|e^{-\beta l} \Rightarrow I(l) = I(0)e^{-2\beta l}$$

这样的选择是有道理的. 下以导电导致的损耗为例: 在不漏电介质中, 麦克斯韦关系本应为:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{H}}{\partial t} = \mathbf{0} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} - \varepsilon \mu \frac{\partial \mathbf{E}}{\partial t} = \mathbf{0} \end{array} \right.$$

这样便有:

$$v = \frac{c}{n} = \frac{1}{\sqrt{\varepsilon \mu}} , \quad k = \frac{\omega}{v} = \omega \sqrt{\varepsilon \mu}$$

但是, 如果介质漏电, 且符合 $\mathbf{J} = \sigma \mathbf{E}$, 那么以上方程就改写为:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{H}}{\partial t} = \mathbf{0} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} - \varepsilon \mu \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \sigma \mathbf{E} \end{array} \right.$$

¹事实上, 只需要亚波长的结构便可造成衰减.

对最后一个式子所引发的不同在光学情况下是很好处理的. 因为单色光总是具有固定的频率 ω , 那么其实以上式子无非是把 $\frac{\partial}{\partial t}$ 变成 $-i\omega$. 在磁导率被认为几乎等于真空磁导率的情形下, 这相当于说:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} - i\omega \mathbf{H} = \mathbf{0} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} + i\omega \varepsilon \mu \mathbf{E} = \mathbf{0} \end{array} \right. \xrightarrow{\varepsilon \rightarrow \varepsilon + i\frac{\sigma}{\omega}} \left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} - i\omega \mathbf{H} = \mathbf{0} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} + i\omega \varepsilon \mu \mathbf{E} - i\sigma \mu \mathbf{E} = \mathbf{0} \end{array} \right.$$

所以对于漏电介质², 通常会有复电容率(complex permittivity) 的说法³, 它就是以上把电容率和电导率合并以后的新的复常数, 用它第四个方程就与真空中的方程没有任何区别了, 除了系数是一个复数:

$$\varepsilon' = \varepsilon + i\frac{\sigma}{\omega} : \quad \nabla \times \mathbf{H} + i\omega \varepsilon' \mu \mathbf{E} = \mathbf{0}$$

这样一个方程的解是可以完全照搬之前的解的, 因为数学上可以证明复数解具有可解析延拓(analytic continuation) 的特性. 从而容易发现, 复波矢就变为:

$$\mathbf{k}^2 = \omega^2 \varepsilon' \mu$$

这样就得到:

$$\alpha^2 - \beta^2 = \omega^2 \varepsilon \mu, \quad 2\alpha \cdot \beta = 2\alpha\beta \cos \theta = \omega \sigma \mu > 0$$

此时传播波矢 α 与衰减波矢 β 就不一定要垂直了, 它们必须夹锐角, 也就是说, 如果电磁波在漏电介质或者导体中传播, 沿传播方向必须要衰减. 我们最后引入光学中使用最多的复折射率(complex refraction index), 按照原来的看法它意味着 $k^2 = n^2 k_0^2 = n^2 \frac{\omega^2}{c^2} = \frac{\omega^2}{v^2}$. 现在要更小心些, 因为 $\mathbf{k} = \alpha + i\beta$ 已经包含两个不贡献的波矢部分. 故我们先对以下表达式开方:

$$\alpha^2 - \beta^2 + 2i\alpha\beta \cos \theta = \omega^2 \varepsilon' \mu$$

$$\Rightarrow \sqrt{\alpha^2 - \beta^2 + 2i\alpha\beta \cos \theta} = n \frac{\omega}{c} = (n_1 + i n_2) \frac{\omega}{c}$$

通过以上两式对比, 我们能得到两个方面. 第一是复折射率实部 n_1 与虚部 n_2 分别是这样依赖于相对介电常数 $\varepsilon_r = \varepsilon/\varepsilon_0$ 和电导率 σ 的:

$$n_1 = \sqrt{\varepsilon_r} \cdot \sqrt{\frac{1}{2} \left(1 + \frac{\sigma^2}{\omega^2 \varepsilon^2} + \sqrt{1 + \frac{\sigma^2}{\omega^2 \varepsilon^2}} \right)}$$

$$n_2 = \frac{\sigma}{\sqrt{\varepsilon_r} \omega \varepsilon_0} \cdot \sqrt{\frac{1}{2} \left(1 + \frac{\sigma^2}{\omega^2 \varepsilon^2} + \sqrt{1 + \frac{\sigma^2}{\omega^2 \varepsilon^2}} \right)^{-1}}$$

第二组关系式如果已知了介质的两个折射率, 那么在介质中传播的波的两个波矢 α 和 β 需要满足的关系为:

$$\alpha^2 - \beta^2 = (n_1^2 - n_2^2) \frac{\omega^2}{c^2}$$

$$\alpha\beta \cos \theta = n_1 n_2 \frac{\omega^2}{c^2}$$

只有当 $\theta = 0$ 才恰有:

$$\alpha = n_1 \frac{\omega}{c}, \quad \beta = n_2 \frac{\omega}{c}$$

²即使不漏电, 也会由于有损耗而等效于有复电容率.

³对于普通的导体, 也一样可以讨论复电容率, 或者更常见地, 复电导率(complex conductivity). 它不仅包含由于原子实部分极化导致的电容性, 还要包含由于电子运动惯性导致的电感性.

28.1.2 经典电子论的解释

下面我们来介绍历史上发挥了重要作用的经典电子论。它虽然不够精确但物理图像十分重要。

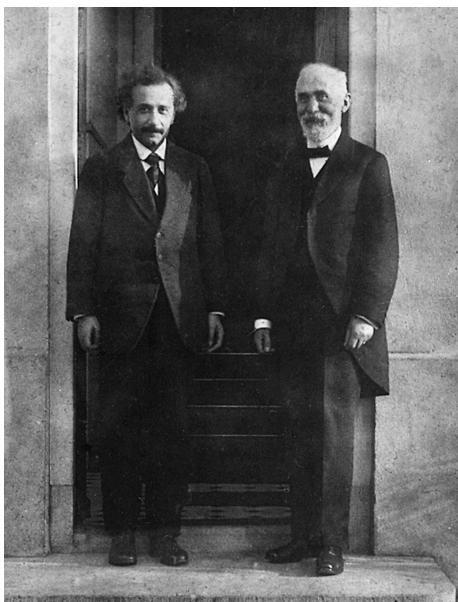


图 28.2: Einstein 与 Lorentz

历史上虽然电子作为粒子的发现是 1897 年汤姆孙 (J. J. Thomson) 的工作。但是早在半个世纪前人们就开始逐渐相信物质由两种带电粒子构成并在导电时可以移动的模型了。人们用这样的模型去想象输运现象, 电磁感应, 热电耦合等现象并取得了丰硕的成果, 经典电磁学早在 1860 年代便在麦克斯韦 (J. C. Maxwell) 的工作中达到顶点进入尾声。剩下来的几十年时间历史让位给了洛伦兹 (H. A. Lorentz), 洛伦兹对电磁学的贡献是深远的。他承前人之大成, 讨论电磁场与电磁介质的相互作用, 讨论电磁学原理与古老的伽利略式相对性原理的结合; 又开近代物理之先河, 电子论自然地导向了量子理论, 而电磁学又自然地促成了狭义相对论的诞生。事实上, 早年间狭义相对论在学界被介绍时便是以洛伦兹和爱因斯坦这一对忘年交命名的洛伦兹-爱因斯坦理论。

电子, 这样一个基本粒子可以说是妇孺皆知的概念, 它带一个单位的负电荷, 其基本参数为:

$$e = 1.602176634 \times 10^{-19} \text{C} \quad m_e = 9.10938356 \times 10^{-34} \text{kg}$$

然而出于各种原因, 在经典电子论中的电子却又是陌生的概念。根据洛伦兹的经典电子论精神, 重的原子实被认为静止, 轻的束缚电子被认为在外场中可以运动, 且主要受电场的影响⁴。电子在外电场中相对原子实的位移造成了电偶极矩。出于好的近似, 将电子视作被绑在连接中心的弹簧的另一端的质点。质量为 m 不再被认为与基本粒子的质量一致, 事实上这个质量的本质实际上是虚无的, 因为当考虑到原子与电磁场相互作用的严格理论时图像将完全是量子的而超出了此处可以讨论的范围, 这样的一个为了解释现象去构造的模型就叫唯象模型 (phenomenological model)。但是其电荷量还是 $-e$, 弹簧劲度系数为 k 。这样子我们就可以写出电子在外电场中的动力学方程:

$$m\ddot{\mathbf{r}} + k\mathbf{r} = -e\mathbf{E}$$

但是我们还忽略了一点, 加速运动的电荷会产生辐射, 也会因为辐射而带走能量和动量, 这就被视作介质对电磁波吸收, 散射的根本原因。也同样为了唯象地描述它, 我们认为电子是受到了一个辐射阻尼力 $\mathbf{f} = -\gamma\dot{\mathbf{r}}$ 。这样以上方程就被修改为:

$$m\ddot{\mathbf{r}} + \gamma\dot{\mathbf{r}} + k\mathbf{r} = -e\mathbf{E}$$

命 $\omega_0 = \sqrt{k/m}$ 代表共振频率 (resonance frequency) 或待会就会说明的吸收频率 (absorption frequency)。再命入射电磁波为频率 ω 的光, 按光学符号约定 $\mathbf{E} = \mathbf{E}_0 e^{-i\omega t}$, 那么很容易解出受迫振动解 $\mathbf{r} = \mathbf{r}_0 e^{-i\omega t}$, 振幅为:

$$\mathbf{r}_0 = \frac{-e\mathbf{E}_0/m}{\omega_0^2 - \omega^2 - i\gamma\omega/m}$$

我们现在就能算出动态的分子极化率来, 它被定义为分子中 Z 个价电子产生的偶极矩 $\mathbf{p} = -Zer$ 与外电场 \mathbf{E} 的比值, 可以发现它也是个复数, 依赖于外光场的频率:

$$\alpha(\omega) = \frac{Ze^2/m}{\omega_0^2 - \omega^2 - i\gamma\omega/m}$$

我们知道, 对于具体介质的极化总是与各个单元的极化有关。我们在此讨论稀薄的无固有偶极矩的气体

⁴根据洛伦兹力公式 $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, 而电磁场中 $E = cB$, 故 $v \ll c$ 时磁场力可以忽略。

在外电磁波中的极化⁵, 此时不难想到介质的介电常数应该直接依赖于分子极化率, 事实上, 极化强度为:

$$\mathbf{P} = \chi \varepsilon_0 \mathbf{E} = (\varepsilon_r - 1) \varepsilon_0 \mathbf{E} = n \mathbf{p}$$

从而得到复电容率的值:

$$\varepsilon' = \varepsilon_0 + \frac{n Z e^2 / m}{\omega_0^2 - \omega^2 - i \gamma \omega / m}$$

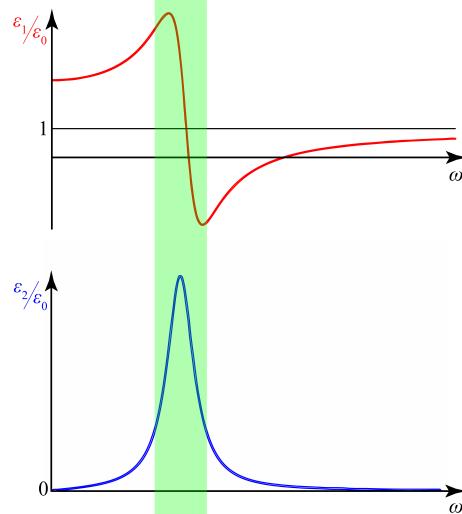
通常, 把 $n Z e^2 / \varepsilon_0 m$ 称作等离子体频率(plasma frequency)的平方 ω_p^2 . 那么上式也作:

$$\varepsilon' = \varepsilon_0 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i \gamma \omega / m} \varepsilon_0$$

出现虚部其实就意味着损耗, 在振动的过程中电场在持续不断地对振子输入能量而导致了电磁波的衰减. 只不过这里的损耗并不是导电行为导致的. 写出复电容率的实部和虚部 $\varepsilon' = \varepsilon_1 + i \varepsilon_2$:

$$\varepsilon_1 = \varepsilon_0 \cdot \left[1 + \frac{\omega_p^2(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + (\gamma \omega / m)^2} \right]$$

$$\varepsilon_2 = \varepsilon_0 \cdot \frac{\gamma \omega_p^2 \omega / m}{(\omega_0^2 - \omega^2)^2 + (\gamma \omega / m)^2}$$



两个函数看上去复杂, 其实都属于近似的洛伦兹型函数(lorentzian function). 图像如右图所示. 在绿色带状区域这个函数有着较大的转折. 这发生在 $\omega \approx \omega_0$ 处. 若做近似, 认为反应损耗的 γ 较小. 则令 $\omega^2 - \omega_0^2 = x, \gamma \omega_0 / m = a$, 那么上式在带状区域附近近似为标准的洛伦兹型:

$$\varepsilon_1 / \varepsilon_0 = 1 - \omega_p^2 \cdot \frac{x}{x^2 + a^2}$$

$$\varepsilon_2 / \varepsilon_0 = \omega_p^2 \cdot \frac{a}{x^2 + a^2}$$

电容率的实部在带状区域由区域外的缓慢增加转为剧烈减少, 而恰好在这样一个区域, 电容率的虚部突然变得很大. 我们习惯上用虚部这种洛伦兹函数增加到最大值的一半的两个点之间的间距作为特征的半高峰宽(full width at half maximum, FWHM). 对于洛伦兹型, 其值恰为 $x = \pm a$ 之间的间距 $2a$, 但是注意到要转化为 ω 对应的间距, 它恰好为:

$$\Gamma = \frac{\gamma}{m}$$

我们可以从这个模型中清楚地看到两个物理现象: 色散(dispersion)与吸收(absorption). 前者就是说折射率依赖于波长, 这和复电容率的实部依赖于频率只是说法不同本质一样. 而复电容率的虚部, 根据之前的讨论, 也就代表吸收. 我们发现在入射光的频率近似为谐振子的固有频率 ω_0 时, 对应经典力学中发生共振的区域附近, 色散和吸收都变得很显著. 我们知道, 经典电子论给出的解释仅仅是一个唯象模型, 所以相关的参数都要根据实验结果来确定. ω_0 可以根据发生强烈吸收与色散的波长来确定, 而 γ/m 则根据吸收峰的半高峰宽 Γ 来确定, 最后等离子体频率 ω_p 则比较有趣, 我们恰好可以根据零频率处的折射率 n_0 来确定它:

$$n_0^2 = \varepsilon_1(0) / \varepsilon_0 = 1 + \omega_p^2 / \Gamma^2$$

那么实际情况是否这么简单呢? 至少通过对右图的常温下水的色散的测量中我们发现, 除了在可见光波段水几乎是透明的而具有大约 1.33 的折射率, 在近红外(0.8-2.5 μm)与中红外(2.5-15 μm)波段的短波区定性上符合以上模型推导出来的结果. 但是中红外到远红外(15-1000 μm)则明显偏离以上结果.

⁵如果是液体或固体, 一方面由于不同的偶极子间的相互影响不可忽略导致公式需要修正, 也因为极化方式还有取向极化等所以需要修正, 但全都只影响定量结果, 定性的图像仍然是适用的.

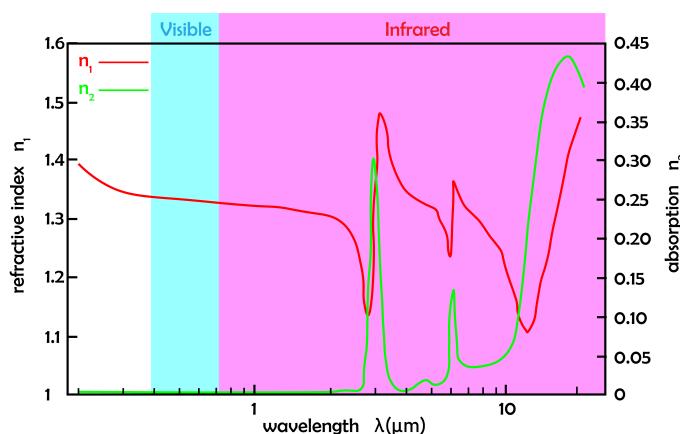


图 28.4: 水的可见-红外色散曲线

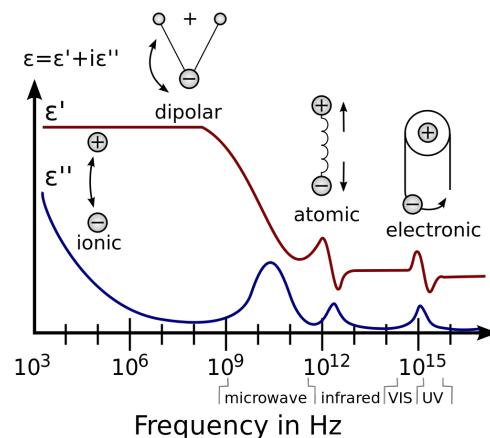


图 28.5: 光与原子作用类型

事实上,在不同的波段,适用于不同的光与物质相互作用的基本模型.微波波段光波长足够长使得晶体的正负离子做大范围的整体相对运动⁶,就是晶格振动.红外则开始使得分子振动,包括转动和振动等形式.再往下的可见波段才是电子共振,也包括原子间的能带共振和原子内的能级共振.最后在紫外波段,电子甚至能直接被电离,这就造成了光与物质相互作用问题的复杂性.

但无论哪种相互作用的机制,我们上述推导得到的描述有着特定共振频率的色散与吸收的最终公式是十分普适的.它只有三个待定的参数:共振频率 ω_0 , 半高峰宽 Γ 和反应共振强度的等离子体频率 ω_p . 我们接下来要做的,是把不同振子,不同吸收效应带来的结果进行求和:

$$\begin{aligned}\varepsilon_1/\varepsilon_0 &= 1 + \omega_p^2 \cdot \sum_i \frac{c_i(\omega_i^2 - \omega^2)}{(\omega_i^2 - \omega^2)^2 + \Gamma_i^2 \omega^2} \\ \varepsilon_2/\varepsilon_0 &= \omega_p^2 \cdot \sum_i \frac{c_i \Gamma_i \omega}{(\omega_i^2 - \omega^2)^2 + \Gamma_i^2 \omega^2}\end{aligned}$$

其中 ω_p 仍然统计各个吸收峰的总强度,而 c_i 系数的和必须为 1 则统计各个吸收峰的相对强度.我们最后计算远离吸收带处的折射率值.忽略各个 Γ_i 后得到:

$$n^2 = 1 + \omega_p^2 \cdot \sum_i \frac{c_i}{\omega_i^2 - \omega^2}$$

上式就是著名的塞尔迈耶尔方程(Selmeier's equation),如果用波长表示则为:

$$n^2 = 1 + \sum_i a_i \cdot \frac{\lambda^2}{\lambda^2 - \lambda_i^2}, \quad a_i = \frac{c_i \omega_p^2 \lambda_i^2}{4\pi^2 c^2}$$

上式依然可以做近似,把共振的各个波长 λ_i 从小到大按 i 来排列.不妨设 $\lambda_i \ll \lambda_{i+1}$,而 λ 恰好接近 λ_i ,那么忽略所有 λ_{i+1} 及之后对应的小量项并将 λ_{i-1} 及之前的项直接近似为 a_i :

$$\begin{aligned}n^2 &\approx 1 + a_1 + a_2 + \cdots + a_{i-1} + \frac{a_i \lambda^2}{\lambda^2 - \lambda_i^2} \\ &= 1 + a_1 + a_2 + \cdots + a_{i-1} + a_i + \frac{a_i \lambda_i^2}{\lambda^2 - \lambda_i^2} \\ &\approx 1 + a_1 + a_2 + \cdots + a_{i-1} + a_i + \frac{a_i \lambda_i^2}{\lambda^2} \\ &= A_0 + \frac{B_0}{\lambda^2}\end{aligned}$$

⁶十分类似于热运动的那种形式,区别在于这是外场诱导的有规律的振动,而且正负电荷位移一定反相,对应热振动的特定光学支.

而进一步近似得到:

$$\sqrt{A_0 + \frac{B_0}{\lambda^2}} = \sqrt{A_0} \left(1 + \frac{B_0}{A_0 \lambda^2}\right)^{1/2} = \sqrt{A_0} \left(1 + \frac{B_0}{2A_0 \lambda^2} - \frac{B_0^2}{8A_0^2 \lambda^4} + \dots\right)$$

一般就总结为如下柯西色散公式(Cauchy dispersion formula):

$$n(\lambda) = A + \frac{B}{\lambda^2}$$

有时为了精度原因再加上四阶项 $+C/\lambda^4$. 而从推导过程中我们发现系数之间的联系:

$$A = \sqrt{1 + \sum_1^i a_i}, \quad B = \frac{a_i \lambda_i^2}{2A}$$

28.1.3 * 辐射阻尼与散射截面

我们此前得到了电荷做加速运动时需要满足的辐射能流和拉莫尔公式:

$$S = \frac{q^2 \ddot{\mathbf{r}}^2}{8\pi^2 \varepsilon_0 c^3 R^2} \sin^2 \theta, \quad P = \frac{q^2 \ddot{\mathbf{r}}^2}{6\pi \varepsilon_0 c^3}$$

从这个式子中可以看出, 在有加速度 $\ddot{\mathbf{r}}$ 的情况下, 辐射场的动量分布是关于电荷中心对称的. 故虽然会造成电荷的动能损失, 但是似乎不造成电荷的动量改变. 但是这一点其实不够准确. 因为偶极辐射本身过程中就包含有取最低阶项的近似成分. 本来可以精确计算的非零发射动量在这个过程中完全有可能被忽略. 而洛伦兹和亚伯拉罕在 1892 年通过将点电荷模型扩大为一个半径很小的球体证明了球体自己产生的辐射场会给自己一个动量的改变. 其中一部分解释为自带场产生的质量对应的跟随电荷运动的动量改变. 另一部分则对应一个辐射阻尼力(radiation damping force), 后人称之为洛伦兹-亚伯拉罕力(Lorentz Abraham force). 后来量子电动力学进一步将其表达式精确化, 它为:

$$\mathbf{f} = +\frac{q^2}{6\pi \varepsilon_0 c^3} \dot{\mathbf{r}}$$

这样一个力却不大可能代表电荷与场这样一个无限多自由度⁷系统的全部. 注意到以下量具有时间量纲, 我们可以把电荷受到一个外力 \mathbf{F} 并与辐射场相互作用下电荷符合的动力学方程在仅考虑辐射阻尼力的情况下简化为:

$$\tau = \frac{q^2}{6\pi \varepsilon_0 m c^3}$$

$$m\ddot{\mathbf{r}} = \mathbf{F} + m\tau \ddot{\mathbf{r}}$$

这样一个简化就会招致祸端: 若在一维情况下, 将受力设置为以下元脉冲函数(unit pulse function):

$$F(t) = \begin{cases} 0 & (t < -0.5) \\ 1 & (-0.5 \leq t \leq 0.5) \\ 0 & (t > 0.5) \end{cases}$$

通过解数值方程 $\tau \ddot{x} - \ddot{x} = -F(t)/m$, 把 $a = \ddot{x}$ 视作待解函数, 可得通解:

$$a(t) = \begin{cases} A_1 e^{t/\tau} & (t < -0.5) \\ A_2 e^{t/\tau} + \frac{1}{m} & (-0.5 \leq t \leq 0.5) \\ A_3 e^{t/\tau} & (t > 0.5) \end{cases}$$

⁷每一点的场强全都是独立的变量. 类似于一张连续的弹性膜, 膜的位移就相当于场的激发.

且三个系数满足使 a 为连续函数的条件:

$$A_1 - A_2 = \frac{e^{0.5/\tau}}{m}, \quad A_3 - A_2 = \frac{e^{-0.5/\tau}}{m}$$

从中可以看出 $A_3 \neq A_1$. 这就导致了一个尴尬的结果: 在 $t < -0.5, t > 0.5$ 两段过程中, 至少有一段过程电荷做非匀速直线运动. 通过因果律来分析, 当力尚未作用在电荷上时, 电荷自然没有道理产生加速度. 故应取 $A_1 = 0$. 那么在这个外力作用于电荷一段时间后并撤销后, 这个理论预言电荷已经被一个扰动后的电磁场挟持以致做以下运动:

$$A_3 = -\frac{e^{0.5/\tau} - e^{-0.5/\tau}}{m}$$

$$a = -\frac{e^{0.5/\tau} - e^{-0.5/\tau}}{m} \cdot e^{t/\tau}$$

无疑, 这是荒诞的, 加速度不仅不是零, 甚至还随时间趋近于无穷大, 这种病态的行为显然没有在实际观测中出现过. 其实, 根据这个辐射阻尼力的推导过程中忽略的高阶项需要足够小的要求来看, 我们上面定义的 τ 恰好界定了这个近似可以进行的条件, 那便是我们关心的过程的 $\ddot{\mathbf{r}}$ 变化的特征时间 T 需要不大于对电子以上定义的特征时间的量级:

$$T \ll (\sim) \tau = \frac{q^2}{6\pi\varepsilon_0 mc^3} = 6.27 \times 10^{-24} \text{ s}$$

在知道这个辐射阻尼力的做法的适用条件之后我们再来看一个常见的简单推导方法. 我们注意到辐射阻尼公式与辐射功率拉莫尔公式系数的高度相似性. 不难去思考两者的联系, 实际上:

$$\frac{q^2 \ddot{\mathbf{r}}^2}{6\pi\varepsilon_0 c^3} = \frac{d}{dt} \left(\frac{q^2 \dot{\mathbf{r}} \cdot \ddot{\mathbf{r}}}{6\pi\varepsilon_0 c^3} \right) - \frac{q^2 \ddot{\mathbf{r}} \cdot \dot{\mathbf{r}}}{6\pi\varepsilon_0 c^3}$$

可见辐射阻尼力对粒子做负功的功率与拉莫尔辐射功率之间仅仅相差一项. 我们指出在我们研究的电磁波照射到振子的经典色散模型中, 电子做周期性运动, 故这一项的周期积分为零:

$$\int_0^T \frac{d}{dt} \left(\frac{q^2 \dot{\mathbf{r}} \cdot \ddot{\mathbf{r}}}{6\pi\varepsilon_0 c^3} \right) dt = \frac{q^2 \dot{\mathbf{r}} \cdot \ddot{\mathbf{r}}}{6\pi\varepsilon_0 c^3} \Big|_0^T = 0$$

故如果一个表达式能代表辐射阻尼, 它必然具有上式具有的形式. 而在 $T \ll (\sim) \tau$ 时, 正如我们之前介绍的那样, 这个做法是合理的.

利用特征时间 τ , 在频率 ω 的电磁波下的谐振子的动力学方程, 考虑辐射阻尼时, 就写作:

$$m\ddot{\mathbf{r}} - m\tau\ddot{\mathbf{r}} + k\mathbf{r} = -e\mathbf{E}$$

这看上去与之前我们给出的谐振子的形式存在出入. 实际上, 在简谐 $\mathbf{E} = \mathbf{E}_0 e^{i\omega t}$ 的情况下, 对时间导数相当于乘以 $i\omega$. 于是上式写为:

$$m\ddot{\mathbf{r}} + m\omega^2 \tau \dot{\mathbf{r}} + k\mathbf{r} = -e\mathbf{E}$$

这就可以看出, 实际上之前的经典色散理论不过是将 $m\omega^2 \tau$ 整体取做 γ 而已. 引入辐射阻尼的现象之后我们就可以看出 γ 与角频率潜在的比例关系 $\gamma \propto \omega^2$.

在这个基础上我们解出单电子某维度上做振动的函数:

$$x = Ae^{i\omega t} = -\frac{eE_x/m}{\omega_0^2 - \omega^2 + i\omega^3 \tau}$$

通过这个振幅我们除了推导色散, 还可以得到两个十分有意义的结果:

第一个结果是重新去计算电子辐射的功率, 即把外电磁场能量单位时间散射的那部分能量. 根据偶极振荡公式:

$$\bar{P} = \frac{e^2 \omega^4 |x|^2}{12\pi\varepsilon_0 c^3}$$

而入射电磁波,如果就是沿 z 方向传播,沿 x 方向偏振的平面波,则具有平均能流:

$$S = \frac{1}{\mu_0 c} \cdot \frac{|E_x|^2}{2} = \frac{\epsilon_0 c |E_x|^2}{2}$$

用这个功率除以这个能流,我们得到一个面积量纲的物理量.实际上就是相当于电子吸收了这个面积里面的辐射并以散射的形式做了新的辐射发射.这个面积称作**散射截面**(scattering cross section),计算可得:

$$\sigma = \frac{\bar{P}}{S} = \frac{8}{3} \pi r_e^2 \cdot \frac{1}{\omega^2 \tau^2 + (1 - \omega_0^2/\omega^2)^2}$$

而其中 r_e 就是我们之前提过的**经典电子半径**(classical electron radius).它恰恰是在洛伦兹等人在研究辐射的问题时提出的.故也称**洛伦兹半径**(Lorentz radius):

$$r_e = \frac{e^2}{4\pi\epsilon_0 mc^2}$$

我们首先注意一点:由于 τ 是个很短的时间,故能够使得 $\omega\tau \sim 1$ 的角频率对应的光波具有以下的量级:

$$\omega \sim 10^{23} \text{ rad/s} , \quad \lambda \sim 12 \text{ fm} , \quad E \sim 105 \text{ MeV}$$

这实际上是一个极高的频率了,到这个能标,光子对电子的影响已经完全不能忽略光子的量子性,当然也要考虑电子的波动性,康普顿散射和量子电动力学的理论时必要的.而物质形成的体系总是有 $\omega_0\tau \ll 1$,故即使在 $\omega \gg \omega_0$,一般也可以把 $\omega\tau$ 视作小量.那么在这种场合下:

$$\sigma \approx \frac{8}{3} \pi r_e^2$$

这个截面称为**汤姆孙截面**(Thomson cross section).反映了在高频电磁波作用下电子接近自由电子散射电磁波的行为.

而在低频下 $\omega \ll \omega_0$,可做近似:

$$\sigma \approx \frac{8}{3} \pi r_e^2 \cdot \frac{\omega^4}{\omega_0^4}$$

这个式子后面会介绍,反映了瑞利散射的基本原理.

最后,在 $\omega \approx \omega_0$ 时,截面表达式的分母很接近零.散射截面远大于汤姆孙截面,电子强烈吸收外电磁场能量,并造成大的散射.这称作**共振散射**(resonance scattering).实验上实际上也观察到了这个现象.但是洛伦兹经典色散理论的弱点就在于它是个经典的理论,并不能很好地解释实验中得到的数据,因为它没有构建出理论模型中谐振子的量子本质.

第二个结果,我们来考察一下这个振子在外界电磁场下产生的振动能量:

$$\begin{aligned} K &= \overline{\frac{1}{2}k\Re e(x)^2 + \frac{1}{2}m\Re e(\dot{x})^2} \\ &= \frac{1}{4}m(\omega^2 + \omega_0^2)|x|^2 \\ &= \frac{e^2}{2\epsilon_0 m \omega^2} \cdot \frac{\omega^2(\omega^2 + \omega_0^2)}{(\omega^2 - \omega_0^2)^2 + \omega^6 \tau^2} \cdot \frac{1}{2}\epsilon_0 |E_x|^2 \end{aligned}$$

这样也许还不足以引起理论上的价值.但是如果设想一个复杂的电磁波场,它总可以分解为简单的平面波场(参考下一节介绍的傅里叶变换理论).那么该点的瞬时 x 方向电场就可以分解为:

$$E_x(t) = \int_0^\infty [A(\omega) \cos \omega t + B(\omega) \sin \omega t] d\omega$$

而对于波 $A(\omega) \cos \omega t$ 和 $B(\omega) \sin \omega t$,在谐振子上激发的能量是:

$$K(\omega) = \frac{e^2}{2\epsilon_0 m \omega^2} \cdot \frac{\omega^2(\omega^2 + \omega_0^2)}{(\omega^2 - \omega_0^2)^2 + \omega^6 \tau^2} \cdot \frac{1}{2}\epsilon_0 [A(\omega)^2 + B(\omega)^2]$$

作为原因,多个谐振电场的叠加产生的结果就是简单的振子的振动的简单叠加,这是因为谐振电场与振子的振动满足的微分方程是线性的.而按照简谐分解的一个数学结果就是,能量也可以直接叠加:

$$\overline{\frac{1}{2}\varepsilon_0 E_x^2(t)} = \int_0^\infty \left[\frac{1}{2}\varepsilon_0 A(\omega)^2 \overline{\cos^2 \omega t} + \frac{1}{2}\varepsilon_0 B(\omega)^2 \overline{\sin^2 \omega t} \right] d\omega = \int_0^\infty \frac{1}{2} \frac{1}{2} \varepsilon_0 [A(\omega)^2 + B(\omega)]^2 d\omega$$

从这个角度来看,振子就会因此具有 x 方向的平均能量:

$$\begin{aligned} K_x &= \int K(\omega) d\omega \\ &= \int \frac{e^2}{2\varepsilon_0 m \omega^2} \cdot \frac{\omega^2 (\omega^2 + \omega_0^2)}{(\omega^2 - \omega_0^2)^2 + \omega^6 \tau^2} \cdot \frac{1}{2} \varepsilon_0 [A(\omega)^2 + B(\omega)]^2 d\omega \end{aligned}$$

如果考虑一个热力学系统,辐射和谐振子达到了一个温度为 T 的热平衡.那么经典理论认为:每一个振子在一个方向上应当具有平均能量 kT .这一点之后由普朗克发现是错误的.但是如果我们将坚持这个做法,由于电磁辐射的内能密度亦可以根据频率展开:

$$\begin{aligned} u &= \overline{\frac{1}{2}\varepsilon_0 E_x^2(t)} + \overline{\frac{1}{2}\varepsilon_0 E_y^2(t)} + \overline{\frac{1}{2}\varepsilon_0 E_z^2(t)} + \overline{\frac{1}{2\mu_0} B_x^2(t)} + \overline{\frac{1}{2} B_y^2(t)} + \overline{\frac{1}{2\mu_0} B_z^2(t)} \\ &= 6 \cdot \overline{\frac{1}{2}\varepsilon_0 E_x^2(t)} \\ &= \int_0^\infty u(\omega d\omega) \end{aligned}$$

这就得到了:

$$u(\omega) = 3 \cdot \frac{1}{2} \varepsilon_0 [A(\omega)^2 + B(\omega)]^2$$

从而之前的积分就可用 $u(\omega)$ 表示.但是又由于 τ 很小,故仅在 $\omega \sim \omega_0$ 附近近似:

$$\begin{aligned} K_x &= \int \frac{e^2}{2\varepsilon_0 m \omega^2} \cdot \frac{\omega^2 (\omega^2 + \omega_0^2)}{(\omega^2 - \omega_0^2)^2 + \omega^6 \tau^2} \cdot \frac{u(\omega)}{3} d\omega \\ &= \int \frac{e^2}{2\varepsilon_0 m \omega_0^2} \cdot \frac{2\omega_0^4}{4\omega_0^4(\omega - \omega_0)^2 + \omega_0^6 \tau^2} \cdot \frac{u(\omega_0)}{3} d\omega \\ &= \frac{e^2 u(\omega_0)}{12\varepsilon_0 m \omega_0^2} \cdot \int \frac{dx}{x^2 + \omega_0^2 \tau^2 / 4} \\ &= \frac{e^2 u(\omega_0)}{12\varepsilon_0 m \omega_0^2} \cdot \frac{2\pi}{\omega_0 \tau} \\ &= kT \end{aligned}$$

代入 τ 的表达式,这就得到了:

$$u(\omega_0) = \frac{3kT \omega_0^3}{2\pi^2 c^3}$$

这就是辐射与振子热力学平衡的理论基础,从此建立的理论可以导出著名的瑞利-金斯公式,从而招致了著名的紫外灾难.普朗克提出了辐射的量子理论成功地解决了这个问题.参见下一章.

第二十九章 量子论

29.1 黑体辐射

- 黑体辐射的性质:

- 总辐射强度:

$$J = \sigma T^4 = \frac{1}{4} u c$$

$$\sigma = 5.67 \times 10^{-8} \text{W}/(\text{m}^2 \cdot \text{K}^4) = \frac{2\pi^5 k^4}{15 h^3 c^2}$$

- 辐射频域谱: 普朗克公式:

$$J(\omega) = \frac{\hbar\omega^3}{4\pi^3 c^2} \frac{1}{e^{\frac{\hbar\omega}{kT}} - 1}$$

- 辐射空间谱: 郎伯定律:

$$\frac{dJ}{d\Omega} = \frac{J}{\pi} \cos \theta$$

- 非黑体满足基尔霍夫定律: 某角频率 ω 下发射率 $e(\omega)$ 与吸收率 $a(\omega)$ 必然相等. 发射率是

$$\frac{dE}{dSdt} = \int E(\omega) d\omega$$

$$e(\omega) = E(\omega)/J(\omega)$$

故基尔霍夫定律发现:

$$e(\omega) = a(\omega)$$

29.2 光粒子性

- 光子概念:

$$E = \hbar\omega , \quad p = \hbar k$$

- 用于解释普朗克公式: 玻色-爱因斯坦统计.

- 用于解释光电效应:

$$E_k = h\nu - W \geq eU \rightarrow I$$

29.3 玻尔原子

- 精细结构常数(fine structure constant):

$$\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c} \approx \frac{1}{137}$$

- 半经典假设: 忽略电子的波动性, 认为仍然有轨道运动, 且为圆周运动, 符合牛顿力学:

$$m \frac{v_n^2}{r_n} = \frac{e^2}{4\pi\varepsilon_0 r_n^2}$$

- 量子化假设: 角动量是量子化的, 或者作用量量子化:

$$mv_n \cdot 2\pi r_n = nh$$

- 定态跃迁假设: 基态 $n = 1$ 是能量最低的稳定态, 其他态与基态彼此之间都有可能发生转变, 转变是瞬间发生的, 伴随着一个光子的吸收或发射, 且需要能量守恒:

$$\begin{aligned} E_n &= \frac{1}{2}mv_n^2 - \frac{e^2}{4\pi\varepsilon_0 r_n} \\ h\nu &= E_n - E_m \end{aligned}$$

- 玻尔理论的解:

$$\begin{aligned} r_n &= n^2 a_0 \\ v_n &= \frac{\alpha c}{n} \\ E_n &= -\frac{E_0}{n^2} = -\frac{\alpha^2 mc^2}{2n^2} \end{aligned}$$

其中 E_0 为第一电离能, 氢原子为 13.6eV. 而 a_0 为玻尔半径, 它与另外两个特征长度: 电子康普顿波长:

$$\lambda_e = \frac{h}{mc}$$

经典电子半径:

$$r_e = \frac{e^2}{4\pi\varepsilon_0 mc^2}$$

符合;

$$\alpha = \frac{\lambda_e/2\pi}{a_0} = \frac{a_0}{r_e}$$

29.4 物质波与波函数

- 匀速运动的粒子对应一个平面波;

$$\psi = A e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}$$

且符合德布罗意关系:

$$E = \hbar\omega$$

$$\mathbf{p} = \hbar\mathbf{k}$$

- 对这个波的诠释: 模方表示粒子在某空间点被探测到的概率分布函数:

$$\begin{aligned} f(\mathbf{r}) &= |\psi|^2 \\ \int f(\mathbf{r}) dV &= 1 \end{aligned}$$

- 测不准原理: 平面波情况下粒子位置是不确定的, 但是动量却唯一. 任意一个波函数可以分解为不同概率的平面波的叠加, 那么某方向上位置和动量的两个不确定度(标准差)满足:

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

第三十章 物理学尺度

30.1 宇观

30.2 宏观

30.3 介观

30.4 微观原子

30.5 微观亚原子