微积分简单入门

by 2516-渚枝惊雀

引言

微积分被费曼称为“上帝的语言！”[1.]．自从牛顿[2.]和莱布尼茨发明了这一奇妙的语言，自然界的许多问题迎刃而解，人们对自然界有了更全面更深入的认识．微积分听起来高深莫测，实际上与我们的生活悉悉相关，学习一点微积分，有助于我们解决一些实际问题．

了解微积分学的基本知识前，我们不妨来看一看这一奇妙语言的发明．

在牛顿和莱布尼茨之前，勒内·笛卡尔[3.]和皮埃尔·德·费马[4.]曾对切线之类的问题产生争执．笛卡尔坚持用他的曲率圆法求切线问题，而来自图卢兹的律师、天才的数学家费马，则使用了类似下面要提到的极限的办法（笛卡尔对于费马先于他并优于他解决了切线问题深感懊恼[5.]）．但是，费马最后并没有将微积分两者统一到一起．在他之后，牛顿和莱布尼茨各自独立地发现了这一伟大的工具，微积分学迎来新的发展．

总的来说，微积分学有**三大基本问题**[6.]（如图0.1）：

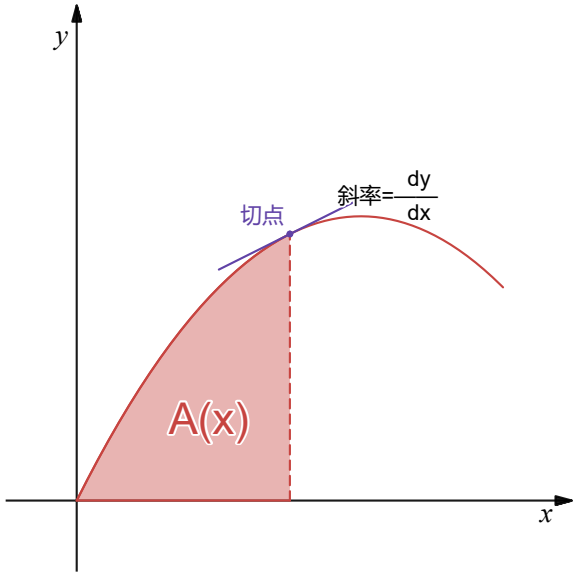


Fig 0.1

① 已知一条曲线，求它各处的斜率（正向问题）；

② 已知一条曲线各处斜率，求这条曲线（反向问题）；

③ 已知一条曲线，求曲线下方的面积（面积问题）.

下面我们来讨论如何用微分和积分解决这些问题.

一、微分

如果你已经学习过或看过高中数学教材选择性必修二，你会了解到**导数**，说到底，导数和微分可以几乎画上等号，它们以不同的形式表示相同的内容，下面看一个例子：

问题一 一物体以加速度 $\vec{a}$ 沿直线做匀加速运动，求物体所受合外力 $\vec{F}$ ？

解：很明显，由牛顿第二定律，我们有：

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

我们来分析一下这个式子，对于此物体，在任意时刻，其所受合外力均等于它的质量乘以其加速度，把这个结论推广一下，对于任一物体，无论它做什么运动，在任意时刻，该方程均成立.

事实上， $\vec{F}=m\vec{a}$ 是一个**微分方程**，它表示的是某一瞬间物体加速度与其受力之间的关系.

从问题一，我们了解到，微分实际上描述的是某一瞬间的状态，我们可以通过解析几何来进一步理解它的实际意义.

问题二 对于抛物线 $y=x^2$ ，求它在 $x=1$ 处的切线斜率.

分析：对于由线 $y=x^2$ ，求其在 $x=1$ 处的斜率实际上是求在 $x=1$ 处与抛物线相切的直线的斜率，我们不妨先研究几种容易计算的情况，来推理一下曲线的切线如何计算.

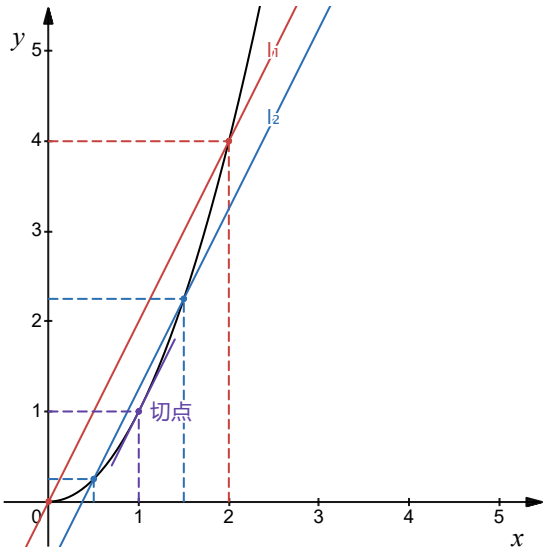


Fig 1.1

如上图1.1，我们取 $(0,0)$ 与 $(2,4)$ 两点，连接起来，得到直线 l_1 那么容易得出 l_1： $y=2x$.

我们可以继续取 $(\frac{1}{2},\frac{1}{4})$ 与 $(\frac{3}{2},\frac{9}{4})$ 两点，连接得到 l_2 ，计算 l_2 的斜率.

同样继续二分求 $(\frac{3}{4},\frac{9}{16})$ 与 $(\frac{5}{4},\frac{25}{16})$ 连线 l_3 斜率.....

做完以上运算的读者，可以猜想，如果取的两点无限接近于 $(1,1)$ ，那么这条直线是否就是在该点的切线呢?

事实上，这是我们在数学上定义导函数的方法，导函数（简称导数）表示了函数图像上任意一定的切线斜率．导函数用 y' 或 $f'(x)$ 表示，以问题二中的函数为例，计算方法如同取两点算直线斜率一样，像下面这样:

在 $y=f(x)$ 上取 $(x,f(x))$ ，$(x+\Delta x,f(x+\Delta x))$ 两点，

则

$$f'(x)=\frac{\Delta y}{\Delta x}=\frac{f(x+\Delta x)-f(x)}{\Delta x}=\frac{(x+\Delta x)^2-x^2}{\Delta x}=\frac{2x\cdot\Delta x+(\Delta x)^2}{\Delta x}$$

由于我们取的两点无限接近，故 $\Delta x\to 0$ ，这样的话， $(\Delta x)^2$ 应是极小的数，可以忽略．原式可以写作:

$$f'(x)=\frac{2x\cdot\Delta x}{\Delta x}=2x$$

导函数的计算通式为:

$$f'(x)=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

其中 $\lim\limits_{\Delta x\to 0}$ 表示极限情况 Δx 趋近于 0 .

这么写有点麻烦，把它换成微分的形式，我们用 $\frac{dy}{dx}$ 表示 $\lim\limits_{\Delta x\to 0}\frac{\Delta y}{\Delta x}$.

那么 $\frac{dy}{dx}=f'(x)$ ，它和导函数表示的意义是相同的，以下列出**常见函数的导函数**:

1. 常函数: $(A)'=0$
2. 幂函数: $(x^n)'=nx^{n-1}$
3. 三角函数:

$$\begin{aligned}(\sin x)'&=\cos x & (\cos x)'&=-\sin x\\ (\tan x)'&=\frac{1}{\cos^2 x} & (\cot x)'&=-\frac{1}{\sin^2 x}\end{aligned}$$

4. 指数函数: $(a^x)'=a^x\ln a$ 特别地, $(e^x)'=e^x$
5. 对数函数: $(\log_a x)'=\frac{1}{x\ln a}$ 特别地, $(\ln x)'=\frac{1}{x}$

这样我们很容易解决问题二:

解: $f'(x)=2x$

$\therefore f'(1)=2\times 1=2$

即 $y=x^2$ 在 $x=1$ 处切线斜率为 2 .

自此三大基本问题中的 ① 被我们解决了.

微分学在数学、物理学、化学、生物学等众多学科中有着重要作用，有的时候，单独一个自变量不能全面地描述因变量的情况，我们使用偏微分来表示多个变量的情况，由于偏微分本身的复杂性，在这里不为大家介绍其计算方法，仅举一例来向读者展示偏微分的符号表示及用途.

问题三 （思考）对于一维的弹性直杆（如图1.2），在它的内部发生质点的平动，且已知质点的振动方向与传播一致，我们把这样传播的波称为一维弹性纵波 (p) ．对于它的波动，我们有以下两个方程[7.]：

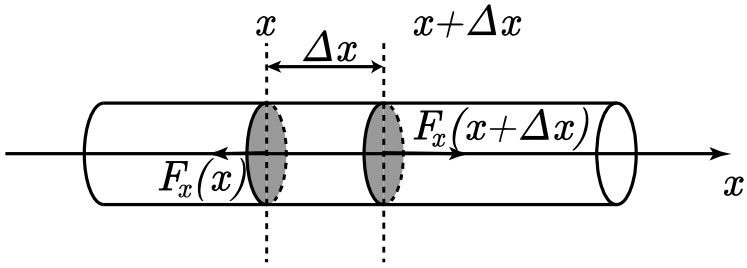


Fig 1.2

$$\rho\frac{\partial^2 u(x,t)}{\partial t^2}=\frac{\partial\sigma_{xx}(x,t)}{\partial x}$$

（一维弹性动力学方程）

$$\frac{\partial^2 u(x,t)}{\partial t^2}={v_p}^2\frac{\partial^2 u(x,t)}{\partial x^2}$$

（一维弹性纵波波动方程）

其中， ρ 为密度， σ 表示应力， v_p 为纵波速度． $\frac{\partial^2 u(x,t)}{\partial t^2}$ 即是函数 u 对 t^2 的二阶偏导形式．

这个方程的证明不予赘述，有兴趣的读者可以自己尝试一下或者来找笔者探讨．

二、积分

如果把微分看成从整体到局部（一个非常小的局部）的话，那么积分事实上就是从局部再回到整体，下面通过一个例子来解释一下．

问题四 （如图2.1）抛物线被一直线所截，截得的弓形内有一面积最大的三角形，其面积为1．求弓形的面积．

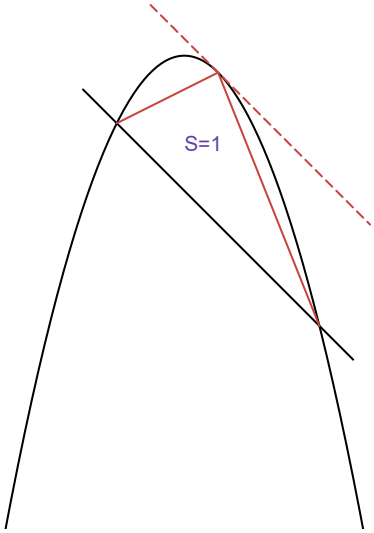


Fig 2.1

这个问题是阿基米德在他的《抛物线求积法》[8.]中提出的．容易得到：面积最大的三角形的顶定应为与直线平行的抛物线切线的切点．

继续分割大三角形外空白部分（如图2.2）．

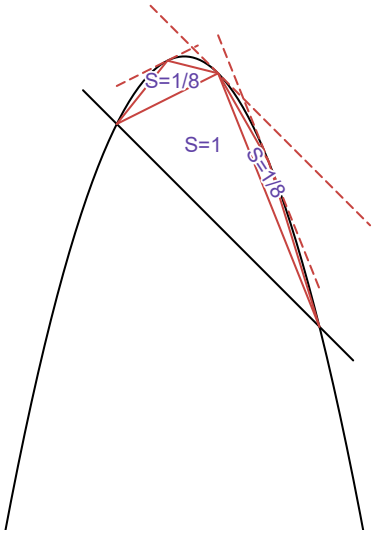


Fig 2.2

阿基米德用数学方法求出两个次一级小三角形的面积各占大三角形的 $\frac{1}{8}$，那么我们称第

二级的两个三角形面积之和为 $\frac{1}{4}$ ．

可以由此得知：之后每一层积的三角形面积之和均为上一层级三角形总面积的 $\frac{1}{4}$ ．

抛线的故抛物线弓形的面积 $S=1+\frac{1}{4}+\frac{1}{16}+\frac{1}{64}+\cdots$

这是一个无穷级数.

对于S的求值，我们应用等比数列的计算方式，那么有：

$$4S = 4 + 1 + \frac{1}{4} + \frac{1}{16} + \cdots = 4 + S$$

$\therefore S = \frac{4}{3}$

在这一过程中，我们先将面积无限分割再求其和，实际上就是先微分再积分的思想．阿基米德的弓形很好地为我们说明了微分与积分的应用逻辑.

- **微积分基本定理**:

把微分和积分联系在一起的是微积分基本定理，它使我们能在短短几分钟内解决基本问题 ② 和 ③ ．要知道，无数数学家花了一辈子也没能解决它们！要理解这个定理，我们先来看一个物理问题.

问题五 一个物体从静止开始以加速度 a 做匀加速直线运动，求它在 t 时间内通过的路程.

解：显然， $x = \frac{1}{2}at^2$ ，这是运动学的一个基本方程.

让我们作出这个运动的 $v - t$ 图像（如图2.3）：

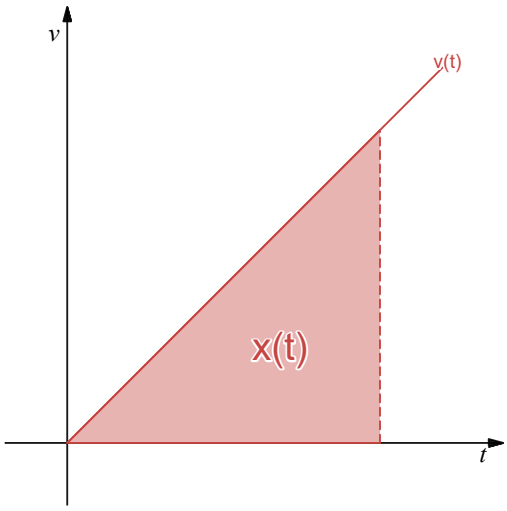


Fig 2.3

这条曲线的解析式容易得到： $v(t) = at$.

那么，阴影部分的面积即为位移，且：

$$x(t)=\frac{1}{2}at^2$$

我们解决了一种最简单的面积问题.

把这种思想扩大到任意曲线（如图2.4）：

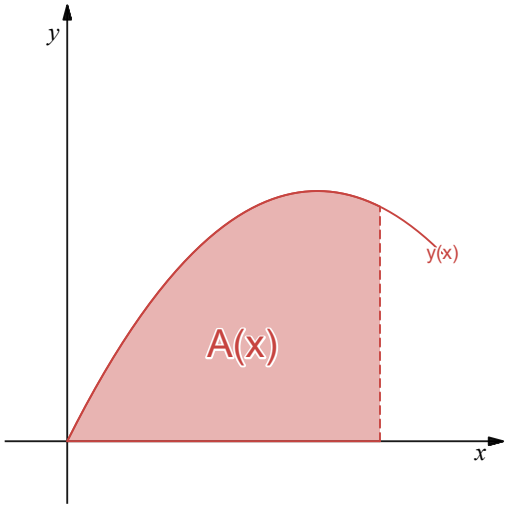


Fig 2.4

我们可以想象[9.]有一个滚筒，从原点出发，始终紧贴 x 轴与曲线顶端将这部分之间的面积刷上油漆（如图2.5），这个滚筒长度的变化应当就是 $y(x)$. 那么，在极短的 $\mathrm{d}x$ 内，滚筒滚过的面积应当是 $y(x)\mathrm{d}x$ ，即 $\mathrm{d}A(x)=y(x)\mathrm{d}x$.

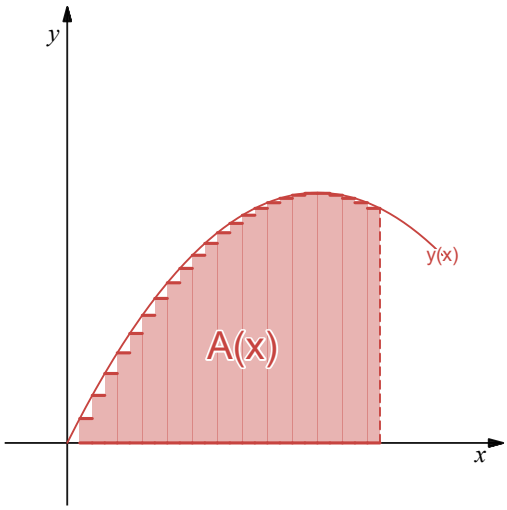


Fig 2.5

或者换一种形式（如图2.6）

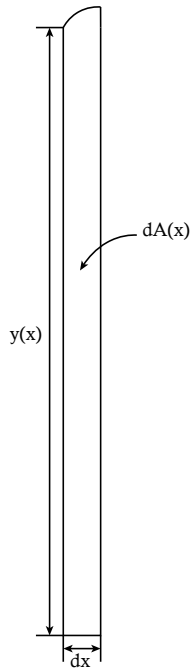


Fig 2.6

因为 dx 极小，顶端的类似曲边小三角形的部分面积可以完全忽略.

那么 $\mathrm{d}A(x) = y(x)\mathrm{d}x$.

换句话说， $y(x)$ 是函数 $A(x)$ 的变化率，也可以写成导数形式，即：

$$\frac{\mathrm{d}A(x)}{\mathrm{d}x} = y(x)$$

这就是微积分基本定理，利用这个定理，我们可以轻松解决基本问题 ② 和 ③ ：

因为：

$$\frac{\mathrm{d}A(x)}{\mathrm{d}x} = y(x)$$

那么求 $A(x)$ 即对 $\mathrm{d}A(x) = y(x)\mathrm{d}x$ 求和

我们把这一过程写作：

$$A(x) = \int_0^x \mathrm{d}A(x) = \int_0^x y(x)\mathrm{d}x$$

其中，符号“ $\int$ ”是积分号，表示极小量求和． $\int_0^x$ 表示从 0 到 x 的 $\mathrm{d}A(x)$ 的和，实际上

$$A(x_{12}) = \int_{x_1}^{x_2} y(x)\mathrm{d}x = A(x_2) - A(x_1)$$

这就是面积问题的解决方法.

至于反向问题，从前面的推导，我们知道， $y(x)$ 就是 $A(x)$ 的导函数. 所以，已知一条曲线各处斜率，求原函数，我们只需对已知的斜率函数求积分即可. 积分的简单计算也要依照这个找原函数的技巧，即不定积分的结果就是原函数. 至此，微积分学的三大基本问题被全部解决.

三、微分拓展：泰勒级数展开$^{<10.>}$

通过已学过的数学知识，我们知道：三维空间内，任意一个向量可以写成这样的形式：

$$\vec{v} = x\hat{i} + y\hat{j} + z\hat{k}$$

其中 $\hat{i}$, $\hat{j}$, $\hat{k}$ 是空间的一组基底.

希尔伯特（Hilbert）将其推广到了无限维欧几里得空间，我们称之为**希尔伯特空间** $\mathscr{H}$.

泰勒将这一概念拓展到函数领域，他说：类似地，可以把任意函数分解成这样的形式：

$$y(x) = A_0x^0 + A_1x^1 + A_2x^2 + \cdots + A_nx^n + \cdots$$

以 $(x^0, x^1, x^2, x^3, \ldots)$ 为基底，所有的 A_n $(n \in \mathbb{N})$ 一定是某个常数

我们把上式简写为

$$y(x) = \sum_{n=0}^{\infty} A_nx^n$$

那么对于任意函数，我们只需要求出 A_n $(n \in \mathbb{N})$ ，这个函数即可确定.

如何求出这些常系数呢?

首先，将 $x = 0$ 代入上式，有：

$$y(0) = A_0$$

这样我们得到了 A_0 .

我们观察到，函数右边部分是幂函数的加和，而对于 $y = x^n$ ，其导数 $y'(x) = nx^{n-1}$ ，相当于降了一次幂.

那么，我们可以这样求解 A_1 ：

对原式求一次导，

$$y'(x) = A_1 + 2A_2x + 3A_3x^2 + \cdots$$

再次取x = 0．我们有：

$$y'(0) = A_1$$

重复以上操作，可以依次得到：

$$A_2 = \frac{1}{2}y''(0) \qquad A_3 = \frac{1}{3 \times 2}y'''(0) \qquad A_4 = \frac{1}{4 \times 3 \times 2}y^{(4)}(0)$$

写出通式：

$$A_n = \frac{1}{n!}y^{(n)}(0)$$

y 右上方的 $^{(n)}$ 表示对 y 进行 n 次求导， $n!$ 表示**阶乘**：

$$n! = \prod_{i=1}^{n} i = n(n-1)(n-2)\cdots 3 \times 2 \times 1$$

特别地， $0! = 1$ ．由此，我们可以将以上函数写成这样的形式：

$$y(x) = \sum_{n=0}^{\infty} \frac{y^{(n)}(0)}{n!}x^n$$

这就是**泰勒级数展开**.

下面来看一道例题：

问题六 试用泰勒级数展开证明：当 x 极小时，存在以下关系：

$$(1 \pm x)^n = 1 \pm nx$$

解：设函数 $f(x)=(1\pm x)^n$ ，计算其常系数.

由 $A_n=\frac{1}{n!}y^{(n)}(0)$ 得：

$$\begin{gathered}A_0=y(0)=1\\A_1=y'(0)=n(1\pm x)^{n-1}|_{x=0}=n\\A_2=\frac{1}{2}y''(0)=\frac{1}{2}n(n-1)(1\pm x)^{n-2}|_{x=0}=\frac{1}{2}n(n-1)\\\cdots\end{gathered}$$

由此类推： $f(x)=1\pm nx\pm\frac{1}{2}n(n-1)\pm\cdots$

$\because x$ 为极小量，

$\therefore x^2$ 及以后都是高阶小量，可以忽略

$\therefore f(x)=(1\pm x)^n=1\pm nx$.

证毕，这个关系被广泛应用于物理学问题.

四、小结

微积分学是一门高深的学问，它囊括了世间万物的运行规律．本文仅作为一篇科普性质的文章向读者介绍微积分学的简单内容．如有必要，在之后数期刊物上，将更加深入地介绍这一学科的内容．如洛必达法则，傅里叶级数展开，罗尔中值定理，拉格朗日中值定理，柯西中值定理等．一个人总是要学点微积分的，毕竟它是“上帝的语言！”.

- ## 引用与注释

  （包含书目推荐）

引言

1. "*It's the language God talks*":

   Wouk, *The language God Talks*, p.5.

2. *Isaac Newton*:

- For biographical information, see Gleik, *Isaac Newton*.

  See also Westfall, *Never at Rest*;

  and I.B.Cohen, "Isaac Newton", in vol.10 of Gillispie, *Complete Dictionary of Scientific Biography*, with amendments by G.E. Smith and W.Newman in vol.23.

- For Newton's mathematics, see Whiteside, *The Mathematical Papers*, vols.1 & 2;

  Edwards, *The Historical Development*; Grattan-Guinness, *From the Calculus*;

  Rickey, "Isaac Newton";

  Dunham, *Journey Through Genius*;

  Katz, *History of Mathematics*;

  Guicciardini, *Reading the Principia*;

  Dunham, *The Calculus Gallery*;

  Simmons, *Calculus Gems*;

  Guicciardini, *Isaac Newton*;

  Stillwell, *Mathematics and its History*;

  and Burton, *History of Mathematics*.

3. *René Descartes*:

- For his life, see Clarke, *Descartes*;

  Simmons, *Calculus Gems* pp.84-92;

  and Asimov, *Asimov's Biographical Encyclopedia* pp.106-108.

- For summaries of his math and physics intended for general readers, see Kline, *Mathematics in Western Culture*, pp.159-181;

  Edwards, *The Historical Development*;

  Katz, *History of Mathematics*, sections 11.1 & 12.1;

and Burton, *History of Mathematics*, section 8.2.

- For a scholarly historical treatment of his work in mathematics and physics, see Michael S. Mahoney, "Descartes: Mathematics and Physics," in Gillispie, *Complete Dictionary of Scientific Biography*;

  also online at *Encyclopedia Britannica*, https://www.encyclopedia.com/science/dictionaries-thesauruses-pictures-and-press-releases/decartes-mathematics-and-physics .

4. *Pierre de Fermat*:

   Mahoney, *Mathematics Career*, is the definitive treatment;

   Simmons, *Calculus Gems*, pp.96-105, is brisk and entertaining about Fermat (just as the author was with everything he writes. If you haven't read Simmons, you must).

5. *Fermat and Descartes locked horns*:

   Mahoney, *Mathematical Career*, ch.4.

6. *Three Main Questions*:

   Steven Strogatz, *Infinite Powers: How Calculus Reveals the Secrets of the Universe* ch.6, p.176.

一、微分

7. Que, Notes of the Earth Science.

二、积分

8. *Quadrature of the Parabola*:

   A translation of Archimedes's original text is in Heath, *The Works of Archimedes*, pp.233-252.

   - For the details I glossed over in the triangular-shard argument,see Edwards, *The Historical Development*, pp.35-39;

     Stein, Archimedes, ch.7;

     Laubenbacher and Pengelley, *Mathematical Expeditions*, section 3.2;

     and Stillwell, *Mathematics and its History*, section 4.4.

There are also many treatments available on the internet. One of the clearest is by Mark Reeder at https://www.bc.edu/mark-reeder/1103quadparas.pdf .

Another is by R.A.G. Seely at https://www.math.mcgill.ca/rags/JAC/NYB/exhaustion2.pdf .

As an alternative, Simmons, *Calculus Gems*, section B.3, uses an analytic geometry approach that you may find easier to follow.

9. For this imaginative statement of fundamental theorem of calculus, see Steven Strogatz, *Infinite Powers: How Calculus Reveals the Secrets of the Universe*, ch.7, p.210.

三、微分拓展：泰勒级数展开

10. The original of this part can be seen at 郑琦，*《强基计划物理一本通：给高中物理加点难度（上册）》*，中国科学技术大学出版社，第1.6章.

妙话均值不等式——“1”的另一种妙用

by 25届-韶华寻梦

需要的前置知识储备：基本不等式

在必修1第2章，我们学习过**基本不等式**：

$$a+b \geq 2\sqrt{ab} \quad (a,b>0,\ eq. \iff a=b) \tag{1}$$

（$eq.$，取等；iff（ $\iff$ ），当且仅当）.

此处我们不加证明地指出其原始形式——**均值不等式**：

$$\frac{\sum\limits_{i=1}^{n} a_i}{n} \geq \sqrt[n]{\prod_{i=1}^{n} a_i} \quad (a_i>0,\ eq. \iff a_1=a_2=\cdots=a_n) \tag{2}$$

这里：

$$\sum_{i=1}^{n} a_i = a_1+a_2+\cdots+a_n$$

$$\prod_{i=1}^{n} a_i = a_1+a_2+\cdots+a_n$$

我们也都学习过“1”的妙用，如下例：

引例 $a,b>0$， $a+2b=1$，求 $\frac{1}{a}+\frac{1}{b}$ 的最小值.

解

$$
\begin{aligned}
(\frac{1}{a}+\frac{1}{b}) &= (\frac{1}{a}+\frac{1}{b})(a+2b) \\
&= 1+2+\frac{a}{b}+\frac{2b}{a} \\
&\geq 3+2\sqrt{2}\ \ (eq. \iff \frac{a}{b}=\frac{2b}{a})
\end{aligned}
$$

故原式最小值为 $3+2\sqrt{2}$.

相信这个技巧是同学们熟悉的.

下面我们给出几道题目，介绍“1”的另几种妙用.

例1 $m,n\in\mathbb{Z}$ ， $1<m<n$. 证明： $(1+m)^n>(1+n)^m$.

证明

$$
\begin{aligned}
(1+n)^m &= \underbrace{(1+n)(1+n)\cdots(1+n)}_{m}\cdot\underbrace{1\cdot1\cdot\cdots\cdot1}_{n-m} \\
&\leq (\frac{m(1+n)+(n-m)}{n})^n \\
&= (1+m)^n
\end{aligned}
$$

$eq. \iff 1+n=1$ ，显然不可能成立.

$$\therefore (1+n)^m<(1+m)^n,\square.$$

本例中，“1”作为“占位符”出现，配凑出我们需要的次数.

例2 $n\in\mathbb{Z}$. 证明： $(1+\frac{1}{n})^n<(1+\frac{1}{n+1})^{n+1}$.

仿照上例显然，读者自证不难. 我们将在文末给出本题的证明过程.

例3　证明 $\lim\limits_{n\to\infty}\sqrt[n]{n}=1$.

证明

$$\begin{aligned}\sqrt[n]{n}&=\sqrt[n]{\sqrt{n}\cdot\sqrt{n}\cdot\underbrace{1\cdot1\cdot\cdots\cdot1}_{n-2}}\\&\le\frac{2\sqrt{n}+n-2}{n}\\&=\frac{2}{\sqrt{n}}+1\ \ (eq.\iff\sqrt{n}=1)\\\therefore\sqrt[n]{n}&<\frac{2}{\sqrt{n}}+1\end{aligned}$$

当 n 无穷大时，$\frac{2}{\sqrt{n}}$ 无限趋近于 0 ，可忽略不计.

$$\therefore\lim_{n\to\infty}\sqrt[n]{n}=1,\square.$$

本例意在让同学们感受此处“1”的用法，严格意义上，这里的证明是有瑕疵的．望各位大佬海涵．

至此正文结束（撒花！），下面是第二题的证明过程：

证明

$$\begin{aligned}(1+\frac{1}{n})^n&=(1+\frac{1}{n})(1+\frac{1}{n})\cdots(1+\frac{1}{n})\cdot1\\&\le(\frac{n(1+n)+1}{n+1})^{n+1}\\&=(1+\frac{1}{n+1})^{n+1}\end{aligned}$$

$eq.\iff 1+\frac{1}{n}=1$ ，显然不可能成立.

$$\therefore(1+\frac{1}{n})^n<(1+\frac{1}{n+1})^{n+1},\square.$$

数、对称性与群

by 2501-Evortexio.7

> *Algebra is generous; she often gives more than is asked of her.*
>
>

计数与对称

我们从很小就接触数字和运算．通过**计数**（counting），我们创造出一系列自然数并发现它们有无穷多个．接着我们在自然数中做加法（多次后继）和乘法（多次相同的加法）如 $3+14=17$，$7\times 4=28$ 等，经过推广发现它们的结果总是落在自然数中：<1.>

$$\begin{aligned}\times &: \mathbb{N}\times\mathbb{N}\to\mathbb{N}\\ + &: \mathbb{N}\times\mathbb{N}\to\mathbb{N}\end{aligned}$$

当我们试图寻找某次加法或乘法之前的数——即解一元一次方程——时，我们发现解并不总是落在自然数中．为此，我们又从减法——加法的逆——定义了负数，从除法——乘法的逆——定义了分数，使一元一次方程总是有解．我们把所有这些数叫做**有理数**（rational number）<2.>．接着人们在求解更高次代数方程的过程中发现了一些数，容易证明它们不是有理数，如 $\sqrt{5}$，$\sqrt[7]{6}$，$\sqrt{7+\sqrt{3}}$ 等；在几何学中，我们发现圆周率 π 不是有理数；在微积分中，我们又用级数展开的方法验证了 e 不是有理数．我们把它们统称为**无理数**（irrational number）<3.>．把所有有理数和无理数合在一起，我们验证了它们具有多种运算的完备性，并把它们叫做**实数**（real number）．

16世纪数学家 Cardano 在《大术》中给出了三次方程的一般解法，却意外得到了$\sqrt{-121}$ 的"不合理的"结果．Descartes 就此发明了**复数**（complex number）．后来 De Moivre 和 Euler 将复数与二维旋转联系起来，极大地扩展了其实际意义．

以上是我们熟知的故事，也大致是我们认识数的历程．

未来学习工科的同学还可能接触到 Hamilton 拓展复数发明的**四元数**（quaternion）<4.>，它们在三维旋转的表示上十分方便．在密码学和抽象代数 <5.> 中还会遇到**有限域**（finite field）中的元素．此外，对计算机有了解的同学一定很熟悉 **p 进数**（p-adic number）．这些都被称为"数"．

著名的 Leonhard Euler 和 Gauss 都曾经研究过一类数字对整数模的代数运算——**模算术**（modular arithmetic）[6.]．它是把**同余**（congurency）作为等价关系的一种新的算术系统．学过竞赛的同学可能知道，我们这里再简单介绍一下．

> Def 1.1 同余 同余类 模运算
>
> 设 a, b, m 为整数，若 m 整除 $a-b$，则称 a 与 b 关于模 m 同余．通常记为 $a \equiv b \pmod{m}$ ．
>
> 关于整数 m 同余的所有整数叫做一个**模-m 同余类**（congurent class modulo m）．
>
> 把取整数 a 除以正整数 m 的余数记作 $a \mod m$ ，$(a+b) \mod m$ 叫做 a 和 b 关于模 m 的加法，$(a \times b) \mod m$ 叫做 a 和 b 关于模 m 的乘法；求解形如 $a+x \equiv c \pmod{m}$ 的方程的运算叫做模减法，求解形如 $ax \equiv c \pmod{m}$ 的方程的运算叫做模除法；此外还可以类比出模乘方和离散对数．它们统称为**模算术**（modular arithmetic）．

举几个例子就可以简单地说明．

$$-4 \equiv 38 \equiv 31 \equiv 3 \pmod{7}$$

显然它们除以 7 的余数都是 3 ，或者说它们任意两者的差都是 7 的倍数．

一共有 3 个模-3 同余类，分别是 [7.]：

$$\begin{aligned}
[0]_3 &= 3\mathbb{Z} = \{\ldots, -6, -3, 0, 3, 6, \ldots\} \\
[1]_3 &= 3\mathbb{Z}+1 = \{\ldots, -5, -2, 1, 4, 7, \ldots\} \\
[2]_3 &= 3\mathbb{Z}+2 = \{\ldots, -4, -1, 2, 5, 8, \ldots\}
\end{aligned}$$

同理对于正整数 m ，也有 m 个模-m 同余类．

模运算类似于钟表的记时．八点以后七个小时是三点，即 $(8+7) \mod 12 = 3$ ．试着将钟表的刻度从 12 个改为任意正整数 m 个，便可以类比出模-m 算术．值得注意的是，模算术中除法、负数幂和对数运算并不是对算术数值取余数，而是模意义下的"逆运算"．

$$\begin{aligned}
5+8 \equiv 6 \pmod{7} &\implies (5+8) \mod 7 = 6 \\
3 \times 7 \equiv 1 \pmod{10} &\implies (3 \times 7) \mod 10 = 1 \\
3 \times 11 \equiv 1 \pmod{8} &\implies (1/3) \mod 8 \\
&\quad = 11 \mod 8 \\
&\quad = 3
\end{aligned}$$

注意模-m 算术的结果必须落在集合 $\{0,1,2,\ldots,m-1\}$ 中，这个集合我们常常记作 $\mathbb{Z}_m$ ；如果算术结果不在 $\mathbb{Z}_m$ 中，则要对结果关于 m 取余数使它落在该集合中．

再来回顾一下小学时学习的“对称”这个概念．

对称（symmetry）指的是一个对象经过某种操作，其形状与原来相重合的性质．我们在数学课上只学习过轴对称和中心对称两种对称性，但是在结构化学和有机化学中，我们遇到过许多其它的对称性．比如正多面体对称性（想想 $\mathrm{CH_4}$ 和 $\mathrm{SF_6}$ ）、n-重旋转轴（绕轴旋转 $2\pi/n$ 后与原图形重合，图中 $n=7$ ）、镜面对称（想想一对光学对映体）、旋转中心对称（顾名思义，其对称元素通常称作反轴）、平移对称性（想想晶体点阵和晶胞）等，下图（图1-1）展示了它们．

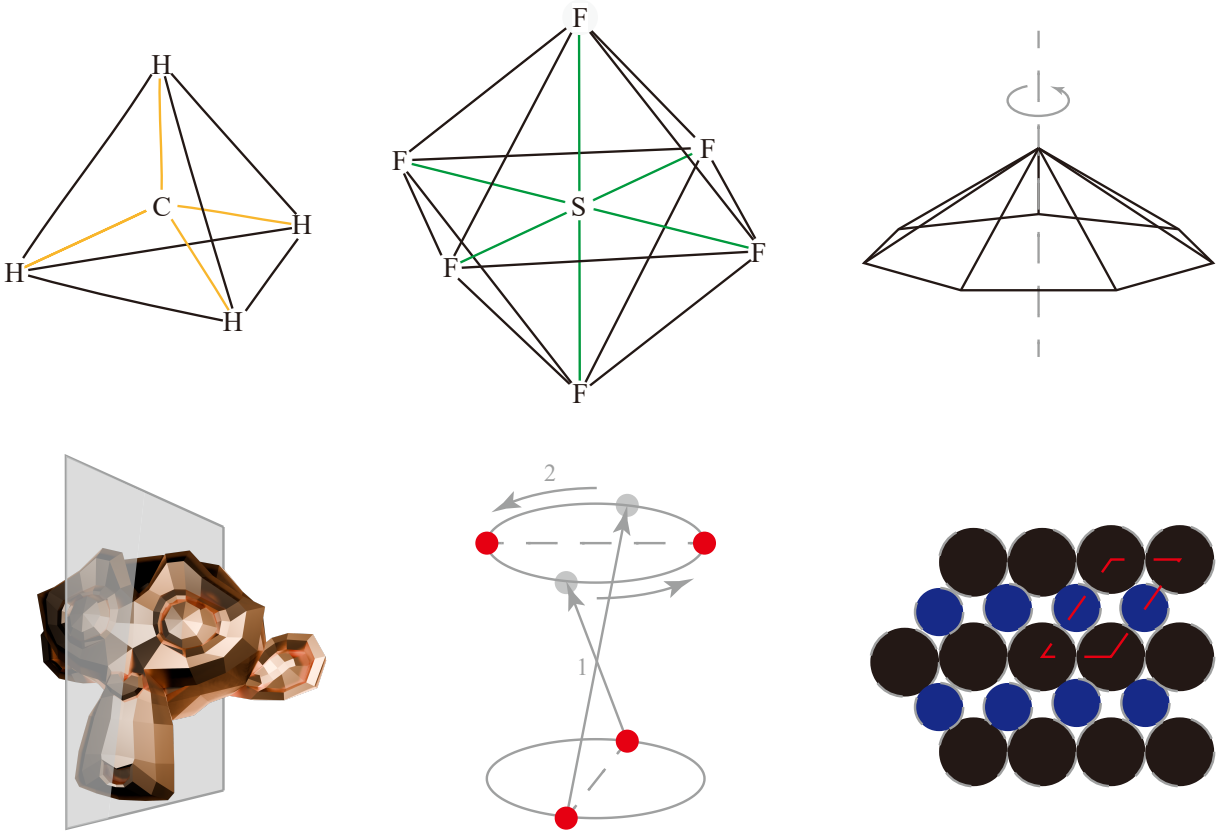


fig_1.1

但是，我们从未系统地研究过它们．

我们不妨从简单的平面正六边形开始．很容易发现它有12种对称操作，分别是（表1-1，图1-2）：

Open Access This article is licensed under CC BY-NC-SA 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-sa/4.0/

不变	**顺时针旋转** 60°	**顺时针旋转** 120°	**顺时针旋转** 180°	**顺时针旋转** 240°	**顺时针旋转** 300°
翻面	**翻面再顺时针旋转** 60°	**翻面再顺时针旋转** 120°	**翻面再顺时针旋转** 180°	**翻面再顺时针旋转** 240°	**翻面再顺时针旋转** 300°

table_1.1

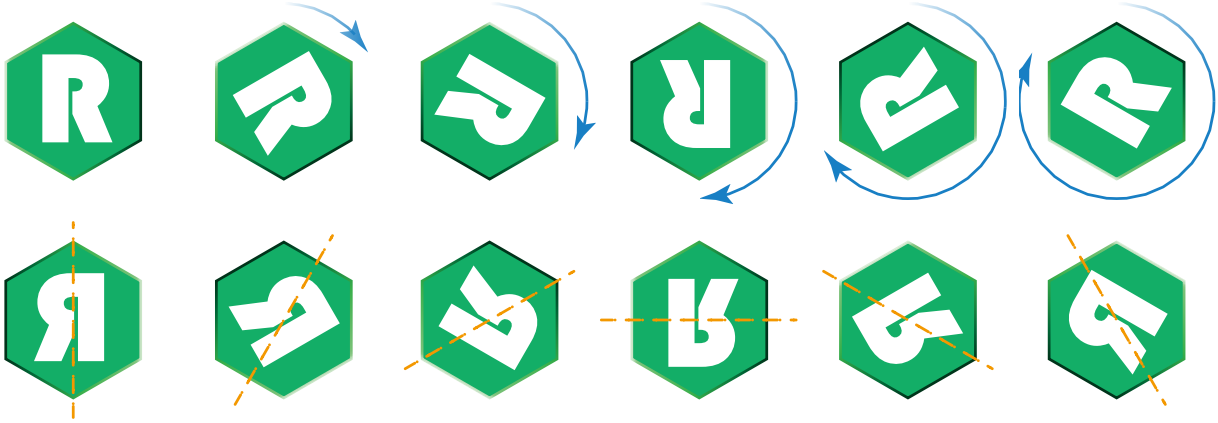


fig_1.2

其中翻面两次相当于不变，顺时针旋转 360° 相当于不变．我们发现，每种对称操作都是由几个“翻面”和“旋转 60° ”复合而成的．而任意两个操作相继进行，结果还是对称操作的一种；每种对称操作都有一个操作可以让它恢复为“不变”．

数与对称都是我们熟悉的东西．但是它们有什么关系呢?

群和它的表示

在解决五次方程通解的过程中，Lagrange 和 Ruffini 等数学家已经意识到了一些朴素的置换群理论的内容．后来在两位英年早逝的天才数学家 Niels Abel [8.] 和 Évariste Galois [9.] 的努力下，群论作为一门数学理论终于得以建立．后来又经过了几代数学家如 Cayley（他发明了后面要讲到的 Cayley 表）、Dedekind（就是这位通过著名的 Dedekind 分割完善了实数理论）、Klein（将群论与几何对称紧密联系）、Sophus Lie（他通过连续群研究微分方程，这类被命名为“Lie 群”的群成为数学和物理学中最重要的元素之一）的研究，群论逐渐完善，并与其他学科如影射几何和拓扑学相结合，筑起了现代数学的大厦．在这个过程中，很多数学家都曾对群提出了自己的定义，而第一个将数论、几何和多项式理论中的群用统一定义表述的是 Klein 的学生 Walther von Dyck，他在1882年首次给出了被认为是“现代的”群的定义．

我们先来给出我们的**群**（group）的定义．这是一种最常见、也相对现代的定义，几乎所有抽象代数的教科书上都使用它来描述群这种数学元素．

> Def 2.1 群
>
> 群是一个集合 G 和在它上面定义的二元运算 <10.> $\circ$（可称作群乘法）组成的二元组 $(G, \circ)$ 使满足以下公理：
>
> - **封闭性**（closure）：对于任意 $a, b \in G$，$a \circ b \in G$
> - **结合律**（associativity）：对于任意 $a, b, c \in G$，$(a \circ b) \circ c = a \circ (b \circ c)$
> - **单位元**（identity）：存在 $e \in G$，使对于任意 $a \in G$，有 $e \circ a = a \circ e = a$
> - **逆元**（inverse）：对于任意 $a \in G$，都存在一个 $b \in G$ 使得 $a \circ b = e$，它通常记作 a^{-1}

注意这里的群乘法" $\circ$ "并不是真正的乘法，而只表示一种**二元运算**（binary operation，$\circ : G \times G \to G$），它可以是数的加减乘除、集合交并差积、函数算子组合等等甚至是全新定义的运算，它并不一定满足交换律．尽管如此，我们在大多数情况下还是把它省略，将 $a \circ b$ 写作 ab，将 $\underbrace{a \circ a \circ \cdots \circ a}_{n}$ 记作 a^n，将 $\underbrace{a^{-1} \circ a^{-1} \circ \cdots \circ a^{-1}}_{n}$ 记作 a^{-n}．如果集合 G 上的运算已经在上文中规定，我们常说 G 是群．

上面的定义似乎有些抽象，我们先用自然语言复述一下．首先封闭性保证了无论哪些、多少个元素在这种运算下的结果都落在群内．结合律代表着运算的一种"稳定性" <11.>，它使得如果几个元素在一定顺序下得到 e，那么它们以同样的顺序连续出现在运算中时可以消去．单位元类似于加法中的 0 或乘法中的 1，它与其它任何元素运算都得到元素自身．而逆元的存在保证了群中每一个元素都可以在某种情况下消去．群乘法" $\circ$ "不一定满足交换律，如果它满足总是交换律则称这个群为 **Abel 群**（Abelian group） <12.>．

应该承认，这样的定义不是很 reasonable，我们很难从中看出这样做的 motif．但是对于数学中很多比较抽象的概念，我们不可能总是要求在读到它们时就完全理解．不如先记住这些定义，并用它们来检验一些例子和练习．在不断使用的过程中，你会发现你慢慢地理解了这些看似抽象的 definition．下面我们就来看一些常见的群的例子吧~

E.g. 2.1.1 整数集 $\mathbb{Z}$ 在普通加法运算下构成 Abel 群．单位元是 0，逆元是相反数．

E.g. 2.1.2 整数集 $\mathbb{Z}$ 在普通乘法下不构成群，因为存在元素没有逆元．

E.g. 2.1.3 所有非零有理数 $\mathbb{Q}^*$ 在普通乘法下构成 Abel 群，单位元是 1，逆元是倒数．

E.g. 2.1.4 实数集 $\mathbb{R}$ 和复数集 $\mathbb{C}$ 既构成加法群也构成乘法群，但是要注意乘法群不包含 0．

E.g. 2.1.5 所有 n-维向量在向量加法下构成 Abel 群.

E.g. 2.1.6 “计数与对称”一节中提到的六边形的 12 种对称操作在“相继操作”的二元运算下构成群，但不是 Abel群，称作 12 阶二面体群，记作 D_6 . 顺时针旋转 60° 再翻面相当于翻面再顺时针旋转 300° ，显然交换律不成立. 更一般地，所有正 n 面体的全部对称操作都可以各自成群，它们是 $2n$ 阶**二面体群**（dihedral group），记作 D_n .

E.g. 2.1.7 如果你学过线性代数，那么你会发现：所有 $n \times n$ 矩阵（实或复）在矩阵乘法下构成群，我们称之为**一般线性群**（general linear group），记作 $GL(n, \mathbb{F})$ （ $\mathbb{F}$ 是实数集或复数集等数域）；所有行列式为 1 的 $n \times n$ 矩阵（实或复）在矩阵乘法下构成群，我们称之为**特殊线性群**（special linear group），记作 $SL(n, \mathbb{F})$. 它们都是非 Abel 群.

读到这里，你应该已经对群有了一个基本的感受. 群有几条比较重要的基本性质，它们都直接从定义中得出.

> $(a^{-1})^{-1} = a$，即逆元关系是“互为逆元”.
>
> 单位元 e 在群中是唯一的.
>
> **消去律**（cancellation）：如果 $ab = ac$ 或 $ba = ca$ ，那么 $b = c$.
>
> 逆元 a^{-1} 对群元素 a 是唯一的.
>
> **鞋袜律**（socks-shoes property）：$(ab)^{-1} = b^{-1}a^{-1}$ ，你可以理解为先穿上袜子再穿鞋再戴鞋套，这一系列行为的反面是先脱掉鞋套再脱掉鞋再脱袜子~

下面证明前三条性质为例. 另外两条性质~~有大托梦给我~~请读者自行尝试证明.

证明 首先证明右消去律. 若 $ba = ca$ ，则 $baa^{-1} = caa^{-1}$ ，由结合律 $b(aa^{-1}) = c(aa^{-1})$ 即 $b = be = ce = c$. ■

注意到命题 $(a^{-1})^{-1} = a$ 等价于 $e = a^{-1}a$ ，已知 $aa^{-1} = e$ ，可得 $a^{-1}aa^{-1} = a^{-1}e = ea^{-1}$ ，由右消去律同时消去 a^{-1} 可得 $a^{-1}a = e$ as desired.

由于 $a^{-1}a = e$ ，与右消去律同理可证得左消去律.

若有两个单位元 e_1 和e_2 ，则 $a = ae_1 = ae_2$ ，由消去律 $e_1 = e_2$. $\mathcal{Q.E.D}$ ~

我们把群中元素的个数叫做群的**阶**（order），记作 $|G|$. 我们前面例子中的群，除去表征对称性的二面体群以外，其它群的阶都是 ∞ ，这样的群叫做**无限群**（infinite group），否则叫做**有限群**（finite group）. 那么是否还有其它有限群呢?

TRIVIALLY!

下面我们来看一看一种特殊但是重要的群——**循环群**（cyclic group）．

在上文模运算的学习中我们认识了一个叫做 $\mathbb{Z}_m$ 的数集，它表示所有小于 m 的自然数，也可以理解为关于 m 的模运算所有可能的结果．容易发现，在模加法下，这个集合满足封闭性的条件；我们可以验证 0 是它的单位元；其它每个元素的逆元都是与它相加等于 m 的那个数（取模后等于零）；显然满足结合律．因此它是一个群．

那么它是什么样的呢？我们让 1 反复和 1 运算如下（可以省略取模符号）：

$$\begin{aligned}
1 &= 1 \\
1+1 &= 2 \\
1+1+1 &= 3 \\
&\cdots \\
\underbrace{1+1+1+\cdots+1}_{m-1} &= m-1 \\
(m-1)+1 &= 0 \\
&\cdots
\end{aligned}$$

接着又从 $1+0=1$ 开始．我们发现，结果总是在 0 到 $m-1$ 之间循环．于是，我们把这种群叫做模-m 加法群（废话），又叫循环群，其中 1 是该群的一个**生成元**（generator）．我们不妨将这个群进一步抽象化：令它的生成元为 a，阶为 k，则我们令 $a^k=e$，循环群可表示为 $\{e, a, a^2, a^3, \ldots, a^{k-1}\}$，其中 a 是一个生成元，也不难发现，对于任何 s 只要 s 与 k 互质，那么 a^s 也是一个生成元．我们来证明一下：

证明　令 $a^{sr}=e=e^n=a^{kn}$，即 $sr=kn$，化为 $r=\frac{kn}{s}$．因为 k 与 s 互质，所以当且仅当 n 是 s 的整数倍时 r 取整，此时 r 是 k 的整数倍．因此，a^s 每自乘 k 次返回到 e，若自乘的次数小于 k 则不为 e，这意味着这 k 次自乘的结果没有重复．而群中恰好有 k 个元素，说明在 k 次自乘的过程中，a^s 的幂遍历了所有元素．这样，$\{e, a^s, (a^s)^2, (a^s)^3, \ldots, (a^s)^{k-1}\}$ 与原来的群只有元素顺序上的不同了．值得注意的是，这并不是一个语言上规范的证明，只是为了有让读者更清楚直观的体会，不过该证明逻辑上是正确完全的．你可以尝试用反证法等写一个规范的证明作为练习（别骂了别骂了~）．

在建立循环群的概念时，我们将数字抽象为了字母所代表的群元素，将模加法抽象为了群运算（群乘法），这种思想使得我们可以用群论的方法来研究数论问题，et vice versa~ 实际上，这里就用到了在代数中极其常见和重要的一种映射：**同构**（isomorphism）．下面先给出它的定义：

阅读提示：文中出现的术语及定理使用钢蓝色标出，并给出英文名称。色彩来自 Sully Prudhomme 的 sonnet，标题为 *Éther*：

> *« et l'azur luit sans tache,*
> *plus vif, comme l'acier qu'un souffle avait terni. »*

> Def 2.2 群同构
>
> 对群而言，同构是一个既是单射又是满射 <13.> 的映射 ϕ ，从一个群 G 到另一个群 $\overline{G}$ 使它保持群结构，即 $\phi(a)\phi(b)=\phi(ab)$ ．同构的逆映射也是一个同构，如果两个群之间存在同构映射，那么我们说这两个群是同构的（isomorphic），记作 $G\cong\overline{G}$ ．

直观来说，群同构就是两个群在代数研究的范围内完全相同 <14.> ，只是换了不同的名字和外观．它们有完全相同的代数性质，这意味着我们只要研究得到一个群的代数性质，就可以将其应用到所有与它同构的群中．

E.g. 2.2.1 实数加法群与正实数乘法群在 $\phi(x)=2^x$ 下同构．根据函数的单调性及值域可以验证它既是单射又是满射，又因为 $2^a2^b=2^{a+b}$ ，所以它满足同构的定义．容易找出它的逆映射为 $\phi^{-1}(x)=\log_2 x$ ，它也是一个同构．

E.g. 2.2.2 集合 $\{0,1,2\}$ 在模-3 运算下与集合 $\{0,n,2n\}$ 在模-$3n$ 运算下形成的群同构．这个例子很简单，请自己尝试运用定义验证一下．

既然有了同构，一个很自然的问题就是我们是否可以用一套统一的系统来表示所有的群．这种想法在后来发展成为了群表示论．这里我们不使用矩阵，也不研究无限群．我们将用置换来完成对有限群的表示．

置换（permutation）是集合 A 到它自身的既是单射又是满射的映射，~~这个定义就像没说一样~~．对于一个没有任何附加的其它数学结构的集合来说，它好像就是恒同映射．然而，对于集合中的每个元素来说，这却是有意义的：置换要么将它映射到它本身，要么映射到集合中的其它元素．这意味着对于任何一个由这个集合中元素生成的数学结构里面，这些元素都进行了一次重新排列．因此，我们通常这样来表示一个有限集的置换（这真的不是矩阵！XD）：

$$\begin{bmatrix} \alpha & \beta & \gamma & \cdots \\ \alpha_1 & \beta_1 & \gamma_1 & \cdots \end{bmatrix}$$

其中上下两行都是集合中的全部元素，只是以不同的顺序排列着．它表示这个置换把第一行的元素换成第二行正下面的元素．举个例子：

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 6 & 3 & 5 & 2 \end{bmatrix}$$

就表示把 2 换成 4 ，把 4 换成 3 ，把 3 换成 6 ，把 6 换成 2 ．其中 1 和 5 保持不变．

等等……看上面这句，怎么像转了一圈？没错，这就是转了一圈！由于 2 换成了 4 ， 4 就不能再出现，必须要换成别的元素……由于 2 必须出现一次，一定有一个元素换成了 2 ，这就闭环了．这种情况，我们通常写作 (2436) （当然你可以写成 (6243) 或者别的，只要圆排列不变就行），把这

种表示叫做一个**轮换**（cycle）．比如这个例子中就是一个 4-cycle．有时候，一个置换不止包含一个轮换，比如这个：

$$(2436)(15) = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 4 & 6 & 3 & 1 & 2 \end{bmatrix}$$

它就至少写成两个轮换的"乘积"．

还有一点我们可以发现：每个轮换都可以写成几个 2-cycle 的乘积；因为显然，每个置换都可以通过几次两两置换完成．比如上例中的 $(2436) = (26)(46)(36)$，先令 2 和 6 互换，再令 4 和 6 互换，最后令 3 和 6 互换．显然这只是其中一种分解方式，但是它总是有效的：令轮换中的每个元素依照轮换的顺序和其中一个元素互换，从该元素的下一个元素开始．想想为什么？

这种方式下，我们可以把所有置换都写成几个 2-cycle 的乘积，注意不一定满足交换律．事实上，当两个轮换中没有出现相同的元素时，它们满足交换律，比如上面的 $(2436)(15)$ 就等于 $(15)(2436)$，这很容易理解（两个轮换相互毫无关系）．

用 2-cycle 表示置换还有一些有趣的性质，比如恒同置换只能表示为偶数个 2-cycle 的乘积；一个置换若能表示为偶数个 2-cycle 的乘积则只能表示为偶数个 2-cycle 的乘积，若能表示为奇数个 2-cycle 的乘积则只能表示为奇数个 2-cycle 的乘积（英文口诀很好记，"Always even or always odd."）．我们把分解为偶数个 2-cycle 的置换叫做**偶置换**（even permutation），反之叫做奇置换．由于偶置换的乘积和逆映射也都是偶置换，容易验证，n 个元素间的全部偶置换在置换复合（相继作用）下构成群．我们将这个群叫做 **n 元交错群**（alternating group of degree n），记作 A_n．类似地，我们把 n 个元素间所有置换（奇或偶）构成的群叫做 **n 元对称群**（symmetric group of degree n），记作 S_n．

下面终于轮到做我们梦寐以求的事情了：用置换表示所有有限群．

> *Theorem* 2.1 **Cayley 定理**（Cayley's theorem）
>
> 任何有限群同构于一个以置换为元素的群．

这个定理看上去很简洁，但意义重大．它的证明用到了构造的方法，并不困难．对于一个群 G 中的任意元素 $g \in G$，定义 $T_g : x \mapsto gx$ 对于所有 $x \in G$．下证 T_g 是一个置换：

对于任意 $x_1, x_2 \in G$，若 $T_g(x_1) = T_g(x_2)$，即 $gx_1 = gx_2$，则由消去律 $x_1 = x_2$．又因为每个 $T_g(x_i)$ 都存在，所以 T_g 是双射（一一映射）．由群的封闭性，T_g 是从 G 到 G 的映射，即 G 的置换．

令 $\overline{G} = \{T_g \mid g \in G\}$，下证 $\overline{G}$ 是一个群：

对于任意 $g, h \in G$ ， $T_g(T_h(x)) = T_g(hx) = g(hx) = (gh)x = T_{gh}(x)$ ，由群 G 封闭性 $gh \in G$ 可知 $T_{gh} \in \overline{G}$ ， $\overline{G}$ 在置换复合下封闭．单位元为恒同置换 $T_e : x \mapsto x$ ． $T_g T_{g^{-1}} = T_{gg^{-1}} = T_e$ ，故有 $(T_g)^{-1} = T_{g^{-1}}$ ，逆元也是良定义的．置换复合本质上是函数复合，故满足结合律．

从 $T_g T_h = T_{gh}$ 中我们已经可以找出同构映射了．令 $\phi : g \mapsto T_g$ ，容易验证 ϕ 是同构．
Finishing the proof as desired !

这是一个漂亮的定理，意味着所有有限群，无论它的元素是什么样的——数字、点、集合，甚至人名和水果——都可以统一地用置换来代替！上述证明中的 ϕ 称为群 G 的一个**左正则表示**（left regular representation）．如果定义 $T_g : x \mapsto xg$ ，则得到的 ϕ 为**右正则表示**（right regular representation）．由于 ϕ 是一个同构，它们都是**忠实**（faithful）表示（了解即可）．

在这个定理的基础上发明了一种有用的工具：**Cayley 表**（Cayley table），又叫群乘法表 （group multiplication table）．具体的做法是：把群运算的符号写在左上角，群中所有元素从单位元开始横着向右列举一遍再竖着向下列举一遍，形成一个 $(|G|+1) \times (|G|+1)$ 二维表．在每个格上写出对应的行上左端元素与列上顶端元素的运算结果．如 4 阶循环群 $\mathbb{Z}_4^+$ （更多时候写作 C_4 ）的乘法表如下（表2-1）：

$+ \pmod 4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

table_2.1

它是沿对角线轴对称的，这说明它是一个 Abel 群．Cayley 表的每一行每一列都不重不漏地包含了所有群元素，可以看作是群上的一个置换．

群的结构——从子群到同态

在数学中我们经常遇到"子（sub-）"这个前缀，如**子集**（subset）、**子流形**（submanifold）、**拓扑子空间**（topological subspace）等．它通常与集合有关，包含了原对象中的一些元素，并继承了一些性质．研究这些子对象的性质对于研究原对象的性质非常有意义，因为它们通常是原对象中"有秩序且合理"的一块．这有点像解剖学的研究方法：去研究某个器官、组织，并阐明它们之间的联系，便能够还原出来整个机体的生理活动过程了．同样地，群也应当存在它的**子群**（subgroup）．下面我们先给出子群的定义：

> Def 3.1 子群
>
> 对于一个群 $(G,\circ)$ ，如果有 $H\subset G$ 使得 $(H,\circ)$ 也是一个群，那么称 H 是 G 的一个子群，记作 $H\leq G$.

这个定义不难理解，它是说，群中的一部分元素，在完全相同的运算下也构成了一个更小的群．显然它包含单位元，且像任何等价关系和其它序关系那样，具有传递性（我的子群的子群，还是我的子群）．下面通过一些例子来认识它：

E.g. 3.1.1 只含单位元的群 $\{e\}$ 和群本身 G 是 G 的子群，称作**平凡子群**（trivial subgroup）．

E.g. 3.1.2 若 s 是 m 的一个因数，那么 s 阶循环群 $\mathbb{Z}_s^+$ 同构于 m 阶循环群 $\mathbb{Z}_m^+$ 的一个子群．就像是从钟表（可看作 12 阶循环群 $\mathbb{Z}_{12}^+$ ）中挑出来 3 点、 6 点、 9 点和 12 点（它们四个成的群同构于 $\mathbb{Z}_4^+$ ）那样．

E.g. 3.1.3 $2n$ 阶二面体群 D_n 的一个子群是由单位元和反映操作（翻面）组成的 $\{E,F\}$ ，它同构于二阶循环群；另一个子群是由单位元和所有旋转组成的 $\{E,R,R^2,\ldots,R^{n-1}\}$ ，它同构于 n 阶循环群．

E.g. 3.1.4 上文中 Cayley 定理的另一种等价表述：任意 n 阶群同构于 S_n 的一个子群．

E.g. 3.1.5 在普通加法运算下，有 $\mathbb{Z}\leq\mathbb{Q}\leq\mathbb{R}\leq\mathbb{C}$ ；在普通乘法运算下， $\mathbb{Q}^*\leq\mathbb{R}^*\leq\mathbb{C}^*$ ．

E.g. 3.1.6 当维度 n 和 数域 $\mathbb{F}$ 都相同时，特殊线性群是一般线性群的子群；旋转群（顾名思义）是特殊线性群的子群．

E.g. 3.1.7 若 $H\leq G$ ，对于任意 $g\in G$ ， $gHg^{-1}:=\{ghg^{-1}\mid h\in G\}$ 是 G 的另一个子群，叫做 H 的一个**共轭子群**（conjugate subgroup）． H 的所有共轭子群（包括它本身）之间互为共轭，构成一个**共轭类**（conjugacy class）．

E.g. 3.1.8 所有三维实向量（欧氏空间 $\mathbb{R}^3$ ）在线性加法下构成群，所有过原点的直线和平面都是它的子群．

要判断一个群的子集是它的子群，我们不需要一一代入它的全部四条定义，只需要满足以下两个条件即可判断为真：

- 该子集在群 G 的运算下封闭；
- 子集中每个元素都有逆元在子集中．

其它条件都可以由这两条推出.

如果我们总是在思考各种群的子群，也许还会发现一个有趣的现象．对于任意 $a \in G$ ，都有 $a^2, a^3, \cdots \in G$ （封闭性），因此所有由 a 多次自乘得来的元素构成一个循环群，它是 G 的子群．显然，存在一个整数 $n \leq G$ 使得 $a^n = e$ ， n 的最小取值即为这个子群的阶数．这个子群我们通常记作 $\langle a \rangle$ ，其中元素个数（子群的阶）也称为元素 a 在 G 中的阶，记为 $|a|$ ．由于任意元素都可以这样生成循环群，所以有限群在某种意义上都可以看作是由几个循环群以一定方式"粘合"起来的.

接下来我们研究子群与原群 <15.> 的关联．我们首先定义**陪集**（coset）：

> Def 3.2 陪集
>
> 对于任意 $a \in G$ ， $H \leq G$ ．令 $aH := \{ah \mid h \in G\}$ ，叫做子群 H 在群 G 中的左陪集；令 $Ha := \{ha \mid h \in G\}$ ，叫做子群 H 在群 G 中的右陪集．其中元素个数记作 $|aH|$ 和 $|Ha|$ ．

可以发现，它用子群中的元素逐个与 a 运算来建立起一个子群和群中任意一个元素的联系，进而建立起它与整个群的联系．如果 $a \in H$ ，显然任何 ah 和 ha 都是 H 的元素，并且能遍历到 H 的每一个元素，使 $aH = Ha = H$ ．而对于 $a \notin H$ ，它产生的陪集又与子群是什么关系呢？下面我们的讨论都以左陪集为例，得出的这些结论在右陪集中也是适用的.

显然对于 $a \notin H$ 且 $a \in aH$ ，所以 aH 与 Ha 不可能完全重合．那么它们是否有相交的部分呢？假设存在 $h_1 \in H \cap aH$ ，那么必然存在 h_2 使得 $h_1 = ah_2 \in H$ ，又由群中每个元素有逆元，有 $ah_2h_2^{-1} = a \in H$ ，与假设不符.

这就是说，陪集与子群要么完全重合，要么交集为空！扩展到两个陪集之间：$aH = Ha$ 或 $aH \cap bH = \varnothing$ ．证明如下：我们将它分成三步来证.

i.**证明：** 对于任意 $a \in G$ ，都有 $|aH| = |H|$.

对于 $h_1, h_2 \in H$ ，若 $h_1 = h_2$ ，则 $ah_1 = ah_2$ ． $\Longrightarrow$ 若 $ah_1 \neq ah_2$ ，则 $h_1 \neq h_2$ ；若 $ah_1 = ah_2$ ，则 $h_1 = h_2$ ． $\Longrightarrow$ 若 $h_1 \neq h_2$ ，则 $ah_1 \neq ah_2$ ．故存在双射 $\phi : H \to aH$ ，即 $|aH| = |H|$ ，证明完毕.

ii.**证明：** $a \in bH \implies aH = bH$.

若 $a \in bH$ ，则可表示为 $a = bh_1$ 其中 $h_1 \in H$ ．因此对于任意的 $h_2 \in H$ ，都有 $ah_2 = bh_1h_2 \in bH$ ，故 $aH \in bH$ ．又有 $|aH| = |bH| = |H|$ ，所以 $aH = bH$ ，证明完毕.

iii.**证明原命题：** 设存在 $c \in aH \cap bH$ ，则 $cH = aH = bH$. ■ .

其实，证明这个命题并不必须分成这三步，也可以更简单一些．但一来这样每一步都很清晰，二来这几个子命题恰好说明了陪集的重要意义：通过子群和它的陪集，可以将一个群不重（如上）不漏（注意到每个元素都可以生成陪集，如果它不属于已知陪集）地划分为大小相等的几块——陪集可以看作是子群的平移．就像这幅图所展示的：

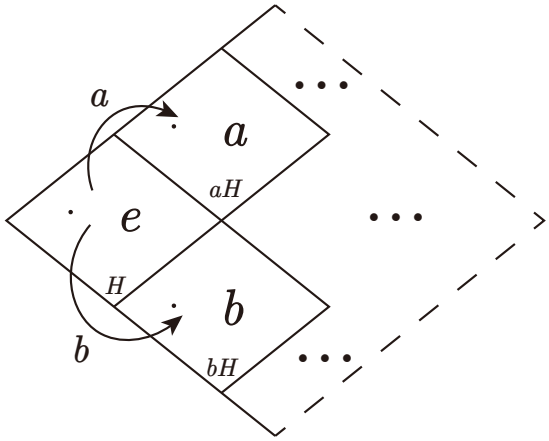


fig_3.1

从几个例子我们可以看得更清楚：

E.g. 3.2.1 对于线性空间 $\mathbb{R}^3$ 在向量加法下构成的群的一个同胚于 $\mathbb{R}^2$ 子群（即过原点的一个向量平面），所有与它平行但不重合的平面都是它的陪集．对于它的一个同胚于 $\mathbb{R}$ 的子群（即过原点的一条向量直线），所有与它平行但不重合的直线都是它的陪集．

E.g. 3.2.2 十二阶二面体群 D_6 的一个子群是单位元和所有单纯的旋转（同构于 $\mathbb{Z}_6^+$ ），它的陪集只有两个，即它本身和所有包含翻面的操作；另一个子群是单位元和翻面，它有六个陪集，分别是它本身和 table_1.1 中它右边的五列．

于是，就有了陪集形式表述的 **Lagrange 定理**（Lagrange's theorem）：

> *Theorem* 3.1 Lagrange 定理
>
> 对于一个有限群 G 和它的子群 H ，总有 $|H|$ 整除 $|G|$.

现在来看，这个定理几乎是显然的．尽管当时并没有群、子群、陪集这些概念，Lagrange 还是以他自己的形式发现了这个定理．从中可以轻松得出一些小结论：

- $|G|$ 整除 $|a|$ ，即元素的阶是群的阶的因数．

证明　注意到 $\langle a\rangle \leq G$.

- $a^{|G|}=e$.

证明　由 $|a|$ 定义得 $a^{|a|}=e$ ，又有 $|G|=k|a|\,(k\in\mathbb{N}^*)$ ，因此 $a^{|G|}=a^{k|a|}=e^k=e$.

以及下面这个在九省联考中出现过的、在竞赛中相当常见、在实际应用中也颇具价值的 Fermat **小定理**（Fermat's little theorem）：

> *Corollary* 3.1.1　Fermat 小定理
>
> 若 p 为素数，对于任意正整数 a 有 $a^p\equiv a\pmod p$.

证明　构造集合 $\mathbb{Z}_p^*=\{1,2,3,\ldots,p-1\}$ ，该集合在模-p 乘法下封闭，单位元为 1 ，取模前即满足乘法结合律．当且仅当 p 为素数，每个元素都存在逆元，下证逆元的唯一存在性：

设 $a\in\mathbb{Z}_p^*$ 的逆元为 x ，则应满足 $ax\equiv 1\pmod p$ ，这是一个同余方程，等价于 $ax=np+1$ ，根据 Bézout 定理 <16.> ，它一定有解．我们要找出它的全部解．下面的字母全部默认为正整数．将 a 移到右边，即 $x=\frac{np+1}{a}$ ，设一个解为 r ，另一个解为 $r\pm k=\frac{ra}{a}\pm k=\frac{ra\pm ka}{a}$ ，要使分子满足 $np+1$ 的形式，应有 $ka=np$ 使 n 为正整数，$n=\frac{ka}{p}$ ，又因为 p 是质数，所以 k 的解为 $k=sp,s\in\mathbb{N}^*$ ．这意味着原同余方程每相邻 p 有且仅有一解，则在 $\{1,2,\ldots,p\}$ 上有且仅有一解．若解为 p ，代入得 $ap=np+1$，显然对于素数 p 不成立．故原同余方程在 $\{1,2,\ldots,p-1\}$上有且仅有一解．逆元的唯一存在性证毕．（练习：找出一种更简洁的逆元存在性证明．）

则 $\mathbb{Z}_p^*$ 是在模-p 乘法下是群．对于 $0\leq a<p$ ，由于 a 在群中，显然 $a^{p-1}=a^{|\mathbb{Z}_p^*|}=1$ ，即 $a^p\equiv a\pmod p$ ．对于 $a\geq p$ ，令 $a=np+r$ 使 $r\in\mathbb{Z}_p^*$，则 $a^p=(np+r)^p$ ，根据高中所学的二项式定理相关知识，多项式展开后唯一不含因数 p 的项为 r^p ，则 $a^p\equiv r^p\equiv r\pmod p$ ．又因为 $a=np+r\equiv r\pmod p$ ，故 $a^p\equiv a\pmod p$ 成立．原命题得证．

需要注意的是，Lagrange 定理的逆命题并不成立．$|G|$ 的每一个因数并不一定都对应着相应阶数的子群．一个典型的例子是四元交错群 A_4 ，尽管 $|A_4|=\frac{4!}{2}=12$ ，但是它并没有 6 阶子群．另一种判断某个阶数子群存在性的方法是著名的 **Sylow 定理**（Sylow's theorems），但相关的内容较多（其实是因为我跳过了轨道和稳定化子すいません~），这里不做介绍，感兴趣的请自己查阅 textbook 或相关讲义和笔记．

根据 Lagrange 定理，$\frac{|G|}{|H|}$ 是一个整数，我们自然也会想到，它是否也对应着一个子群的阶数．答案是肯定的．从这一想法，我们定义了**商群**（quotient group），又叫**因子群**（factor group）．我并不想像抽象代数的教科书那样直接给出概念；我希望能从头开始一步步地建立它．我们知道 $\frac{|G|}{|H|}$

是 H 陪集的个数，一个显然的想法是将每个陪集作为这个新群的一个元素．为了让它的结构与 G 相似，必须有 $\phi: aH \times bH \mapsto (ab)H$ ，即将这个群保持 G 的运算．自然地，它就有了单位元 $eH = H$ ，以及各个元素的逆元（自分で確認してください）．我们也许会想到使用一个同构 $a \mapsto aH$ ，但是，它并不完备：我们发现对于一个确定的陪集 aH ，其中任何一个元素都可以是 a ，用来表示这个陪集．用数学语言表达就是：$\forall h \in H, aH = (ah)H$ ．假如我们仅仅用 $a \times b \mapsto ab$ 来表示这个群的运算，那么由于任意 ah 也可以用来代替 a ，就等于承认了 $ahb = ab$ ，经过消去律就成了 $h = e$ ，这显然不能符合 H 是非平凡子群的情况．所以我们不得不放弃用某个元素来代替一个陪集的想法．为了让 aH 中的某个元素和 bH 中的每个元素的运算结果都落在 $(ab)H$ 中，需要满足以下陈述：

$$\forall h_1, h_2 \in H, \exists h_3 \in H, ah_1bh_2 = abh_3 \tag{cl3.1}$$

这难住了我们：我们不会处理这样的结构．我们会想：如果这个群是 Abel 群就太好了，因为 $ah_1bh_2 = abh_1h_2 \in (ab)H$ ．但并不是每个子群都是 Abel 群．为了最大限度地适用于更多子群，同时严格地满足上述命题，我们会想到让任意 $hb = bh'$ 即可（注意 h' 不一定等于 h ），这样 $ah_1bh_2 = abh_1'h_2 \in (ab)H$ ．由于要求对所有 $h \in H$ 成立，这等价于 $Hb = bH$ ，即子群 H 关于 b 的左陪集与右陪集相同．为了让任意 $b \in G$ 都满足上述 $(\mathrm{cl}3.1)$ ，要求子群 H 的所有左陪集和右陪集都相同．这就是最早由 Galois 提出的**正规性**（normality）的概念．下面我们给出**正规子群**（normal subgroup）的定义：

> Def 3.3 正规子群
>
> 对于一个 $N \leq G$ ，若对于任意 $a \in G$ ，都有 $aH = Ha$ ，则称 N 是 G 的一个正规子群，记作 $N \lhd G$ ．

不要误以为 H 中的元素都与其它元素可交换，而是 $ah_1 = h_2a$ ，只要 h_1 与 h_2 都在 H 中即可，它们可以相等也可以不等．这是一个比交换律要弱的性质，因此更容易满足，但很多时候"它就像交换律一样好"．

通常我们验证正规子群的方法是验证它的任意一个共轭子群都是它的子集（利用定义就可以发现这与定义是等价的），即：

$$N \lhd G \iff \forall g \in G, x \in N, gxg^{-1} \in N$$

容易验证所有共轭子群的元素个数是一样多的，因此，共轭子群是子群的子集就等于是说这个子群的所有共轭子群都是它本身．此外，对于一些特殊的情况，我们还可以更容易地判断一个子群是否正规：

- 如果相应有限阶数的子群有且仅有一个，那么它是正规子群．

证明　注意到它的共轭子群与它的阶数相等，这意味着它的所有共轭子群都是它本身．

- 如果 $\frac{|G|}{|N|}=2$ ，那么 $N \lhd G$ ．

证明　注意到这个子群的非平凡左陪集和右陪集都是 G 中去掉 N 以后所有元素的集合．

简单地举几个例子吧．

E.g.　3.3.1　平凡子群是正规的．

E.g.　3.3.2　如果子群是 Abel 群，那么它是正规的．我们定义群的**中心**（center，或者德语 Zentrum）为 $Z(G):=\{a \in G \mid \forall g \in G,\, ag=ga\}$ ，它是群中最大的 Abel 子群，包含了所有与其它元素交换的元素．

E.g.　3.3.3　当 n 与 $\mathbb{F}$ 一致时，特殊线性群 $SL(n,\mathbb{F}) \lhd GL(n,\mathbb{F})$ ，这可以直接由行列式的线性性得出（ $\det(A)\det(B)=\det(AB)$ ）．

基于此，我们便可以给出 well defined 的商群定义了．

> $Def\ \ 3.4$　商群（因子群）
>
> 对于一个群 G 和它的正规子群 $N \lhd G$ ，记 $G/N:=\{aH \mid a \in G\}$ ，即 H 所有左陪集（或右陪集）的集族．它是一个群，称为 G 关于 H 的商群（因子群），其群乘法为 $aHbH=(ab)H$ ．

商群也有一些基本性质，比如循环群的商群都是循环群，Abel 群的商群都是 Abel 群等．其中有一个比较有趣，你可以尝试证明一下：如果 $G/Z(G)$ 是循环群，那么 G 是 Abel 群．

我们观察一个群和它的商群，它们之间保持了运算，但是其中元素的对应关系并不像同构那样严格．这种关系比同构要弱许多，我们称之为**同态**（homomorphism）．

> $Def\ \ 3.5$　群同态
>
> 对群而言，同态是一个映射 τ ，从一个群 G_1 到另一个群 G_2 使它保持群运算，即 $\tau(a)\tau(b)=\tau(ab)$ ．

对比同构的概念，我们发现：首先它既不要求单射，即可以把好几个元素映射到一个；也不要求满射，也就是说可以只把 G_1 映射到 G_2 的一部分上，或者说 G_1 的同态像 $\tau(G_1)$ （也可以记作 $\operatorname{Im}\tau$ ）只要是 G_2 的子集即可．如果一个同态是单射，我们就把它叫做单同态；如果一个同态是

满射，我们就把它叫做满同态；如果一个同态既是单射又是满射，那么它是一个同构．我们注意到一个群到它的商群存在一个满同态 $\tau_1 : a \mapsto aH$ ，这个同态看起来如此自然且典范~（其实并不只是因为“看起来”，更深层次的原因请自行探索）以至于我们把它叫做**自然同态**（natural homomorphism）或**典范同态**（canonical homomorphism）．

在一个同态中，有一些元素被映射为 G_2 中的单位元 e' ，这些元素的集合叫做同态 τ 的**核**（kernel），记作 $\operatorname{Ker}\tau := \{g \mid g \in G_1, \tau(g) = e'\}$ ．单同态的核是 G_1 中的单位元．下面这幅图形象地展示了同态、满同态、单同态和同构．

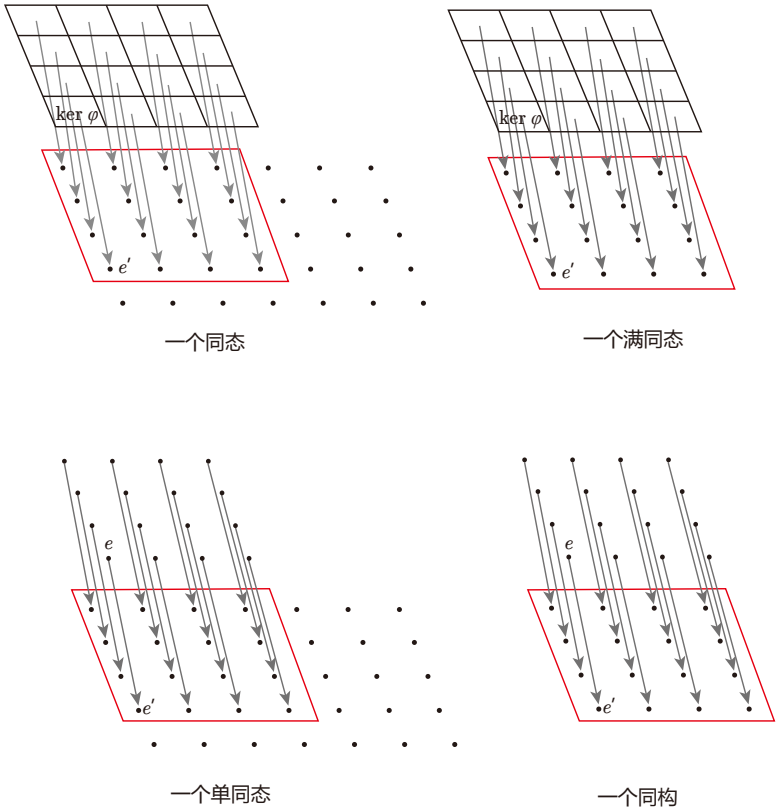


fig_3.2

同态只能说明两个群在结构上有一定的相似之处，它们相互只能部分地反映对方的特征．但是，由于容易满足，同态的应用十分广泛，尤其是在群表示论中．下面有一个我个人认为非常有趣的例子：

E.g. 3.5.1 所有实系数多项式的集合 <17.> 记作 $\mathbb{R}[x]$ ，它在加法运算下是一个 Abel 群．则求导是它到自身的一个同态，其同态核为整数加法群 $\mathbb{Z}$ ．

既然群到商群的映射是同态，那么反过来同态是否都具有类似于群到商群映射的结构呢？这个问题最初由 Jordan 研究，后来 Emmy Noether <18.> 在她的论文 *Abstrakter Aufbau der Idealtheorie in algebraischen Zahl- und Funktionenkörpern* 中第一次给出了广泛性的（适用于多种代数结构）答

案：是的．在这篇论文中，Noether 总结出了一系列有关同态中的同构的定理，被后人称为 Noether's isomorphism theorems，它们成为了现代代数学的最重要基础之一．我们介绍其中最基础的**群同态基本定理**（fundamental homomorphism theorem，FHT），又叫做**群的第一同构定理**（the first isomorphism theorem for groups）．

> *Theorem* 3.2　群同态基本定理
>
> 若 $\phi: G \to G'$ 是一个同态，则有 $\mathrm{Ker}\ \phi \lhd G$ 且 $G/\mathrm{Ker}\ \phi \cong \phi(G)$ ．

证明　记 G' 中单位元为 e' ，首先证明 $\mathrm{Ker}\ \phi \lhd G$ ．先证 $\mathrm{Ker}\ \phi \leq G$ 如下：

显然 $\mathrm{Ker}\ \phi \subseteq G$ ．由上文子群的判定条件需证封闭性及逆元存在．对于任意 $e_1, e_2 \in \mathrm{Ker}\ \phi$ ，有 $\phi(e_1e_2) = \phi(e_1)\phi(e_2) = e'$ ，即 $e_1e_2 \in \mathrm{Ker}\ \phi$ （**封闭性**）．对于 G 的单位元 e ，对任意 $a \in G$ ，有 $\phi(a) = \phi(ae) = \phi(a)\phi(e)$ ，由消去律 $\phi(e) = e'$ ，可得 $e \in \mathrm{Ker}\ \phi$ ．对于任意 $e_1 \in \mathrm{Ker}\ \phi$ ，总有 $\phi(e_1)\phi(e_1^{-1}) = \phi(e_1e_1^{-1}) = \phi(e)$ ，即 $\phi(e_1^{-1}) = e'$ ，可得 $e_1^{-1} \in \mathrm{Ker}\ \phi$ （**逆元**）．

综上所述，有 $\mathrm{Ker}\ \phi \leq G$ 成立．对任意 $e_1 \in \mathrm{Ker}\ \phi$ 且 $a \in G$ ，总有 $\phi(ae_1a^{-1}) = \phi(a)\phi(e_1)\phi(a^{-1}) = \phi(aa^{-1}) = \phi(e) = e'$ ，即 $ae_1a^{-1} \in \mathrm{Ker}\ \phi$ ．由上文正规子群的判断条件可得 $\mathrm{Ker}\ \phi \lhd G$ ．

写出 $G/\mathrm{Ker}\ \phi = \{a\mathrm{Ker}\ \phi \mid a \in G\}$ ， $\phi(G) = \{\phi(a) \mid a \in G\}$ ．构造映射 $\sigma: G/\mathrm{Ker}\ \phi \to \phi(G), a\mathrm{Ker}\ \phi \mapsto \phi(a)$ ．但是，它其实并不一定是一个映射：因为陪集的 a 可以是陪集中的任意元素，它们的同态像相同并不是显然的，这里必须要验证 σ 是良定义的映射，即将每个陪集只映射到 $\phi(G)$ 中的唯一确定元素：

若 $a\mathrm{Ker}\ \phi = b\mathrm{Ker}\ \phi$ ，有 $b \in a\mathrm{Ker}\ \phi$ ，可以将 b 唯一地表示为 ae_3 其中 $e_3 \in \mathrm{Ker}\ \phi$ ．那么一定有 $a^{-1}b = e_3 \in \mathrm{Ker}\ \phi$ ，即 $\phi(a^{-1}b) = \phi(a^{-1})\phi(b) = e'$ ．要处理这个结构，我们注意到 $\phi(a^{-1})\phi(a) = \phi(aa^{-1}) = e'$ ，则 $\forall a \in G$ 有 $\phi(a^{-1}) = \phi(a)^{-1}$ ．代入得 $\phi(a^{-1})\phi(b) = \phi(a)^{-1}\phi(b) = e' = \phi(a)^{-1}\phi(a)$ ，由消去律得 $\phi(a) = \phi(b)$ ．这就证明了 σ 是一个映射：对于每个确定的陪集，它只将它映射到唯一确定的 $\phi(a)$ ．

然后我们就开始证明它是同构：首先它是个同态，因为 $\sigma(a\mathrm{Ker}\ \phi\ b\mathrm{Ker}\ \phi) = \sigma(ab\mathrm{Ker}\ \phi) = \phi(ab) = \phi(a)\phi(b) = \sigma(a\mathrm{Ker}\ \phi)\sigma(b\mathrm{Ker}\ \phi)$ ；其次它是个单射，假设存在 $\sigma(a\mathrm{Ker}\ \phi) = \sigma(b\mathrm{Ker}\ \phi) = \phi(a)$ ，那么 $\phi(a) = \phi(b)$ ，两边一起左乘 $\phi(a^{-1})$ ，得到 $\phi(a^{-1}b) = e'$ ，即 $a^{-1}b \in \mathrm{Ker}\ \phi$ ．记 $b = ag$ 则 $a^{-1}ag = g \in \mathrm{Ker}\ \phi$ ，即 $b \in a\mathrm{Ker}\ \phi$ ，也就是说 $a\mathrm{Ker}\ \phi = b\mathrm{Ker}\ \phi$ ．最后证明它是一个满射，对于每个 $\phi(a)$ ，它都由某个 a 映射得到，它一定属于一个陪集 $a\mathrm{Ker}\ \phi$ ，因此它是满射．综上所述，它是一个同构．

至此，原命题就证明完毕了（这可能就是某些教材省略的“繁而不难”的证明吧#$...%@~）

如果你学过线性代数（再次），你可能会想到**秩-零化度定理**（rank nullity theorem），即线性变换的秩和零化度（零空间的维度）的和总是等于原像的维度，看起来非常直观．它就是同态基本定理在线性空间的表现形式．

至此，我们的文章主体部分就要告一段落了．我们从简单的模运算和对称性开始，构建起了一系列概念，最终得到了如此美妙的结果．这可能就是数学的 charme 所在吧．

探索仍在继续...

如果你完完整整地读到了这里，并且理解了上面所说的全部，那么恭喜你，你对代数的感觉真的很好；即使你未能完全理解，能读下来也颇为不易：如果感兴趣不妨再多读几遍，花些时间多想一想（作者学会这些内容也差不多花了整一个月）．事实上，群论只是抽象代数的一个相对不那么现代的小小分支——离入门抽象代数这门学科，你还差得很远．但是，在读完上面所有内容以后，你已经有了阅读一些代数书籍和论文的能力，并掌握了一些常用的代数研究方法．如果你真的想系统性地学习更多群论和抽象代数，这里推荐几本我用过的教材：

入门级别（前两本书差不多，挑一个）：

- Dummit 的 *Abstract Algebra, 3rd.*（那个合著者经常被忘记bs，还有据说出第四版了）
- Gallian 的 *Contemporary Abstract Algebra, 10th.*（有十分详细的前置知识，甚至不要求之前认识复数和最大公约数）
- Nathan Carter 的 *Visual Group Theory*，好看，但是内容只包括群论，可以作为早期学习阶段的辅助用书．有在线的网页可交互学习资源，好像需要梯子．

物理方向（通俗易懂）：

- 李新征老师的《*〈群论一〉讲义*》，是北京大学物理学院内部资料（官网上可直接下载，相当好的简体中文讲义！）．

代数领域综合，有一定深度：

- GTM73 和 GTM211，别的也有评价不错的我没看过，这两本我也没看完，但是经典作品确实是漂亮，从群环域模到表示论一气呵成．如果不知道 GTM 是什么就可以不用考虑了，先入门再说．

其实我再这方面看过的资料也不是很多，大家可以自己探索（或者说，一定要自己探索）．另外，代数是一个非常吃概念的学科，一定要非常重视例子和练习，运用你学到的概念去检验它们，其实也就并没有那么“抽象”了．

法国的布尔巴基学派从公理化集合论出发，把数学研究的对象分为代数结构、序结构和拓扑结构（现在看来并不完备）；抽象代数这门学科研究的核心对象就是代数结构．其实群只是其中最典

型、也相对简单和丰富的一种代数结构，这幅图展示了各种常见的代数结构的关系（图4-1）：

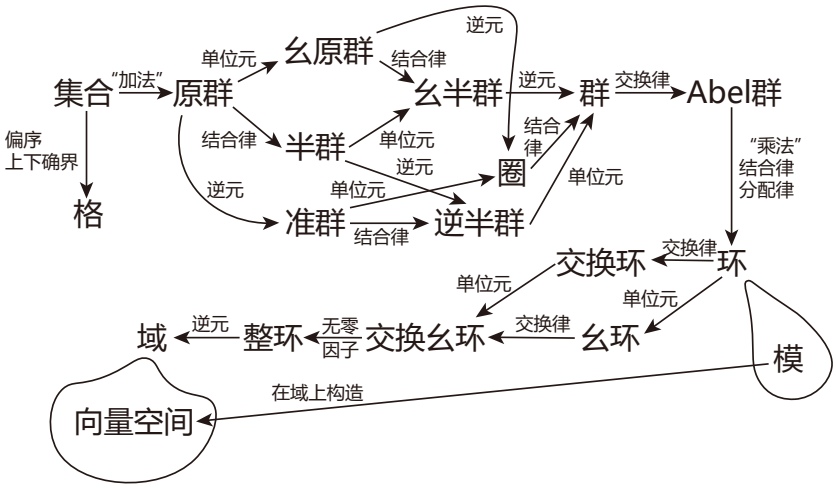


fig_4.1

前面讲历史的时候提到过 **Lie 群**（Lie group），它把数学里面最漂亮的两个东西（别跟我说 p 进几何里钻石、棱镜那一堆抽象的玩意）——群和流形联系在了一起，成为微分几何和代数拓扑的重要研究对象．如果你知道**流形**（manifold）是什么（形象来说其实就是嵌入高维空间里面的一个局部同胚于较低维的欧氏空间的拓扑空间），那么可以来看看这个大家都耳熟能详的 Lie 群的定义：

> Def 4.1 Lie 群
>
> Lie 群既是流形也是群，满足以下两个条件：
>
> 1. 群乘法是光滑的（无穷阶可导 C^∞）；
> 2. 求逆元的映射 $x \mapsto x^{-1}$ 是光滑的（无穷阶可导 C^∞）．

容易验证，上文说到过的一般线性群和特殊线性群在数域为 $\mathbb{R}$ 或 $\mathbb{C}$ 时都是 Lie 群．实际上所有表示 n-维旋转的矩阵组成的群——**正交群**（orthogonal group） $SO(n)$ 也是一类常见的 Lie 群，此外还有**辛群**（symplectic group）和**酉群**（unitary group），它们分别是辛矩阵和酉矩阵组成的群．

我们前面给出过商群的定义，很自然地会想到，既然有商也应该有一种乘法：事实上我的很多代数书里都是先介绍的**外直积**（external direct product），再介绍的正规性和商群．两个群的外直积类似于集合的笛卡尔积，定义如下：

$$G_1 \oplus G_2 := \{(g_1, g_2) \mid \forall g_1 \in G_1,\ g_2 \in G_2\}$$

它是一个阶数为 $|G_1||G_2|$ 的群，群乘法为 $(g_1, g_2)(g_1' g_2') = (g_1 g_1', g_2 g_2')$ ．它就像把两个群"粘合"起来了．相似地，还有**内直积**（internal direct product），它定义在 G 的两个交集为 $\{e\}$ 的正规子群 N 和 H 间：

$$N \times H = \{nh \mid \forall n \in N,\ h \in H\}$$

它与 N 和 H 的外直积是同构的．在外直积的基础上，有著名的**有限 Able 群基本定理**（fundamental theorem of finite Abelian groups）．它表述了所有有限 Abel 群的分类和结构：

> *Theorem* 4.1　有限 Abel 群分类定理
>
> 每个有限Abel群都可以被唯一地分解为有限个素数幂阶循环群的外直积．即Abel群 $G \cong \bigoplus_{i=1}^{n} \mathbb{Z}_{p_i^{m_i}}^{+}$ ，其中 p_i 是素数， m_i 是整数．

这种分类思想在数学的发展中产生了很大的影响，后来，数学家不仅创立了**范畴论**（category theory），而且创建了**有限单群周期表**（the periodic table of finite simple groups）（表4-1）：

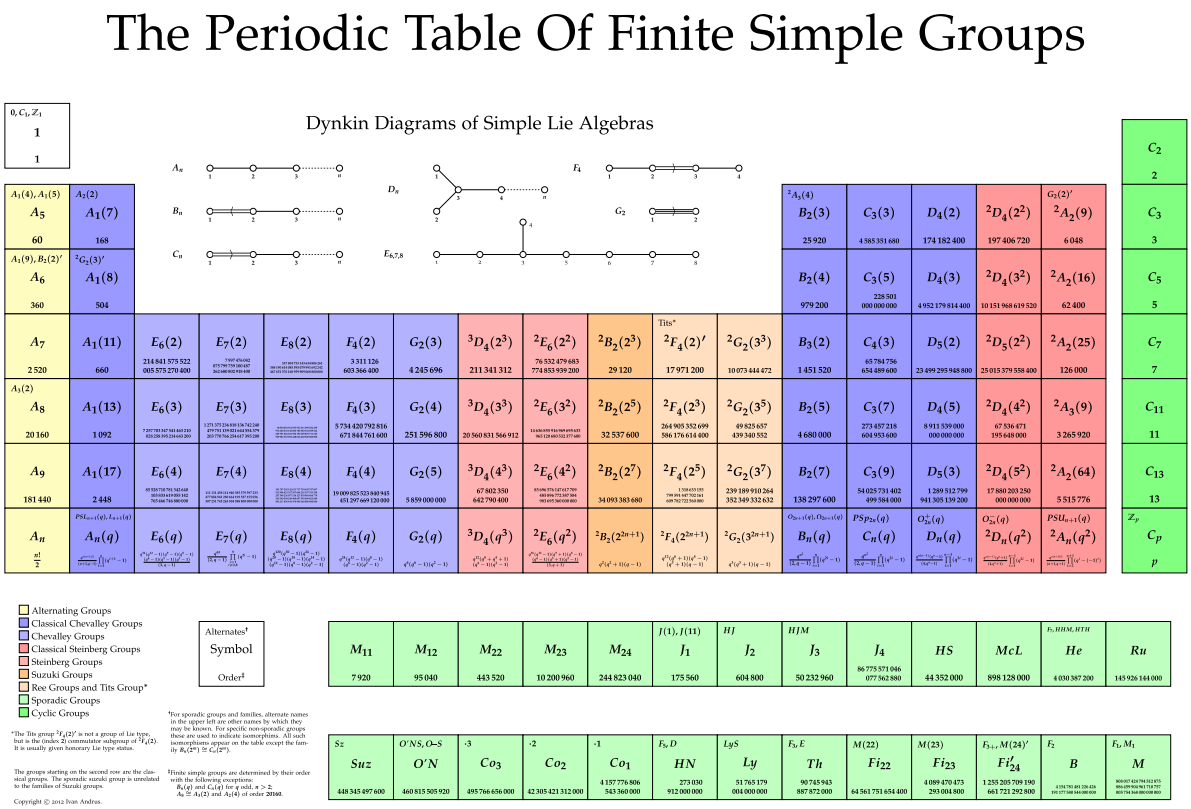


table_4.1

打印出来你当然看不清，里面的字太小了，比如 Lie 群 $E_8(4)$ 的阶数为:

$$
\begin{array}{l}
191\ 797\ 292\ 142\ 671\ 717\ 754\ 639\ 757\ 897 \\
512\ 906\ 421\ 357\ 507\ 604\ 216\ 557\ 533\ 558 \\
287\ 598\ 236\ 977\ 154\ 127\ 870\ 984\ 484\ 770 \\
345\ 340\ 348\ 298\ 409\ 697\ 395\ 609\ 822\ 849 \\
492\ 217\ 656\ 441\ 474\ 908\ 160\ 000\ 000\ 000
\end{array}
$$

它的结构图由18个数学家共同绘制，其计算结果，包括所有的信息及表示，其总容量达到了60GB，可以铺满整个曼哈顿岛．不过上面的周期表是开源的，你可以在网上很容易地找到 `.pdf` 格式的文件放大来看．

所谓**单群**（simple group），就是没有非平凡正规子群的群，也即它的正规子群要么是 $\{e\}$ 要么是它本身．下面两行是无法归类进几大类的一些特殊的单群，称作**散在单群**（sporadic group），其中右下角的 M 是最大的一个，被命名为**魔群**（monster group）．这个群线性既约表示需要的维数从小到大分别是 $1, 196883, 21296876, \ldots$，其前 n 项和恰好等于代数数论模形式中的 j-不变量（一个著名的模函数）的 Fourier 展开形式中规范化后（去掉常数项）第 n 项的系数．基于此，普林斯顿大学的 Conway 和剑桥大学的 Norton 提出了**魔群月光猜想**（Monstrous Moonshine）．该猜想认为可以由魔群构建一个无限维代数结构，通过魔群的不可约线性表示，给出 j-不变量的每个 Fourier 系数，而魔群每个元素在这个代数结构上的作用，都自然给出了与某个群相关的模形式．之所以叫"月光"，是因为这在当时看来实在太不可思议．其内容摘录如下：

> **Conway-Norton's Monstrous Moonshine Conjecture（1979）**: There is a natural infinite dimensional graded representation $V = \bigoplus_{n=0}^{\infty} V_n$ of $\mathbb{M}$, such that for any $g \in \mathbb{M}$, the series $\sum \mathrm{Tr}(g|V_n)q^{n-1}$ is the q-expansion of a modular function $T_g(\tau)$ on the complex upper half-plane $\mathfrak{H}$ that satisfies the following properties:
>
> (1) $T_g(\tau)$ is invariant under a discrete subgroup Γ_g of $SL_2(\mathbb{R})$ that contains $\Gamma_0(12|g|)$.
> (2) $T_g(\tau)$ generates the field of meromorphic functions on the smooth compactification of the quotient Riemann surface $\mathfrak{H}/\Gamma_g$, i.e., $T_g(\tau)$ is a Hauptmodul for Γ_g. In particular, $\mathfrak{H}/\Gamma_g$ is a genus zero complex curve with finitely many punctures.

别说你看不懂，我学完基本的抽代也还是看不懂．这里涉及的知识太深，还有待我们慢慢探索．该问题最终由加州伯克利大学的 Borcherds 借用物理学里面的共形场论和弦论中的方法解决．

代数可能是你面前几道可怕的 IMO 不等式真题；但是任何一门数学学科远不仅仅是技巧的堆砌，真正的美感来源于从看似简单的定义中得到如此丰富的结果；它挑战我们逻辑思维和形式化语言的边界．总之，Algebra 是我认为最美的学科，仅展示如上．

Appendix-i-注释

<1.>两个 $\mathbb{N}$ 之间的乘号是**笛卡尔积**（Cartesian product）又称**直积**（direct product），它生成全部有序元素组的集合．定义为：

$$X \times Y := \{(x, y) \mid x \in X, y \in Y\}$$

<2.>有人认为更好的翻译也许是"比例数"或"可比数"，此处存在争议．

<3.>更详细地，能通过解有理式代数方程得到的数叫做**代数数**（algebraic number），表示为 $\mathbb{A}_R$；而无法用根式表示的那些数，比如 π^n，e^n 与 $\sin 1\ (rad)$，称作**超越数**（transcendental）．

<4.>四元数 $\mathbb{H}$ 表示为 $\alpha + \beta i + \gamma j + \delta k$ 使 $i^2 = j^2 = k^2 = ijk = -1$，它的乘法不总是满足交换律，感兴趣的同学可以自己研究．之后 Cayley（后面还会经常讨论到他）和 Hamilton 的朋友 Graves 各自独立地研究了**八元数**（octonion）$\mathbb{O}$ 的性质．

<5.>其实我们这里讨论的就是抽象代数的基础内容．

<6.>事实上模算数最早起源于古代的中国和印度．

<7.>同余类中任意一个数放在直角括号内可以表示这个同余类．

<8.>尼尔斯·亨利克·阿贝尔（Niels Henrik Abel，1802年8月5日—1829年4月6日），挪威数学家，在很多数学领域做出了开创性的工作．他最著名的一个成果是首次完整地给出了高于四次的一般代数方程没有一般形式的代数解的证明．这个问题是他那个时代最著名的未解决问题之一，悬疑达250多年．他也是椭圆函数领域的开拓者．尽管阿贝尔成就极高，却在生前没有得到认可，他的生活非常贫困．

<9.>埃瓦里斯特·伽罗瓦（Évariste Galois，1811年10月25日—1832年5月31日），法国数学家，群论的创立者．利用群论彻底解决了根式求解代数方程的问题，并由此发展了一整套关于群和域的理论，人们称之为伽罗瓦理论，并把其创造的"群"叫作伽罗瓦群．他因强烈支持共和主义受到保皇势力迫害入狱，20岁时死于一场"爱情和荣誉"的决斗．他的传记很值得一读．

<10.>其实我们将封闭性放在二元运算的定义中，即必须满足封闭性才是二元运算．

<11.>相应地，交换律代表着运算的一种"对称性"，分配律代表着两种运算间的优先级和一种"穿透性"．范畴论的学习可以帮助你加深对它们的体会．idea 来自：交换律、结合律、分配律的本质是什么？ - 知乎用户的回答 - 知乎 https://www.zhihu.com/question/285971671/answer/446968492

<12.>没错就是用那个英年早逝的维京人 Niels Abel 名字命名的．

<13.>既是单射又是满射，就是一一对应的映射．对于每个 x，总有唯一的 $\phi(x)$ 与之对应，反之亦然．

<14.>同构并不是群之间最强的等价关系，在群表示论中有**等价表示**（equivalent representation），要求两个置换群或线性群同为一个对称群或一般线性群的子群且互为共轭（下一章会说到）.

<15.>没见过我这么不严谨的叫法吧哈哈哈，**原群**（magma，hot liquid rock found just below the surface of the earth...但是它真叫这个名字）是带有封闭二元运算的集合，比群的定义要宽松许多．当然这里我指的是"子群所属于的那个我们一开始研究的群".

<16.>Bézout 定理是初等数论中的重要定理，译作裴蜀定理或贝祖定理．表述为设 a, b 是不全为零的整数，对任意整数 x, y，满足 $\gcd(a, b) | ax + b$，且存在整数 x, y，使得 $ax + by = \gcd(a, b)$．其中 $\gcd(a, b)$ 表示 a 与 b 的**最大公约数**（greatest common divider）．其证明可以参考：https://oi-wiki.org/math/number-theory/bezouts/

<17.>它其实是一个整环，叫做**多项式环**（polynomial ring）．它是交换代数和代数几何等学科中十分重要且基础的研究对象.

<18.>艾米·诺特（Emmy Noether，1882年3月23日—1935年4月14日），德国数学家．她彻底改变了环、域和代数的理论；在物理学方面，**Noether 定理**（Noether's theorem）解释了对称性和守恒定律之间的根本联系．她还被称为"现代数学之母"，她允许学者们无条件地使用她的工作成果，也因此被人们尊称为"当代数学文章的合著者"．她的 *Idealtheorie in Ringbereichen*（《环中的理想论》）这篇重要论文标志着抽象代数学真正成为一门数学分支．然而，巨大的声誉并未改善她受到性别歧视的处境，在不合理的制度下，她的生活非常艰苦.

Appendix-ii-练习

练习对于抽象代数这样重视概念和抽象对象的理解的学科而言极其重要．例子可以帮助你构建起对于概念的第一印象，而练习则可以让你真正地将概念理解为你自己的东西．下面给出了几道练习题，当然，这是远远不够的，你必须学会对你遇到的各种结构使用所学去检验，甚至凭借自己想象构造一些结构．下面有些练习难度可能较大，量力而行.

1. 【模运算的计算】计算：

$$\begin{aligned} 100 \mod 7 &= \\ (16 \times 11) \mod 13 &= \\ (2/9) \mod 10 &= \end{aligned}$$

2. 【群的定义的理解】证明 $\mathbb{Z}_n^* := \{x \in \mathbb{Z}_n \mid \gcd(x, n) = 1 \wedge x > 0\}$ 在模-n 乘法下为群.
3. 【群的定义的理解】（如果你学过线性代数）确定 $GL(2, \mathbb{Z}_2)$ 是不是Abel群并给出证明.

4. 【同构的理解】**自同构**（automorphism）指的是一个将群中的每个元素映射到群中元素的同构映射．所有自同构映射的集合记作 $\mathrm{Aut}(G)$ ．**内自同构**（inner automorphism）指的是一个映射将群中每个 $x \in G$ 映射到 axa^{-1} ．所有内自同构的集合记作 $\mathrm{Inn}(G)$ ．证明它们在映射复合运算下为群．如果有能力，证明 $\mathrm{Aut}(\mathbb{Z}_n^+) \cong \mathbb{Z}_n^*$ ．

5. 【置换群的理解】求对称群 S_4 和它的子群 A_4 的阶数．如果对 n 元对称群 S_n 和它的子群 A_n ，它的阶数又该如何求得？试用排列组合的知识解决．

6. 【Cayley 表】画出 $\mathbb{Z}_6^+$ 和 $\mathbb{Z}_7^*$ 的乘法表，观察它们是否同构．

7. 【子群的理解】如果一个子集元素个数有限，最少需要证明几条群公理就能证明它是子群？分别是哪几条？证明你的猜想．

8. 【Fermat 小定理】计算 $17^{97} \mod 13$ ．

9. 【正规性&商群】证明 $G/Z(G) \cong \mathrm{Inn}(G)$ ．

10. 【同态和同态定理】找出合适的群乘法使 $\mathbb{R}/\mathbb{Z} \cong S^1$ 成立并证明．其中 S^1 是一维球面，即圆．

11. 【同态和同态定理】证明第二同构定理：已知 $N \lhd G$ 且 $H \leq G$ ，则 $H/H \cap N \cong HN/N$ ．其中 $HN := \{hn \mid h \in H \wedge n \in N\}$ ．

12. 【外直积与分类】（拓展题）证明任意阶数为 p^2 的群要么同构于 $\mathbb{Z}_{p^2}$ ，要么同构于 $\mathbb{Z}_p \oplus \mathbb{Z}_p$ ．

13. 【计算机技术与应用】把具有部分素数性质的合数叫做**拟素数**（quasi prime number）．用 Python 或其它你熟悉的编程语言写一个程序，找出以 $a = 3, 24, 37$ 的 Fermat 小定理为判据时最小的 3 个拟素数．

14. 【数学与生活】魔方是布达佩斯大学建筑学院的 Ernő Rubik 教授发明的一种益智游戏，自从面世以来广受欢迎（一个标准的三阶魔方如图）．试用置换表示一个二阶魔方的任意操作，思考它们是否构成一个群．查阅相关资料，了解魔方、群论与计算机的关系．你也可以尝试自己解决二阶魔方的一些问题（如：至少需要几步操作可以还原任意一种打乱的二阶魔方？）．

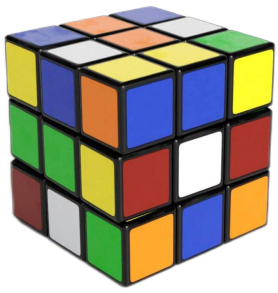


fig_2a.1

Appendix-iii-参考材料

按照字母顺序：

David S. Dummit & Richard M. Foote, *Abstract Algebra, 3rd*. John Wiley & Sons, 2003.

東雲正樹 (2020, Nov. 30), 群论 (Group Theory) 终极速成 / 物理系零基础火箭级 notes, https://zhuanlan.zhihu.com/p/643534515

J. Conway & S. Norton, *Monstrous Moonshine*. Cambridge, Bull. Lond. Math. Soc., 11:308-339, 1979. DOI: 10.1112.

John M. Lee, *Introduction to Smooth Manifolds, 2nd*. New York, Springer New York, NY, 2012.

Joseph A. Gallian, *Contemporary Abstract Algebra, 10th*. Taylor & Francis Group, LLC, 2021.

J. Stillwell, *Mathematics and Its History, 3rd*. New York, Springer New York, NY, 2010.

J. Voight, *Quaternion Algebras*. Cham, Springer Nature Switzerland AG, 2021.

李新征, *《群论一》讲义*. 北京大学物理学院, 2018.

N. Carter, *Visual Group Theory*. Washington, DC, Mathematical Association of America, 2009.

S. Carnahan & S. Urano, *Monstrous Moonshine for Integral Group Rings*, arXiv:2111.09404v2 [math.RT], 2023. DOI: 10.48550.

S. Lang, *Algebra, 3rd.(revised)*. New York, Springer New York, NY, 2002.

S. Roman, *Advanced Linear Algebra, 3rd*. New York, Springer-Verlag New York 2008.

Thomas W. Hungeford, *Algebra*. New York, Springer-Verlag New York, 1974.

V. Tatitscheff, *A short introduction to Monstrous Moonshine*. Strasbourg, arXiv:1902.03118v4 [math.NT], 2021. DOI: 10.48550.

如何深入浅出地讲解Maxwell方程组?

by 研研仔（答主）-知乎

original link https://www.zhihu.com/question/36766702/answer/70426471

题主：如何尽可能使用形象朴素的语言讲解麦克斯方程组。微积分可以不做深入解释，但是尽可能使用高中知识解释方程组的推导及其物理意义？

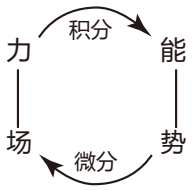
答主：题主简直坑爹。不讲微积分怎么给你讲麦克斯韦方程组？你不知道麦克斯韦方程组里面每个方程都是一个积分或者微分么？？那既然这样，我只能躲躲闪闪，不细谈任何具体的推导和数学关系，纯粹挥挥手扯扯淡地说一说电磁学里的概念和思想。

1. 力、能、场、势

经典物理研究的一个重要对象就是**力**force。比如牛顿力学的核心就是 $\mathbf{F}=m\mathbf{a}$ 这个公式，剩下的什么平抛圆周简谐运动都可以用这货加上微积分推出来。但是力有一点不好，它是个**向量**vector（既有大小又有方向），所以即便是简单的受力分析，想解出运动方程却难得要死。很多时候，从能量的角度出发反而问题会变得简单很多。**能量**energy说到底就是力在空间上的积分（能量＝功＝力×距离），所以和力是有紧密联系的，而且能量是个**标量**scalar，加减乘除十分方便。分析力学中的拉格朗日力学和哈密顿力学就绕开了力，从能量出发，算运动方程比牛顿力学要简便得多。

在电磁学里，我们通过力定义出了**场**field的概念。我们注意到洛仑兹力总有着 $\mathbf{F}=q\left(\mathbf{E}+\mathbf{v}\times\mathbf{B}\right)$ 的形式，具体不谈，单看这个公式就会发现力和电荷（或电荷×速度）成正比。那么我们便可以刨去电荷（或电荷×速度）的部分，仅仅看剩下的这个“系数”有着怎样的动力学性质。也就是说，场是某种遍布在空间中的东西，当电荷置于场中时便会受力。具体到两个电荷间的库仑力的例子，就可以理解为一个电荷制造了电场，而另一个电荷在这个电场中受到了力，反之亦然。类似地我们也可以对能量做相同的事情，刨去能量中的电荷（或电荷×速度），剩下的部分便是**势**potential。

一张图表明关系：



具体需要指出，这里的电场（标为 $\mathbf{E}$ ）和磁场（标为 $\mathbf{B}$ ）都是向量场，也就是说空间中每一个点都对应着一个向量。如果我们把 x, y, z 三个分量分开来看的话，这就是三个标量场。而能量和势是标量（电磁学中的势其实并不是标量，原因马上揭晓），放到空间中也就是一个标量场。在力/场和能量/势之间互相转化的时候，我们是在三<->一个标量场之间转化，必然有一些信息是丢掉了的。怎么办？

一个显而易见的答案是**保守力场**conservative force field。在这样一个场中，能量（做功）不取决于你选择什么样的路径。打个比方，你爬一座山，无论选择什么路径，只要起点和终点一样，那么垂直方向上的差别都是一样的，做的功也一样多。在这种情况下，我们对力场有了诸多限制，也就是说，我假如知道了一个保守力场的 x 一个分量，那么另两个分量 y, z 就随之确定了，我没得选（自由度其实只有一个标量场）。有了保守力场这样的额外限制，向量场$\mathbf{F}$ （三个标量场）和（一个）标量场 V 之间的转化便不会失去信息了。具体而言，二者关系可以写作 $\mathbf{F} = -\nabla V$ 。这里不说具体细节，你只要知道 ∇ 是一种固定的、把一个标量场变成三个标量场的算法就可以了（叫做**算符**operator）。

那么我们想问，电场和磁场是不是保守力场呢？很不幸，不是。在静电学中，静止的电场是保守的，但在电动力学中，只要有变化的电场和磁场，电场就不是一个保守力场了；而磁场从来都不是保守力场。这也就是说明，在电磁学中，我们很少涉及能量这个概念，因为它不能完整地描述一个电磁场。我们更多时候只关注“场”这个概念，尽管因此我们不得不涉足很多向量微积分，但我们没有办法，这是不让信息丢掉的唯一办法。那么，既然势也是标量，它是否也是一个没什么用的概念呢？恰恰相反，在电动力学中我们定义出了**向量势**vector potential，以保留额外的自由度。后面我会更具体地谈到这一点。

总而言之，我想说明一点，那就是电磁学的主要研究对象是电场和磁场，而麦克斯韦方程组就是描述电场和磁场的方程。势（包括电势和磁向量势）也是有用的概念，而且不像引力势是一个标量，在电磁学中势不得不变成一个向量。

2. 麦克斯韦方程组

前边说到，**麦克斯韦方程组**Maxwell equations**是描述电场和磁场的方程**。前边也说到，因为电磁场不是保守力场，它们有三个标量场的自由度，所以我们必须用向量微积分来描述电磁场。因此，麦克斯韦方程组每个式子都出现了向量微积分，而整个方程组也有**积分形式**和**微分形式**两种。这两种形式是完全等价的，只是两种不同的写法。这里我先全部写出。

积分形式：

$$
\begin{aligned}
&(1\text{-}1) \quad \oint_{\partial V} \mathbf{E} \cdot \mathrm{d}\mathbf{a} = \frac{Q_V}{\epsilon_0} \\
&(1\text{-}2) \quad \oint_{\partial S} \mathbf{E} \cdot \mathrm{d}\mathbf{l} = -\frac{\mathrm{d}}{\mathrm{d}t} \int_S \mathbf{B} \cdot \mathrm{d}\mathbf{a} \\
&(1\text{-}3) \quad \oint_{\partial V} \mathbf{B} \cdot \mathrm{d}\mathbf{a} = 0 \\
&(1\text{-}4) \quad \oint_{\partial S} \mathbf{B} \cdot \mathrm{d}\mathbf{l} = \mu_0 I_S + \mu_0 \frac{\mathrm{d}}{\mathrm{d}t} \int_S \mathbf{E} \cdot \mathrm{d}\mathbf{a}
\end{aligned}
$$

微分形式：

$$
\begin{aligned}
&(2\text{-}1) \quad \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \\
&(2\text{-}2) \quad \nabla \times \mathbf{E} = -\frac{\partial}{\partial t} \mathbf{B} \\
&(2\text{-}3) \quad \nabla \cdot \mathbf{B} = 0 \\
&(2\text{-}4) \quad \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \mathbf{E}
\end{aligned}
$$

这里 $\mathbf{E}$ 表示电场， $\mathbf{B}$ 表示磁场， ϵ_0 和 μ_0 只是两个常数暂时可以忽略。积分形式中 Q 是电荷， I 是电流， V 表示一块体积， ∂V 表示它的表面，而 S 表示一块曲面， ∂S 表示它的边缘。微分形式中 ρ 是电荷密度（电荷/体积）， $\mathbf{J}$ 是电流密度（电流/面积）， $\nabla\cdot$ 和 $\nabla\times$ 是两个不同的算符，基本可以理解为对向量的某种微分。

先不说任何细节，我们可以观察一下等式的左边。四个方程中，两个是关于电场 $\mathbf{E}$ 的，两个是关于磁场 $\mathbf{B}$ 的；两个是曲面积分 $\int \cdots \mathrm{d}\mathbf{a}$ 或者散度 $\nabla\cdot$ ，两个是曲线积分 $\int \cdots \mathrm{d}\mathbf{l}$ 或者旋度 $\nabla\times$. 不要管这些术语都是什么意思，我后面会讲到。但光看等式左边，我们就能看出四个式子分别描述电场和磁场的两个东西，非常对称。

3. 电荷->电场，电流->磁场

这一部分和下一部分中，我来简单讲解四个式子分别代表什么意思，而不涉及任何定量和具体的计算。

我们从两个电荷之间的库仑力讲起。**库仑定律**Coulomb's Law是电学中大家接触到的最早的定律，有如下形式：

$$
(3) \quad \mathbf{F} = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{r^2} \hat{\mathbf{r}}
$$

其中 Q 是电荷， r 是电荷之间的距离， $\hat{\mathbf{r}}$ 是表示方向的单位向量。像我之前说的，把其中一个电荷当作来源，然后刨去另一个电荷，就可以得到电场的表达式。

高中里应该还学过**安培定律**Ampere's Law，也就是电流产生磁场的定律。虽然没有学过具体表达式，但我们已经能看出它与库仑定律之间的区别。库仑定律描述了“两个”微小来源（电荷）之间的“力”，而安培定律是描述了“一个”来源（电流）产生的“场”。事实上，电磁学中也有磁场版本的库仑定律，描述了两个微小电流之间的力，叫做**毕奥-萨伐尔定律**Biot-Savart Law；反之，也有电场版本的安培定律，描述了一个电荷产生的磁场，叫做**高斯定律**Gauss's Law。这四个定律之间有如下关系：

\	电场	磁场
两个微小来源之间的力	库仑定律	毕奥-萨伐尔定律
单个来源产生的场	高斯定律	安培定律

数学上可以证明库仑定律（毕奥-萨伐尔定律）和高斯定律（安培定律）在静电学（静磁学）中是完全等价的，也就是说我们可以任意假设一个定律，从而推导出另一个定律。然而如果我们想从静止的静电学和静磁学推广到电动力学，前者是非常不便的而后者很却容易，所以尽管库仑定律在中学中常常提到，麦克斯韦方程组中却没有它，有的是高斯定律和安培定律。这两个定律分别是麦克斯韦方程组里的(1)和(4)的第一项，即：

高斯定律（积分、微分形式）：

$$
\begin{aligned}
&(4\text{-}1) \quad \oint_{\partial V} \mathbf{E} \cdot \mathrm{d}\mathbf{a} = \frac{Q_V}{\epsilon_0} \\
&(4\text{-}2) \quad \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}
\end{aligned}
$$

安培定律（积分、微分形式）：

$$
\begin{aligned}
&(5\text{-}1) \quad \oint_{\partial S} \mathbf{B} \cdot \mathrm{d}\mathbf{l} = \mu_0 I_S \\
&(5\text{-}2) \quad \nabla \times \mathbf{B} = \mu_0 \mathbf{J}
\end{aligned}
$$

我们继续推迟讲解数学关系，单看这几个式子本身，就能看到等式的左边有电场 $\mathbf{E}$ （磁场 $\mathbf{B}$ ），而右边有电荷 Q （电流 I ）或电荷密度 ρ （电流密度$\mathbf{J}$ ）。看，**电荷产生电场，电流产生磁场**！

4. 变化磁场->电场，变化磁场->电场

然而这不是故事的全部，因为事实上电磁场是可以互相转化的。法拉第发现了电磁感应，也就是说变化的磁场是可以产生电场的，这就是**法拉第定律**Faraday's Law。类似地，麦克斯韦发现安培定律的描述并不完善，除了电流以外，变化的电场也可以产生磁场，这被称为**安培-麦克斯韦定律**Ampere-Maxwell Law。这两个定律分别是麦克斯韦方程组里的(2)和(4)的第二项，即：

法拉第定律（积分、微分形式）：

$$(6\text{-}1)\quad \oint_{\partial S} \mathbf{E}\cdot \mathrm{d}\mathbf{l} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{B}\cdot \mathrm{d}\mathbf{a}$$

$$(6\text{-}2)\quad \nabla\times\mathbf{E} = -\frac{\partial}{\partial t}\mathbf{B}$$

安培-麦克斯韦定律（积分、微分形式）：

$$(7\text{-}1)\quad \oint_{\partial S} \mathbf{B}\cdot d\mathbf{l} = \mu_0\epsilon_0\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{E}\cdot \mathrm{d}\mathbf{a}$$

$$(7\text{-}2)\quad \nabla\times\mathbf{B} = \mu_0\epsilon_0\frac{\partial}{\partial t}\mathbf{E}$$

同样地，等式的左边有电场 $\mathbf{E}$ （磁场$\mathbf{B}$ ），而右边有磁场 $\mathbf{B}$ （电场 $\mathbf{E}$ ）的导数 $\frac{\mathrm{d}}{\mathrm{d}t}$ 或偏导 $\frac{\partial}{\partial t}$ 。看，**变化磁场产生电场，变化电场产生磁场**！

需要指出的是，我这样的说法其实是不准确的，因为并不是真的某一个场"产生"的另一个场。这两个定律只是描述了电场（磁场）和磁场（电场）的变化率之间的定量关系，而不是因果关系。

小结一下，我们已经搞清楚了麦克斯韦方程组里每一项的意思，基本就是指出了电磁场的来源和变化电磁场的定量关系。下一步便是往我们这些粗浅的理解中加入数学，具体看看这些方程到底说了什么。在这之前，我们必须花一点时间了解一下向量微积分的皮毛。

5. 向量积分

普通的单变量微积分基本可以理解为乘法的一种拓展。我们想计算一个矩形的面积，我们用长 x 乘宽 y ，即 xy 。如果宽不是一个定值而是根据长而变化的（也就是说宽是一个长的函数，即宽 $= y(x)$ ），那么我们就需要积分，记为 $\int y(x)\,\mathrm{d}x$ 。这样的想法也很容易推广到更高的维度，比如在一块体积 V 内，若电荷密度为 ρ ，那么这块体积内的总电荷就是 $Q=\rho V$ ；如果 ρ 在空间中每一点都不一样，是个关于坐标的函数 $\rho(\mathbf{x})$ ，那么就要变成积分 $Q=\iiint \rho(\mathbf{x})\,\mathrm{d}V$ （这里 $\iiint$ 表示是一个三维的积分，很多时候也可以省略写为一个 $\int$ ）。

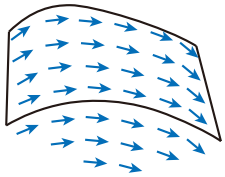
在向量场中，这个事情比较麻烦。首先两个向量的乘积的定义稍显复杂，必须使用**点乘**dot product，即 $\mathbf{u}\cdot\mathbf{v}$ ，它暗示着两个向量之间的角度，也就是有多么平行。如果 $\mathbf{u}$ 和 $\mathbf{v}$ 完全平行，它们的点乘是一个正值；如果方向相反，则是一个负值；如果垂直，那么为 0 。另一方面，我们不一定要像上一个电荷的例子一样积上整个体积 V ，我们可以只积一个曲面 S 或者一条曲线 γ 。这就是所谓的曲面积分和曲线积分的概念。

曲面积分surface integral有如下形式：

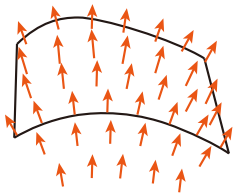
$$(8)\quad \int_S \mathbf{F}\cdot \mathrm{d}\mathbf{a}$$

其中 S 表示我们需要积的曲面，$\mathbf{F}$ 是我们想要积的向量场，$\cdot$ 代表点乘，$\mathbf{a}$ 指向垂直于 S 的方向。因此，我们看到，如果 $\mathbf{F}$ 和 S 是平行的，那么点乘处处得 0，这个曲面积分也为 0。换句话说，**曲面积分表示着向量场 $\mathbf{F}$ 穿过曲面 S 的程度**，因此也很形象地叫做**通量flux**。下图为两个简单的例子：

曲面积分（通量）为 0：



曲面积分（通量）不为 0：

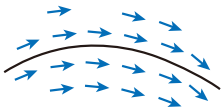


那么**曲线积分**line integral也很类似，只不过我们不积一个曲面 S 而是一个一维的曲线 γ。它有如下形式：

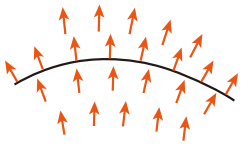
$$\int_{\gamma} \mathbf{F} \cdot \mathrm{d}\mathbf{l} \tag{9}$$

其中 γ 表示我们需要积的曲线，$\cdot$ 代表点乘，$\mathbf{l}$ 指向曲线 γ 的方向。不难看出，**曲线积分表示着向量场 $\mathbf{F}$ 沿着曲线 γ 的程度**。下图为两个简单的例子（线——表示曲线 γ）：

曲线积分不为 0：



曲线积分为 0：



特别地，如果曲线是闭合的（首尾相连的），那么我们可以在积分符号$\int$上画一个圈，表示闭合，然后这个特殊的曲线积分叫做**环量**circulation，因为是积了一个环嘛。很显然，如果 $\mathbf{F}$ 是个保守力场，那么我随便找一个闭合曲线，做的功都一定为 0（这就是保守力场的定义啊），所以**保守力场的任意环量都为** 0 。最后一提，“环量”这个名字很少使用，一般就直接叫做“闭合曲线的积分”。

定义一个通量所使用的曲面 S 则不一定要是闭合的，任何曲面都可以。如果这个曲面很特殊恰好是闭合的，我们也可以在积分符号 $\iint$ 上画上一个圈，代表闭合，但这个量则没有一个特殊的名字了。

总结如下表：

\	**曲面积分**	**曲线积分**
表示向量场	通过曲面	沿着曲线
又叫做	通量	—
若为闭合	—	环量

6. 麦克斯韦方程组的积分形式

我非常不严谨地描述了曲面积分和曲线积分分别是什么。我们回头看看麦克斯韦方程组的积分形式，我们应该都能看懂了。

$$
\begin{aligned}
&(1\text{-}1) \quad \oint_{\partial V} \mathbf{E}\cdot \mathrm{d}\mathbf{a} = \frac{Q_V}{\epsilon_0} \\
&(1\text{-}2) \quad \oint_{\partial S} \mathbf{E}\cdot \mathrm{d}\mathbf{l} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{B}\cdot \mathrm{d}\mathbf{a} \\
&(1\text{-}3) \quad \oint_{\partial V} \mathbf{B}\cdot \mathrm{d}\mathbf{a} = 0 \\
&(1\text{-}4) \quad \oint_{\partial S} \mathbf{B}\cdot \mathrm{d}\mathbf{l} = \mu_0 I_S + \mu_0 \frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{E}\cdot \mathrm{d}\mathbf{a}
\end{aligned}
$$

(1) **高斯定律**：电场 $\mathbf{E}$ 在闭合曲面∂V 上的通量，等于该曲面包裹住的体积 V 内的电荷 Q （乘上系数 $\frac{1}{\epsilon_0}$ ）；
(2) **法拉第定律**：电场 $\mathbf{E}$ 在闭合曲线 ∂S 上的环量，等于磁场 $\mathbf{B}$ 在该曲线环住的曲面 S 上的通量

的变化率（乘上系数 -1 ）；
(3) **高斯磁定律**：磁场 $\mathbf{B}$ 在闭合曲面 ∂V 上的通量，等于 0；
(4) **安培麦克斯韦定律**：磁场 $\mathbf{B}$ 在闭合曲线 ∂S 上的环量，等于该曲线环住的曲面 S 里的电流 I（乘上系数 μ_0 ），加上电场 $\mathbf{E}$ 在该曲线环住的曲面 S 上的通量的变化率（乘上系数 $\mu_0\epsilon_0$ ）。

虽然在我看来，这样的描述已经是非常通俗、没有任何数学了，但对于没有学习过微积分的同学来说，显然还是太晦涩了一点。那么我来举几个例子吧。

(1) **高斯定律：**

例子1：假设我们有一个点电荷 Q ，以其为球心作一个球，把这块体积称为 V ，那么 ∂V 就是这个球的表面。这个电荷 Q 产生了一些电场，从中心的 Q 向外发射，显然电场线都穿过了球的表面 ∂V ，所以“闭合曲面 ∂V 的通量”是个正数，不为 0 ，而“该曲面包裹住的电荷”为 Q ，也不为 0。

例子2：假设我们把电荷 Q 替换为 $-Q$ ，那么所有的电场线方向都反过来了， ∂V 的通量（记得通量中的点乘吗？）也因此获得了一个负号，所以“闭合曲面 ∂V 的通量”变成了负数，而“该曲面包裹住的电荷”为 $-Q$ ，也变成了负数。等式再一次成立。

例子3：假设我们把这个球的半径扩大为原来的 2 倍，这个球的表面积就变成了原来的 4 倍。与此同时，由于库仑力的反比平方定律，由于球表面与球心电荷 Q 的距离变成了原来的 2 倍，在球表面 ∂V 的电场强度也变成了原来的 $1/4$ 。通量（电场和面积的积分）获得一个系数 4 ，又获得一个系数 $1/4$ ，所以“闭合曲面 ∂V 的通量”没有变，而“该曲面包裹住的电荷”显然仍然为 Q ，也没有变。

例子4：事实上，我们随便怎么改变这一块表面积的大小、体积，算出来的通量都不会变（尽管会非常难算），因为等式的右边“该曲面包裹住的电荷”一直都没有变。

例子5：假设我们把电荷移到这个曲面外面，那么电场线会从这个球的一面穿透进去，然后从另一面出来，所以当我们做积分的时候，两个方向的通量抵消了，整个“闭合曲面 ∂V 的通量”为 0 ，而此时我们的曲面没有包裹住任何电荷，所以“该曲面包裹住的电荷”也为 0 。等式成立。

(2) **法拉第定律：**

例子6：一圈闭合导线，环住了一块曲面 S ，则记这个曲线的位置为 ∂S ，那么经过 ∂S 的电场 $\mathbf{E}$ 的环量其实就是导线内的电势（电压）。垂直于 S 通过一些磁场 $\mathbf{B}$ ，则通过 S 的磁通量不为 0 。然而此时导线内并没有电流，也就是说，并没有电压，“闭合曲线 ∂S 的环量”为 0 。这是很显然的，因为磁通量并没有变化，没有电磁感应，换句话说，“曲面 S 上的通量的变化率”为 0 。

例子7：这个时候我突然增加磁场，所以磁通量变大了，“磁通量的变化率”为正，不为 0 。因此，等式的左边“闭合曲线 ∂S 的环量”也为正，不为 0 ，也就是说，导线内产生了一些电压，继而产生了一些感应电流。这正是大家熟悉的法拉第电磁感应。

例子8：如果我不是增加磁场，而是减小磁场，那么磁通量变小了，“磁通量的变化率”为负。那么等式左边“闭合曲线 ∂S 的环量”也获得了一个负号，换句话说，感应电流的方向反了过来。

(3) **高斯磁定律：**

例子9：随便选择一个闭合曲面，整个曲面上的磁通量一定为 0 。这和电场的情况迥然不同，因此说明，不像有可以产生电场的“电荷”，这个世界上是没有能单独产生磁场的“磁荷”（也就是“磁单极子”）的（编者注：现代理论认为可能存在磁单极子，但与高斯磁定律并不矛盾）。

(4) **安培-麦克斯韦定律：**

例子10：假设我们有一个电流 I ，以其为轴作一个圆，把这个圆称为 S ，那么 ∂S 就是这个圆的边缘。这个电流 I 产生了一些磁场，（按照右手定则）绕着导线。显然磁场线和 ∂S 都是“绕着导线”，方向一致，所以“闭合曲线 ∂S 的环量”是个正数，不为 0 ，而“该曲线环住的电流”为 I ，也不为 0 。

例子11：假设我们改变电流方向，即把 I 变成 $-I$ ，那么所有的磁场线方向都反过来了， ∂S 的环量也因此获得了一个负号，所以“闭合曲线 ∂S 的环量”和“该曲线环住的电流”均获得一个负号。等式再一次成立。

例子12：和高斯定律很像，我们随便怎么改变这一个环的大小、面积，只要环住的电流不变，算出来的环量都不会变（尽管可能会非常难算）。而若电流在这个环外面，尽管仍然有磁场存在，但在计算环量时相互抵消，使得等式两边都变成 0 。

例子13：“变化的电场产生磁场”（即第二项）的例子非常难找，这也正是安培当年没有自己发现、非要等到麦克斯韦帮忙才发现的原因。我这里不妨不再细述，读者只要接受这个设定就好。有兴趣的读者可以自己思考一个这种情况的例子。

最后，还记得我们之前说过“保守力场的任意环量都为 0 ”吗？显然，要想让磁场的环量为 0 ，那就只能既没有电流（方程(4)中的第一项），也没有变化的电通量（第二项），那么磁场只能为 0 。换言之，任何磁场都不是保守力场。想让电场的环量为 0 还比较简单，只需要令磁通量不变（方程(2)）就好了。换言之，只有在静电学（电磁场均静止不变）中，静电场才是保守力场。

7. 向量微分

麦克斯韦方程组描述了所有的电磁现象，从每个方程的名字也可以看出，方程组总结、整合了前人（库仑、高斯、安培、法拉第等）发现的各种现象和其方程（在麦克斯韦以前这样的方程可能有数十个），而麦克斯韦把它们总结归纳到了一起，用短短四个公式涵盖了所有现象，非常了不起。然而平心而论，积分形式仍然显得颇为繁琐，原因有二：1. 积分是很难算的，虽然每一个方程的左右两边都必然相等，但随便给你一个场和一个曲面/曲线，想把左侧的积分算出来极为困难；2. 也正因为如此，我们尽管有可以描述电磁场的方程，但给定一个特定的来源（比如天线中一个来回摇摆

的电荷），我们想算出具体的$\mathbf{E}$ 和 $\mathbf{B}$ 也是极为困难，因为我们只知道 $\mathbf{E}$ 和 $\mathbf{B}$ 在某个特殊曲面/曲线上的积分。

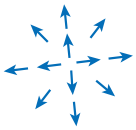
这就是微分形式的好处。首先，计算一个给定向量场的微分（散度和旋度）是很简单的，只要使用之前提到过的 $\nabla\cdot$ 和 $\nabla\times$ 算符就好，而这两个算符都有一套固定的算法。其次，散度和旋度代表着一个向量场的两种不同的自由度，有着非常直接的几何意义，从这两个量中恢复出向量场也是比较直观的过程。当然，我们又需要再准备一些向量微积分的知识，其中的重点就是散度和旋度。

散度divergence，顾名思义，是**指一个向量场发散的程度**。一个向量场 $\mathbf{F}$ 的散度是一个标量场（向量场的每一点有一个自己的散度），写作 $\nabla\cdot\mathbf{F}$ （这个写法也很直白，因为点乘就是标量）。如果一个点的散度为正，那么在这一点上 $\mathbf{F}$ 有向外发散的趋势；如果为负，那么在这一点上 $\mathbf{F}$ 有向内收敛的趋势。

旋度curl则**指一个向量场旋转的程度**。一个向量场 $\mathbf{F}$ 的旋度是一个向量场（向量场的每一点有一个自己的旋度，而且是一个向量；这是因为旋转的方向需要标明出来），写作 $\nabla\times\mathbf{F}$ （这个写法也很直白，因为叉乘就是向量）。如果一个点的旋度不为 0 ，那么在这一点上 $\mathbf{F}$ 有漩涡的趋势，而这个旋度的方向表明了旋转的方向。

举些例子，以下是两个向量场的例子。其中第一个向量场往外发散，但完全没有旋转扭曲的趋势；第二个向量场形成了一个标准的漩涡，但没有任何箭头在往外或往里指，没有发散或收敛的趋势。

散度不为 0 、但旋度为 0 的向量场：



旋度不为 0 、但散度为 0 的向量场：

因此，如你所见，散度和旋度描述的都是非常直观的几何性质。只要知道一个向量场的散度和旋度，我们就可以唯一确定这个向量场本身（这是亥姆霍兹定理，我要是有兴致可以以后简单谈谈）。

麦克斯韦方程组的微分形式，就是要描述电磁场的散度和旋度。我前边说到，微分形式和积分形式是完全等价的，我很也可以很轻松地从一个形式推导出另一个形式，用的是（不要和高斯定律混淆、又叫散度定理）和斯托克斯定理。

高斯定理Gauss's Theorem：一个向量场 $\mathbf{F}$ 在闭合曲面 ∂V 上的通量，等于该曲面包裹住的体积 V 里的 $\mathbf{F}$ 全部的散度（ $\mathbf{F}$ 的散度的体积积分）。这是可以想象的，毕竟通量就是在计算有多少场从这个闭合曲面里发散出去了，也就是总共的散度（散度的积分）。

斯托克斯定理Stokes' Theorem：一个向量场 $\mathbf{F}$ 在闭合曲线 ∂S 上的环量，等于该曲线环住的曲面 S 上的 $\mathbf{F}$ 全部的旋度（ $\mathbf{F}$ 的旋度的曲面积分）。这也是可以想象的，毕竟环量就是在计算有多少场和这个环方向一样（有多少场在沿着这个环旋转），也就是总共的旋度（旋度的积分）。

总结如下表：

\	**曲面积分**	**曲线积分**
积分形式	通量	环量
联系	高斯定理	斯托克斯定理
微分形式	散度	旋度

8. 麦克斯韦方程组的微分形式

了解了散度和旋度的概念之后，我们便可以读懂麦克斯韦方程组的微分形式了。

$$
\begin{aligned}
&(2\text{-}1) \quad \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \\
&(2\text{-}2) \quad \nabla \times \mathbf{E} = -\frac{\partial}{\partial t}\mathbf{B} \\
&(2\text{-}3) \quad \nabla \cdot \mathbf{B} = 0 \\
&(2\text{-}4) \quad \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial}{\partial t}\mathbf{E}
\end{aligned}
$$

(1) **高斯定律**：电场 $\mathbf{E}$ 的散度，等于在该点的电荷密度 ρ （乘上系数 $\frac{1}{\epsilon_0}$ ）；
(2) **法拉第定律**：电场 $\mathbf{E}$ 的旋度，等于在该点的磁场 $\mathbf{B}$ 的变化率（乘上系数 -1 ）；
(3) **高斯磁定律**：磁场 $\mathbf{B}$ 的散度，等于 0；
(4) **安培麦克斯韦定律**：磁场 $\mathbf{B}$ 的旋度，等于在该点的电流密度 $\mathbf{J}$ （乘上系数 μ_0 ），加上在该点的电场 $\mathbf{E}$ 的变化率（乘上系数 $\mu_0\epsilon_0$ ）。

我们可以看出，电荷和电流对电场和磁场干的事情是不一样的：电荷的作用是给电场贡献一些散度，而电流的作用是给磁场贡献一些旋度。然而变化的电磁场对对方干的事情是一样的，都是给对方贡献一些旋度。

想看一些具体例子的同学要失望了。微分形式的例子比较难举，因为微分形式主要是让计算更加简便，在数学上比较有优势，而应用到具体的现象上则不那么显而易见。不过，至少静电磁场的例子

还是可以举的。比如，我们知道电场线总是从正电荷出发、然后进入负电荷，这正是在说电场的散度在正电荷处为正，在负电荷处为负。再例如我们知道磁场线总是绕着电流，而不会进入或发源于电流，这也就是在说磁场有旋度而一定没有散度。

9. 电磁波

我刚刚提到，微分形式的主要好处是数学上处理起来很简便，我现在就给一个例子，也就是著名的光速。想象我们在真空中，周围什么都没有。这个时候，显然电荷密度和电流密度均为0，所以麦克斯韦方程组的微分形式变成了：

$$
\begin{aligned}
&(12\text{-}1) \quad \nabla \cdot \mathbf{E} = 0 \\
&(12\text{-}2) \quad \nabla \times \mathbf{E} = -\frac{\partial}{\partial t}\mathbf{B} \\
&(12\text{-}3) \quad \nabla \cdot \mathbf{B} = 0 \\
&(12\text{-}4) \quad \nabla \times \mathbf{B} = \mu_0\epsilon_0\frac{\partial}{\partial t}\mathbf{E}
\end{aligned}
$$

这四个公式简直太对称了！而且它们的含义也很清晰，基本就是说，变化的电场产生磁场，而变化的磁场产生电场。这就是**电磁波**electromagnetic wave的方程，电磁波也就是电场和磁场此消彼长、相互转化、向前传播的形式。

想要具体解出这个方程的解，还是需要玩儿一会儿微积分的，但是我们注意到两个式子分别有系数 -1 和 $\mu_0\epsilon_0$ 。如果你了解波动方程的话，从这两个系数就可以算出这个波传播的速度，为

$$
(13) \quad c = \frac{1}{\sqrt{\mu_0\epsilon_0}}
$$

然而！ μ_0 和 ϵ_0 这两个常数是真空的性质（分别叫做**真空电容率**vacuum permittivity和**真空磁导率**vacuum permeability），是个定值。换句话说，**电磁波传播的速度（光速）也是一个定值**！也就是说，在任何参考系里观察，光速都应该是一样的 c ！这根据伽利略速度相加原理是不可能的（静止的你认为火车的速度是 $50m/s$ ，那么如果你以 $1m/s$ 的速度往前走你就会认为火车的速度只有 $49m/s$ ，显然不会仍然是 $50m/s$ ），但是电磁学却实实在在地告诉我们光速是不会变的。呐，这就是相对论的由来了。

10. 方向性

可能有同学已经发现，我们的讨论中似乎忽略了很重要的一部分就是方向性。毕竟初高中学电磁的时候，出现了各种左手、右手定则（插一句，请一定一定忘掉左手定则，使用左手简直反人类，在正统的向量微积分和电磁学里**只有右手定则**）。在之前对于麦克斯韦方程组的诠释中，我们似乎很少提及方向。麦克斯韦方程组描述了方向性吗？

答案是肯定的。方向或者说手性（为什么是“右手”定则而不是“左手”定则？）来自于叉乘的定义和面积的向量微分元素的定义。我们定义叉乘 $\mathbf{u} \times \mathbf{v}$ 是一个向量，指的方向是垂直于 $\mathbf{u}$ 和 $\mathbf{v}$ 的方向；但显然有两个不同的方向均满足这个条件，而我们选择了其中特定的一个，把选择的这个规则叫做“右手定则”。类似地，一个曲面 S 也有两个方向（即其微分元素 $\mathrm{d}\mathbf{a}$ 是向量）。注意到曲线积分也是有方向性的（即其微分元素 $\mathrm{d}\mathbf{l}$ 也是向量），因此我们把 S 的 $\mathrm{d}\mathbf{a}$ 和 ∂S 的 $\mathrm{d}\mathbf{l}$ 联系起来，这个联系的规则也叫做“右手定则”。

上面这些情况中，选择“右手”是非常随意的；原则上我也可以全部选择左手，那么我得到的数学体系和原来的是完全等价的。当然，磁场 $\mathbf{B}$ 会和原来的磁场指的方向完全相反，但是没有关系，因为我们又不能直接看到磁场，所有的定律的手性都变了之后，描述的物理是不变的。但是，选择右手是约定俗成的，也就没必要再纠结为什么了。

11. 梯度、二次导数

我在之前说到保守力场的时候，偷偷塞进来过这样一个式子： $\mathbf{F} = -\nabla V$ 。这里 $\mathbf{F}$ 是个向量场， V 是个标量场。我们看到，这个神奇的倒三角不但可以表示散度（把向量变成标量）和旋度（把向量变成向量），还可以这样把一个标量场变成一个向量场！数学上这个倒三角叫**Nabla算符**，而 ∇V 叫做一个标量场 V 的梯度。

什么叫做梯度呢？其实相比于散度和旋度，这应该是更加熟悉的概念。**梯度**gradient就是**一个标量场变化的程度**。我们可以把一个标量场想象成一个山坡，每一点的梯度是一个向量，指的方向是上坡的方向，大小则是坡的陡峭程度。

总结一下我们见到的三种向量微分吧：

\	**梯度**	**散度**	**旋度**	
作用在一个	标量	向量	向量	**场上**
表示这个场	变化	发散	旋转	**的程度**
得到一个	向量	标量	向量	**场**
写作	∇V	$\nabla \cdot \mathbf{F}$	$\nabla \times \mathbf{F}$	

于是从 $\mathbf{F} = -\nabla V$ 这个公式我们看到，保守力场（比如引力场）可以表示为某个标量场（比如引力势能）的梯度。之前说过，保守力场的环量/旋度一定为 0 。这也就是说，梯度的旋度一定为 0。这是可以想象的，梯度指的是上坡的方向，而如果它有旋度，就意味着它们的指向可以形成的一个环，在这个环上可以一直上坡。这就像彭罗斯楼梯，是不可能的情形。

还有一个类似的定理，是说旋度的散度一定为 0 。我们也来想一下几何上这意味着什么。如果旋度有散度，就意味着在某个球上散度都在往球外指，也就意味着在球上每个点这个场都是逆时针旋转

的。想想也知道这是不可能的。所以我们得到了两个重要的结论：

1. **任意标量场 V 的梯度 ∇V 都是没有旋度的，也就是 $\nabla \times (\nabla V) = 0$；** 2. **任意向量场 $\mathbf{F}$ 的旋度 $\nabla \times \mathbf{F}$ 都是没有散度的，也就是 $\nabla \cdot (\nabla \times \mathbf{F}) = 0$ 。**

我说过，这些"X度"都可以认为是场的一种微分，那么这些"X度的X度"就可以认为是二次导数了。我们看到，有两种二次导数都自动为0，不必我们深究。还有一种二次导数也很有名，也就是梯度的散度，它甚至有了一个专门的花哨的名字，叫**拉普拉斯算符**Laplacian。在此我不作展开，大家只要知道它挺重要的就行。

12. 电荷守恒

从麦克斯韦方程组中可以直接推出电荷守恒。这个推导十分简单，且颇为有趣，可以让大家看到向量微积分的方便之处，我就简要写一下：

首先我们有安培-麦克斯韦定律：

$$(14\text{-}1) \quad \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \mathbf{E}$$

两边同时取散度：

$$(14\text{-}2) \quad \nabla \cdot (\nabla \times \mathbf{B}) = \mu_0 \nabla \cdot \mathbf{J} + \mu_0 \epsilon_0 \nabla \cdot \left(\frac{\partial}{\partial t} \mathbf{E} \right)$$

注意到左边是磁场的旋度的散度，而旋度的散度一定为 0 ，故左边为 0 。右边交换散度和时间导数，并约掉 μ_0 ，得：

$$(14\text{-}3) \quad 0 = \nabla \cdot \mathbf{J} + \epsilon_0 \frac{\partial}{\partial t} (\nabla \cdot \mathbf{E})$$

使用高斯定律：

$$(14\text{-}4) \quad \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$$

代入原式，约掉 ϵ_0 ，得：

$$(14\text{-}5) \quad 0 = \nabla \cdot \mathbf{J} + \frac{\partial}{\partial t} \rho$$

这个就是电荷守恒的公式。用语言说，就是**电流密度的散度加上电荷密度的变化率一定为 0 。**如果这比较抽象，我们可以对两项同时体积积分，再对 $\mathbf{J}$ 那项使用高斯定理变成面积积分，则结论变成：
一块体积 V 内的电荷的变化率加上通过表面 ∂V 的电流一定为 0 。

举个栗子，如果一块体积内的电荷 Q 变少了，其变化率为负，根据上述结论，通过表面的电流一定为正，也就是说有电流从这块体积内流出去了。这就是非常明显的电荷守恒了，给出了电荷和电流的关系，这个公式也叫**连续性方程**continuity equation。连续性方程在流体力学里十分重要，甚至在量子力学里的概率也遵守这个方程（电荷->概率，电流->概率流）。

13. 电势、磁向量势

我之前提到过，知道了一个向量场的散度和旋度，就能解出这个向量场。毕竟散度和旋度描述了一个向量场在空间中每一点发散和旋转的趋势，还能漏掉什么几何性质呢？（这个说法不准确，因为需要有**边界条件**boundary conditions。但是我们忽略这些细节，假定边界条件一直很棒，不必烦神）于是，可以想象，我们可以把任意一个向量场 $\mathbf{F}$ 拆成两部分，第一部分只有散度没有旋度，第二部分只有旋度没有散度，然后 $\mathbf{F}$ 就是这两个部分加在一起（这就是**亥姆霍兹定理**Helmholtz theorem）。比如静电场没有旋度，那它就只有第一部分、第二部分就为 0；磁场没有散度，则只有第二部分、第一部分为 0 。

注意到第11部分我提到，梯度的旋度总是为 0 ，而旋度的散度总是为 0 。因此，我们可以把没有旋度的第一部分写作某个标量场 $\tilde{\Phi}$ 的梯度（这样它就自动没有旋度了），然后把第二部分写作某个向量场 $\tilde{\mathbf{A}}$ 的旋度（这样它就自动没有散度了）。也就是说，我们可以把任意一个向量场 $\mathbf{F}$ 写成：

$$(15) \quad \mathbf{F} = -\nabla\tilde{\Phi} + \nabla \times \tilde{\mathbf{A}}$$

其中负号不重要只是个习惯，$\tilde{\Phi}$ 是个标量场叫做**标量势**或者就简称**势**potential， $\tilde{\mathbf{A}}$ 是个向量场叫做**向量势**vector potential。叫它们“势”的原因是我们对它微分（梯度或旋度）就可以得到我们想要的场，符合我们一般的习惯。

在引力场和静电场这种保守力场中，旋度是 0 ，于是第二部分 $\nabla \times \tilde{\mathbf{A}}$ 也为 0 ，这就是我们以前没见过“向量势” $\tilde{\mathbf{A}}$ 的原因。但是现在我们看到，在讨论更一般的场的时候，我们必须加入这个向量势 $\tilde{\mathbf{A}}$ ，来描述这个场的旋度。

在电磁学中，我们专门地把电场 $\mathbf{E}$ 的标量势 Φ 叫做**电势**electric potential，把磁场 $\mathbf{B}$ 的向量势 $\mathbf{A}$ 叫做**磁向量势**magnetic vector potential或简称**磁势**。注意到磁场没有标量势，而电场的向量势可以用磁向量势表达出来（利用法拉第定律），所以我们省了一个变量。那么根据这个定义，电场和磁场就变成了：

$$(16\text{-}1) \quad \mathbf{E} = -\nabla\Phi - \frac{\partial}{\partial t}\mathbf{A}$$
$$(16\text{-}2) \quad \mathbf{B} = \nabla \times \mathbf{A}$$

我们甚至可以把这两个式子看成是**电场和磁场的定义**，而把电势和磁向量势看作是**更基本**的物理量。

有的同学现在可能非常地不爽：电场和磁场已经够绕的了，为什么要引入这么两个新的场？还加入了好多新的微分？什么梯度旋度看得我头都晕了！现在我来列举好处：

1. 减少了自由度。电场 $\mathbf{E}$ 和磁场 $\mathbf{B}$ 都是向量场，也就意味着总共有 $3+3=6$ 个标量场的自由度。而电势 Φ 和磁势 $\mathbf{A}$ 一个是标量场一个是向量场，只有 $1+3=4$ 个自由度。这就意味着用电场和磁场来描述电磁场的状态时，有冗余的信息（用电势和磁势来描述的话，也会有冗余的信息，没办法，但能减少一些自由度总是好的）。

2. 减少了麦克斯韦方程组的方程数量。电场和磁场有更多的自由度，也就意味着必须要有更多的方程来约束它们。事实上，麦克斯韦方程组里的法拉第定律和高斯磁定律正是用来约束电磁场的自由度的，因为你看，它们俩不涉及来源（电荷与电流），只是说明了电场和磁场之间的相互关系，换句话说，就是**恒等式**identities。而使用电势和磁势之后，电场和磁场的定义（公式 (16)）直接**保证**了这两个恒等式必然成立，我们便可以把它们踢出麦克斯韦方程组，从而只剩下两个方程。

3. 可以组成有四个分量的**四维向量**four dimensional vectors，即电势加三个磁势。对相对论比较熟悉的同学可能知道，电磁学和相对论是完全相容的，而相对论里全都是四维向量（时间+三个空间、能量+三个动量、如此等等）。引入这样的形式可以让相对论和电动力学结合得更加紧密。

S1. 附录：省略掉的各种公式和定义

库仑定律：

$$(\mathrm{S1}) \quad \mathrm{d}\mathbf{a} = \frac{1}{4\pi\epsilon_0}\frac{\mathrm{d}Q\hat{\mathbf{r}}}{r^2}$$

毕奥-萨伐尔定律：

$$(\mathrm{S2}) \quad \mathrm{d}\mathbf{B} = \frac{\mu_0}{4\pi}\frac{I\mathrm{d}\mathbf{l}\times\hat{\mathbf{r}}}{r^2}$$

Nabla算符：

$$(\mathrm{S3}) \quad \nabla = \frac{\partial}{\partial x}\hat{\mathbf{x}} + \frac{\partial}{\partial y}\hat{\mathbf{y}} + \frac{\partial}{\partial z}\hat{\mathbf{z}}$$

梯度：

$$(\mathrm{S4}) \quad \nabla V = \frac{\partial V}{\partial x}\hat{\mathbf{x}} + \frac{\partial V}{\partial y}\hat{\mathbf{y}} + \frac{\partial V}{\partial z}\hat{\mathbf{z}}$$

散度：

$$\text{(S5)} \quad \nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

旋度:

$$\text{(S6)} \quad \nabla \times \mathbf{F} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\hat{\mathbf{x}} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right)\hat{\mathbf{y}} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\hat{\mathbf{z}}$$

高斯定理:

$$\text{(S7)} \quad \iint_{\partial V} \mathbf{F} \cdot \mathrm{d}\mathbf{a} = \iiint_V (\nabla \cdot \mathbf{F}) \mathrm{d}V$$

斯托克斯定理:

$$\text{(S8)} \quad \oint_{\partial S} \mathbf{F} \cdot \mathrm{d}\mathbf{a} = \iint_S (\nabla \times \mathbf{F}) \mathrm{d}\mathbf{a}$$

真空中的电磁波:

$$\text{(S9)} \quad \nabla^2 \mathbf{E} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}, \quad \nabla^2 \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{B}}{\partial t^2}$$

任意规范的势的麦克斯韦方程组:

$$\text{(S10-1)} \quad \nabla^2 \Phi + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = -\frac{\rho}{\epsilon_0}$$

$$\text{(S10-2)} \quad \Box \mathbf{A} + \nabla \left(\nabla \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial}{\partial t}\Phi\right) = \mu_0 \mathbf{J}$$

芳香性—— $4n+2$ 的背后

——芳香性的基本概念和特征

by H3PO4⑨ (from Bilibili, UID: 16076021)

凯库勒的那天晚上，到底做了什么梦呢?

- 本文为关于芳香性的简介科普性质的文章。会分为四期发出，本文为第一期，讲述芳香性的历史，基本概念和特征。
- 本文为完全原创，部分图片来自 Wiki 百科、有机化学教科书。
- 本文并不是完全的科普文，某些知识性可能较深。可以作为资料查阅，储备目的。
- 没有绚丽的实验图片，后面几期可能会有。

在有机化学中，这是一个极为极为常见的结构。就算对于化学并不是很了解的人来说，也一定在各种药物说明书，各种读物上看到过这个结构。这就是苯环。

很多人看到这个分子就会记住它的特征：标准的六元环，单键和双键构成了一个像是表情包的神秘图案。可是对这个分子本身的探索和研究，却发展出了有机化学中，一个非常非常重要的概念——芳香性。

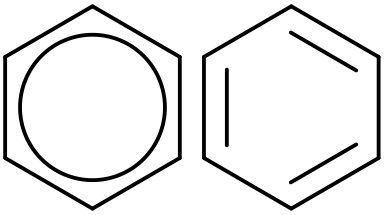


图 1　苯环的结构式，左边为了表示共轭结构写成了圆环[1.]

苯的发现和芳香性的早期定义

苯最早是**法拉第**发现的。1825 年，法拉第在裂解鲸鱼油脂的时候，发现一种无色透明的有特殊气味的液体，经过简单的元素分析，法拉第发现这种化合物只由碳元素和氢元素构成，是一种碳氢化合物，且碳原子数：氢原子数$=1$。1834 年，Mitscherlich 通过干馏安息香酸（苯甲酸）得到同样的化合物，接下来 1845 年，Hofmann 从煤焦油中发现了苯，并在接下来提出：“**苯环**”**是含有该结构的化学物质产生香味的根源，因此将苯环的这种特殊性质称为 芳香性。**从此，苯环就和芳香性这个概念不可分割的绑定在一起。

1834年，Mitscherlich 在蒸馏出苯的同时，确定了苯分子中 6 个碳和 6 个氢原子的经验式（C_6H_6）。这说明**苯环具有高度的不饱和度**（可以理解为相对于其他碳氢化合物，苯的氢原子个数显得特别的少）。但是苯却和常见的不饱和化合物不同，它的化学性质较为惰性。当时的人们发现，**苯环不容易发生加成反应，但容易发生取代反应，与苯环连接的基团具有特殊的性质。**这使得化学家们大为困惑，在那个有机化学尚不完善的时代，无数的化学家尝试提出了无数种解说来回答苯的结构的问题。

1865 年，弗里德里希·凯库勒做了那一个著名而神奇的梦，梦中梦到一条蛇咬住自己的尾巴。“据此”，凯库勒**提出了苯环单、双键交替排列的结构**，即现在所谓“凯库勒式”。为了解释苯的高度对称性，以及邻二甲苯只有一种同分异构体的现象，凯库勒提出**苯分子中的单双键交替排列，双键和单键交替不断变换。**这造成了苯分子的高度稳定性和对称性。在凯库勒之后，仍然有许多化学家提出不同的结构和猜想试图来解释苯环的结构的特殊性。但都没有从根本上揭示出苯环结构的特殊性。

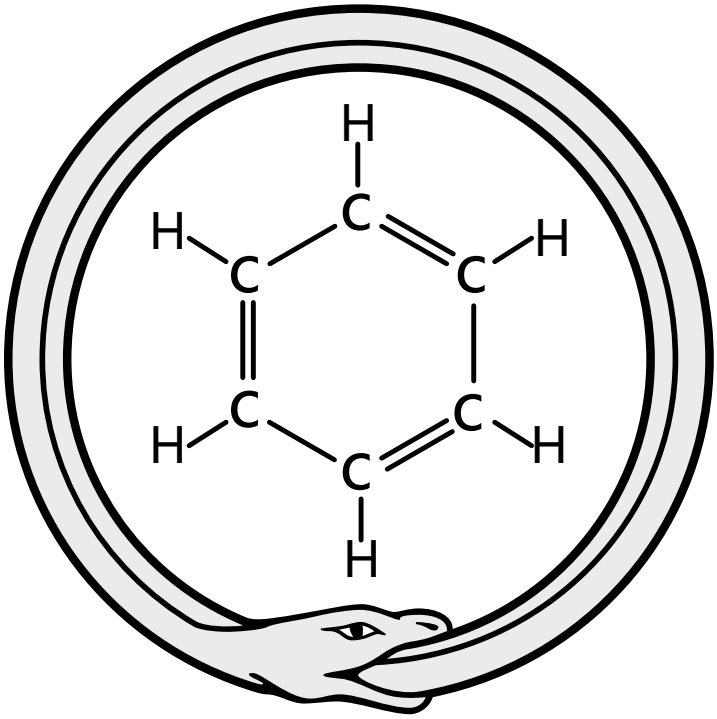


图 2　凯库勒的梦和苯环的凯库勒结构[2.]

1925 年，Robinson 提出“芳香六电子组态概念”，将问题的关键锁定在**苯环的电子数**。这之后，1931 年，休克尔提出芳香性的量子力学理论，根据原子轨道线性组合理论提出了第一个芳香性的量子力学模型，并经过简化导出了大家熟知的 $\mathbf{4n+2}$ **的芳香性判据**。这个理论成为芳香性理论发展史上的里程碑，也成为后续的芳香性理论研究的基础。

到今天为止，人们已经针对芳香性的问题进行了无数的理论研究，做了无数的实验，并合成发现了无数种新的化合物，芳香性的概念也大大的扩展和延伸。但最初关于苯环的结构式的解释在今天仍然具有重要的意义。

从轨道到大 π 键

苯有六个碳原子，在苯的分子中，苯的碳原子以 sp^2 的形式和 H 原子以及邻近的碳原子成键，在碳原子的垂直方向，剩余的一个 p 轨道电子的电子云互相平行，相互叠加。形成一个特殊的，离散在六个碳原子上方的化学键，称为：离域大 π 键。苯环因为离域大 π 键的存在产生了额外的稳定性。

量子力学认为：原子的特定原子轨道除了具有固定的形状，也具有特定的能量和特定的相位，在原子轨道相互叠加时，遵循一定的规则（对称性匹配原则；最大重叠原则；能量相近原则），这样会因为叠加方式的不同而形成多组轨道。

例如，氢分子中两个氢原子形成 σ 键时，两个氢原子的 s 轨道分别以同相位和反相位的方式重合，当轨道以同相位结合时，得到的新轨道能量低于原来的 s 轨道，这种轨道称为成键轨道；而当轨道以反相位重合时，得到的新轨道能量高于原来的 s 轨道，这种轨道称为反键轨道。来自于两个氢原子 s 轨道的两个氢原子，按照电子填充原则填入能量较低的成键轨道，这使得两个电子的能量降低，这样我们说，它们形成了化学键。

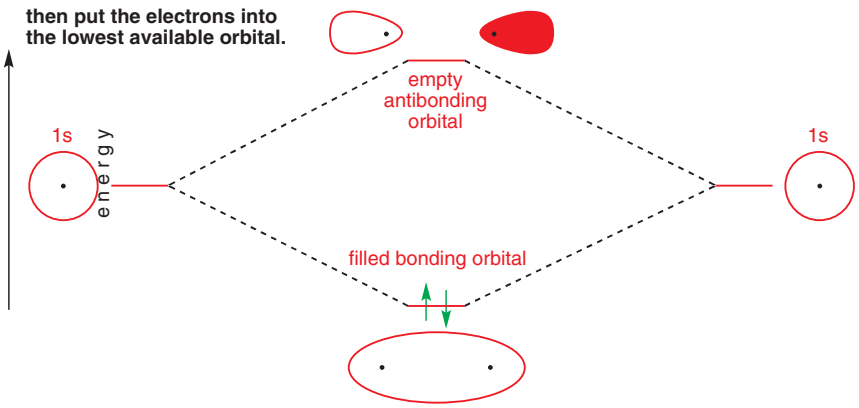


图 3　$\mathrm{H_2}$ 氢气分子的分子轨道图

苯的情况要复杂的多，苯有六个碳原子，每个碳原子都可以提供一个 p 轨道，形成的 π 键包含 6 个轨道，如下：

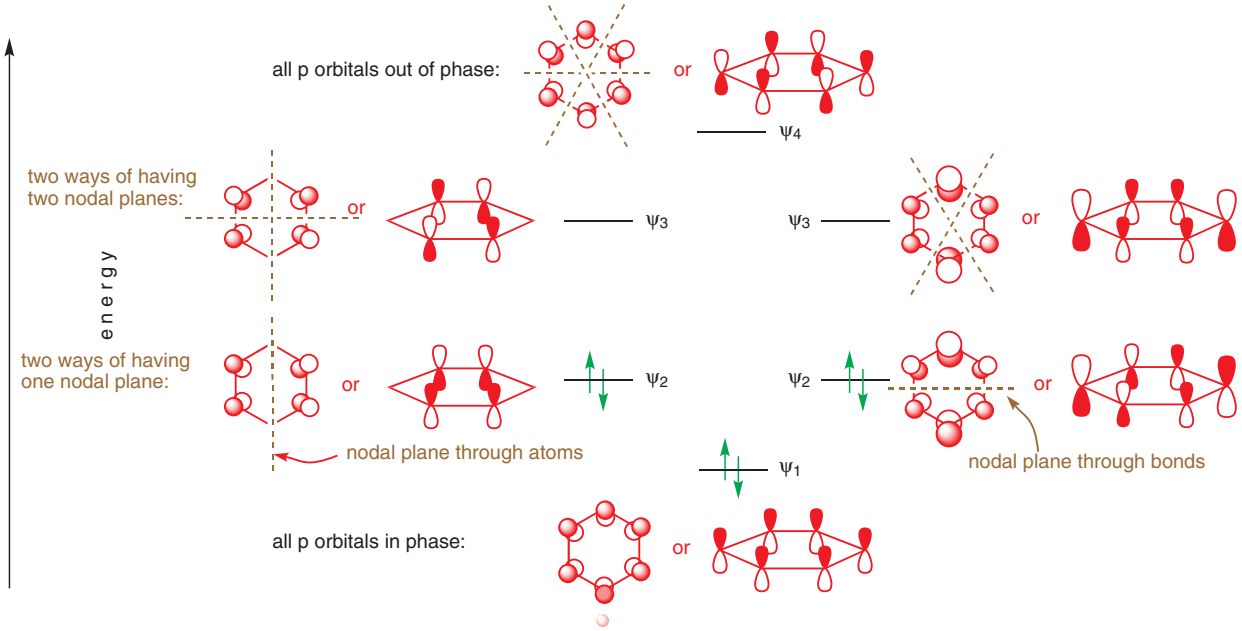


图 4　苯环的分子轨道[3.]

图中由于不同的相位对称性组合出的 6 个轨道分为 4 组：ϕ_1，ϕ_2，ϕ_3，ϕ_4，其中 ϕ_1 对称性最高，能量最低；ϕ_2 次之，ϕ_3 再次，ϕ_4 能量最高。这之后，6 个 π 电子按照轨道填充原则依次填入轨道，6 个电子刚好填满能量最低的三个轨道 ϕ_1，ϕ_2，这样形成了一种极为稳定的"刚好填满成键轨道"的结构，这也是苯分子芳香性的根源。

利用类似的方法，可以计算得到其他类似的分子的分子轨道图如下：

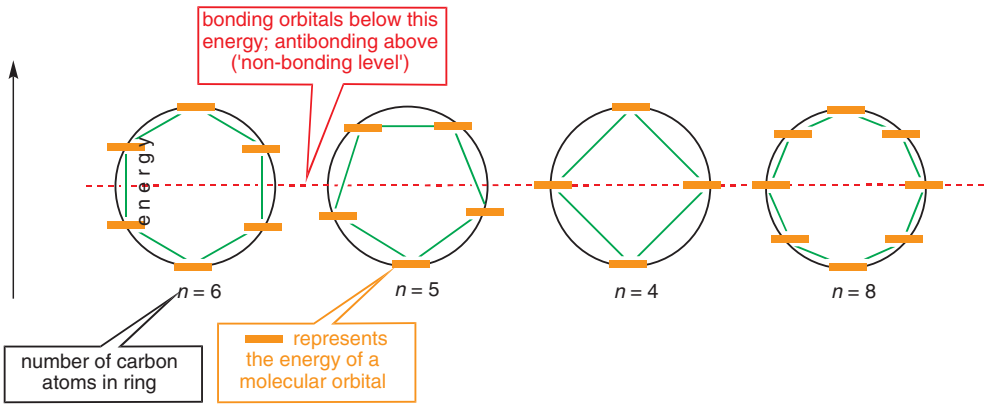


图 5　各种环状多烯烃的轨道[3.]

可以看到，要形成“刚好填满成键轨道”的化合物，电子除了要填满能量最低的成键轨道（需要 2 个电子），还要填充满所有能量较高的成键轨道或非键轨道，这样的轨道每组需要 4 个电子才能充满。因此这些分子的共同特点就是共轭电子数为 $4n+2$（n 为不小于零的整数）。

基于此，休克尔提出著名的休克尔判据：

芳香性是指，由不饱和键、孤对电子和空轨道组成的共轭系统具有特别的、仅考虑共轭时无法解释的稳定作用。

在单环 平面 共轭 体系的分子，共轭电子数为 $\mathbf{4n+2}$ 时，该分子分子具有芳香性。

这是芳香性的研究史上一个极为重要的理论结果，也是到现在为止大家所最为熟悉，应用最广泛的一种芳香性的判据。虽然在这之后，不断有新的判据被提出，新的“芳香性种类”被发现，但是休克尔判据仍然是使用最为广泛也最为基本的芳香性判断方法。

芳香化合物的特殊性

芳香性的化合物由于其电子结构的特殊性，具有许多与其他化合物不同的特征。

1. **从结构上来讲，芳香性化合物键长趋于平均化。**苯的碳碳键长度是 139 pm，六个碳碳键是均等长度的。这个数值介于碳碳单键 $\mathrm{C-C}$ 的 154 pm（1 pm=1/1000 nm）和碳碳双键 $\mathrm{C=C}$ 的 134 pm 之间，这是分子形成离域体系的重要特征。高中教科书中说的苯环形成了“**介于单键与双键之间的键**”也是说的这个。苯是最典型的芳香体系，对于其他芳香化合物，其键长不一定完全相等，但都有趋向于平均化的趋势。

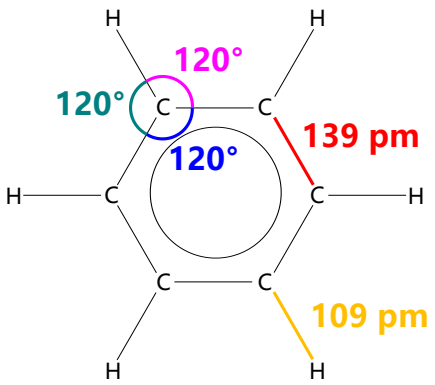


图 6　苯环的结构简式[4.]

再来看如下的分子。可以看到左边两个分子（萘[nài]和薁[ào]）不存在明显的键长交替现象，是芳香性的，右边的分子则有明显的键长交替现象，是非芳香性的。

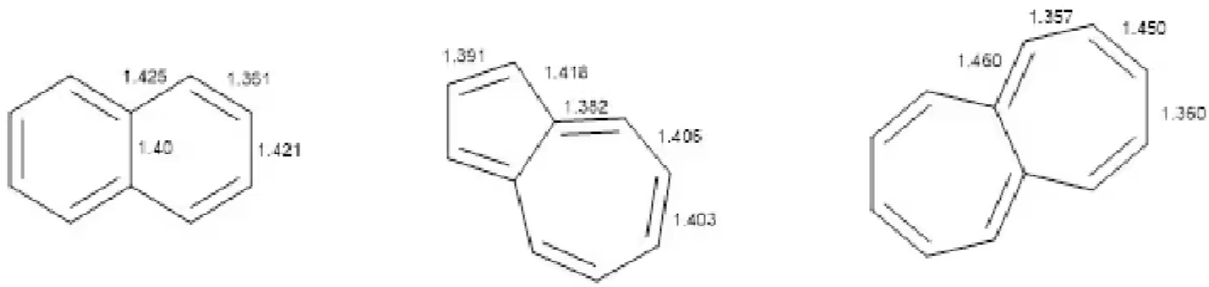


图 7　萘，薁，庚搭烯的键长（单位 $\mathring{A}$，1 Å=0.1 nm=100 pm）

2. **芳香性分子往往具有平面性的结构。**由于电子的共轭需要形成对称的轨道，因此分子具有一定平面性是必须的。

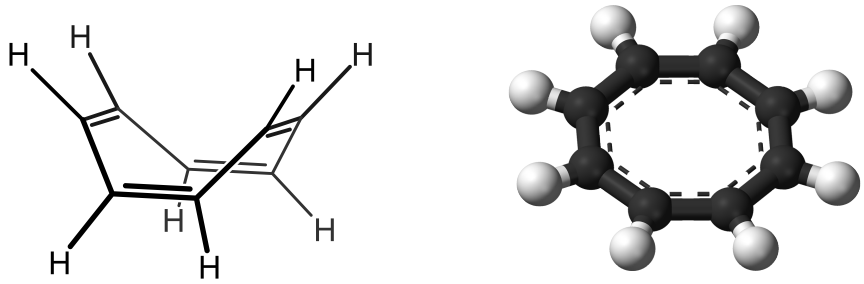


图 8　环辛四烯和负离子

环辛四烯（左图）是非芳香性的分子，它的结构上是澡盆型的非平面形式，但是当它被还原变成环辛四烯 $^{2-}$ 离子后，就变成了芳香性的离子（10 个电子，满足 $10=4\times2+2$），其结构也变成平面结构。

目前芳香环扭曲的极限是名古屋大学的Kenichiro Itami课题组合成的一种化合物，萘环扭转了 69.5°，苯环单元扭转变形达到了 35.3° <5.>。

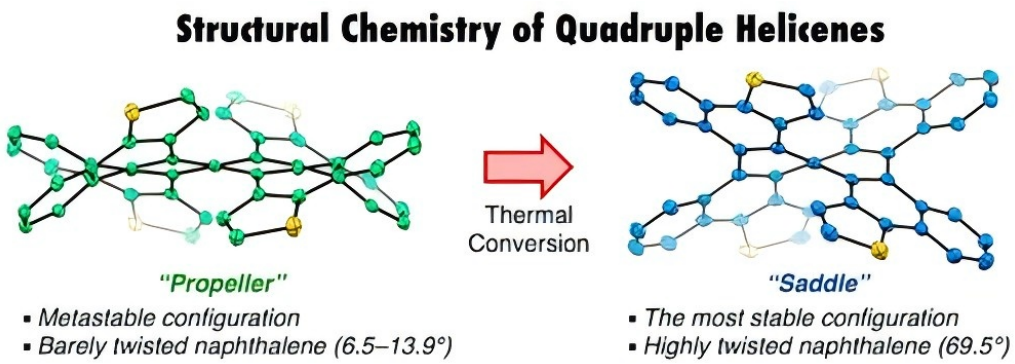


图 9　扭转最大的芳香化合物[5.]

3. **从热力学性质上来说，芳香性化合物由于其特殊的结构，可以产生特殊的稳定性**，这可以从芳香族化合物的氢化热看出。

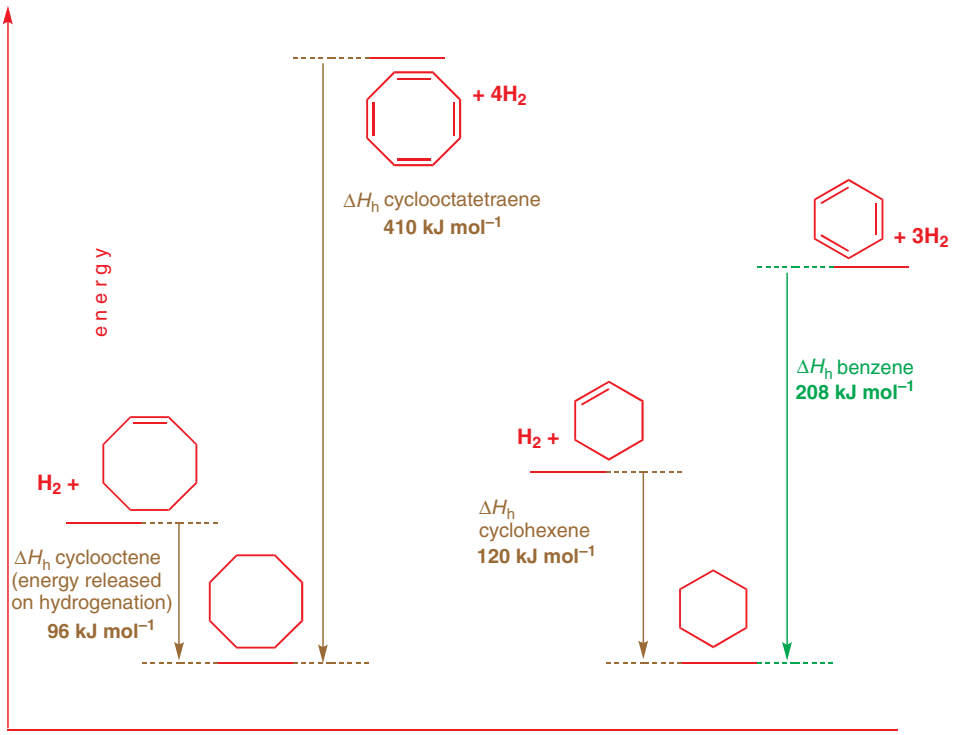


图 10　八元环和六元环的氢化热数据[3.]

如图所示，左侧的环辛烯在发生与氢气的加成反应时，每 mol 放出 96 kJ 能量，类似的，有 4 个双键的环辛四烯在氢化为环辛烷的时候，放出 410 kJ 的能量，这几乎相当于环辛烯的 4 倍。

而同样的情况不适用于苯环。每摩尔环己烯在氢化的时候放出 120 kJ 的能量，但是表观上有 3 个双键的苯环在氢化的时候，本应该放出前面的 120 的三倍也就是 360 kJ 的能量，但是事实上实验测得只有 208 kJ 的能量，这失去的差不多 150 kJ 的能量。

事实上就是当苯环形成时，由于芳香性化合物特殊的电子结构，使得整个分子稳定下来，能量已经比三个双键的情况低了。这造成了氢化热的下降。

4. **化学性质上，苯环具有与常见的不饱和化合物不同的特点。**

传统的不饱和化合物容易氧化，聚合，发生加成反应。而苯环不容易发生加成反应，反而容易发生取代反应。

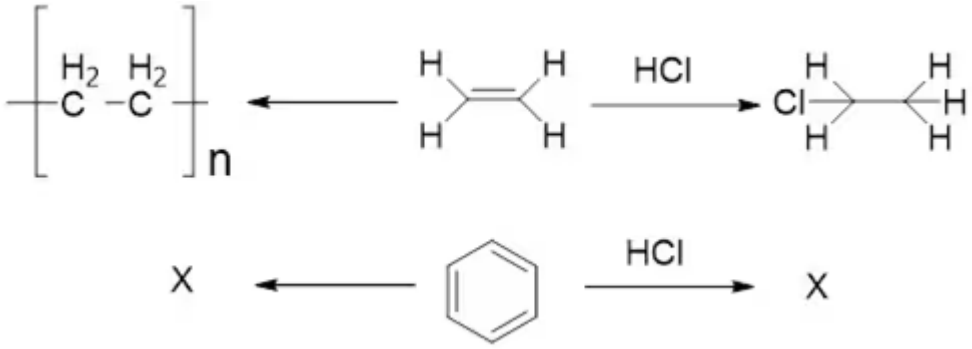


图 11　苯环和乙烯的反应

当许多基团与苯环相连时，性质会发生变化。例如羟基 $-\mathrm{OH}$ 连接烷基时（这种化合物称为醇），其 H 原子具有一定的活性（酸性），但是这种酸性是很弱的，而当羟基连接苯环（形成酚）时，其H原子的酸性会增强。

又像环戊二烯是一个碳氢化合物，但是相比于其他碳氢化合物，其结构中的 $-\mathrm{CH_2}-$ 的一个 H 原子很容易解离（可以理解为这个化合物酸性"很强"）这也是因为环戊二烯在放出 H^+ 生成环戊二烯负离子之后，能够形成了芳香性的离子而稳定下来，因此对于环戊二烯本身，具有很强的放出 H^+ 的趋势。

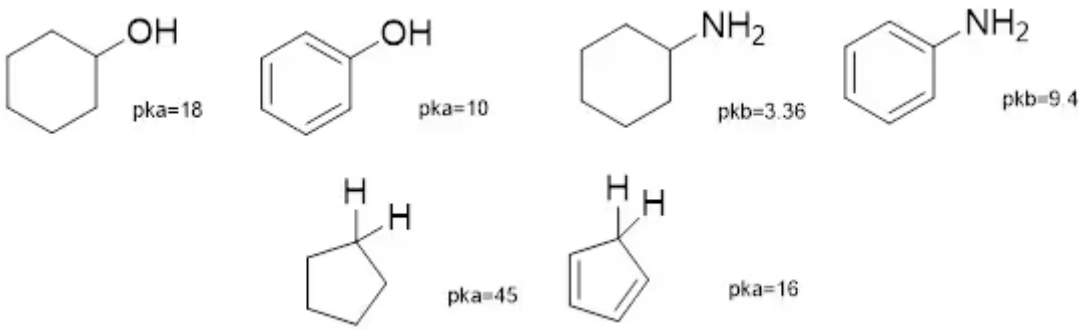


图 12　酸碱性对比（pK_{a} 越小酸性越强，pK_{b} 越小碱性越强）

5. **芳香性化合物有特殊的光谱特性。**我们以核磁共振氢谱（$^1\mathrm{H}$-NMR）来举例。

核磁共振波谱（NMR）是一种极为常用的有机化合物表征手段。在极强的磁场中，化合物中的某些特定的原子在强磁场的作用下会发生定向排列，当外界给这些原子一定的能量时，可以发生能量跃迁，这就是核磁共振检测的基本原理。

一般来说对于同一种原子，在固定的磁场中，其共振频率是相等的。但是，对于处在特定分子中的原子核，由于其电子云受到周围的原子的影响，其感受到的磁场会有所不同，这样这个原子的共振频率会与标准频率产生偏差。我们把这个偏差叫做化学位移，这是核磁分析的关键。一般来说，对于核磁共振氢谱（$^1\mathrm{H}$-NMR）我们采用一种特殊的物质叫做四甲基硅烷（TMS），这种物质中的 H 原子周围的电子云密度很高，利用这个物质来作为核磁共振检测的 0 点。将检测到的频率和 TMS 的频率对比，采用差值为最终数据。

利用核磁共振技术，**可以通过峰的数量来判定分子有几种化学环境的氢，通过化学位移的值和峰的形状可以推算出这个峰代表的原子所处的环境，对于核磁共振氢谱还可以通过峰面积的比值推算不同化学环境的氢数量的比值。**

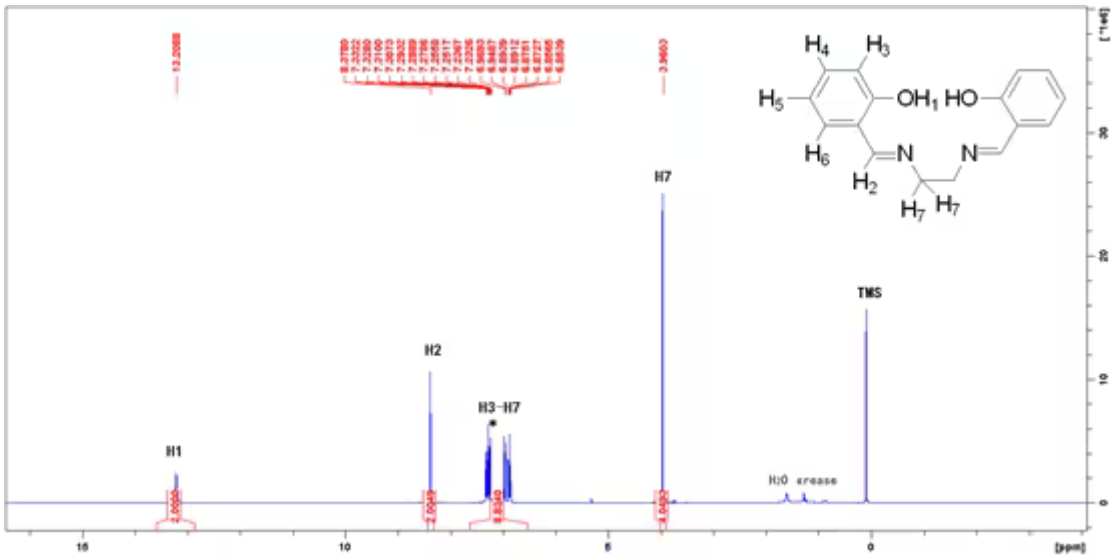


图 13　化合物 salen 的核磁共振氢谱（Bruker, $400.00\mathrm{MHz}$, $\mathrm{CDCl_3}$）

如图所示，图中的化合物叫 salen（结构式画在右上角）。salen 的核磁共振氢谱（NMR）如图，图中标明了峰的位置和峰面积，通过对以上信息的分析可以推导出 salen 的分子结构并对峰进行归属。

（salen 的酚羟基氢原子由于分子内氢键，处于极低场的位置 $\delta = 13.2$，亚胺基团的氢原子H2属于正常范围 $\delta = 8.3$，之后芳香区氢原子集中于 $\delta = 7$ 左右，亚甲基氢原子位移为$\delta = 3.9$ 左右，各个峰积分比值为 $2:2:8:4$，符合结构式。*为溶剂峰，$\delta = 1.0 - 1.5$的小峰为水峰和油脂峰）

芳香族化合物对核磁信号有特殊的响应。芳香性的化合物在强磁场中，**由于电磁感应效应会在该化合物的环中产生环电流，并由此产生一个逆于外磁场方向的感应的磁场。这个感应的磁场会使得处于环外面的氢原子的电子密度降低，屏蔽作用减少**，化学位移增高，一般来说，对于芳香化合物，**环外氢原子的化学位移处于 6.5-8.5 之间**，这是一个特征的区域与频率。而对于某些环内有氢原子的芳香族化合物，其屏蔽作用会很高，化学位移会很低。如图

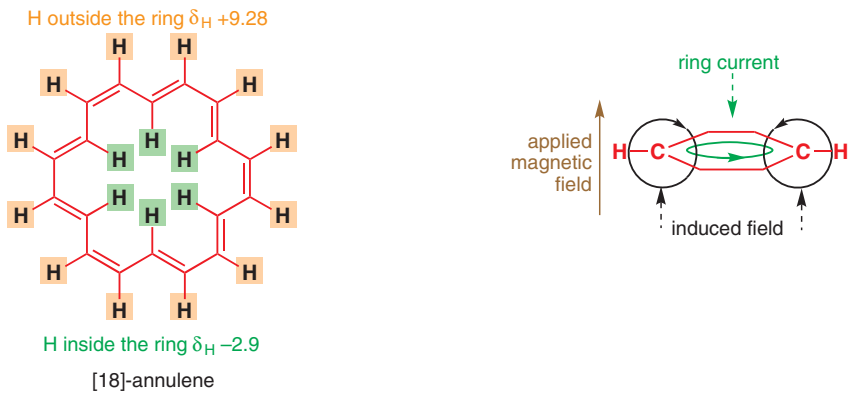


图 14　苯环的环电流[3.]

图中的化合物叫 18-轮烯，这是一个芳香性的化合物，这个化合物在环外的 12 个氢原子属于去屏蔽区，化学位移为正常的芳香性化合物的氢的范围，但是环内的氢原子由于属于屏蔽区，化学位移很低，低至不可思议的负值（-2.9）。这个化合物的核磁共振信号是对核磁共振原理和芳香性的本质的最好的阐释。

6. **理论计算方法**

人们开发了几种基于量子力学基础的能够用于判定芳香性的计算方法，可以模拟一个未知物质是否有芳香性。

例如**异构稳定化能（ISE）和核独立化学位移方法（NICS）**

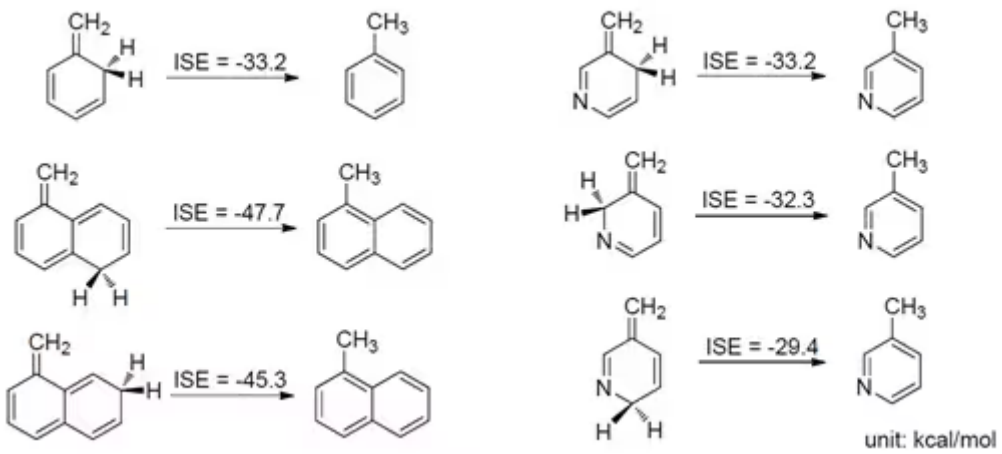


图 15　异构稳定化能[6.]

ISE是考虑如图所示的这种（亚甲基—芳香环）转化“虚拟反应”，通过量子力学软件考察这些反应的放热情况。**如果一个分子是芳香性的，那么在形成芳香性的环时势必会放出一定的能量而得到稳定，这可以从图中的数据显示出来。**

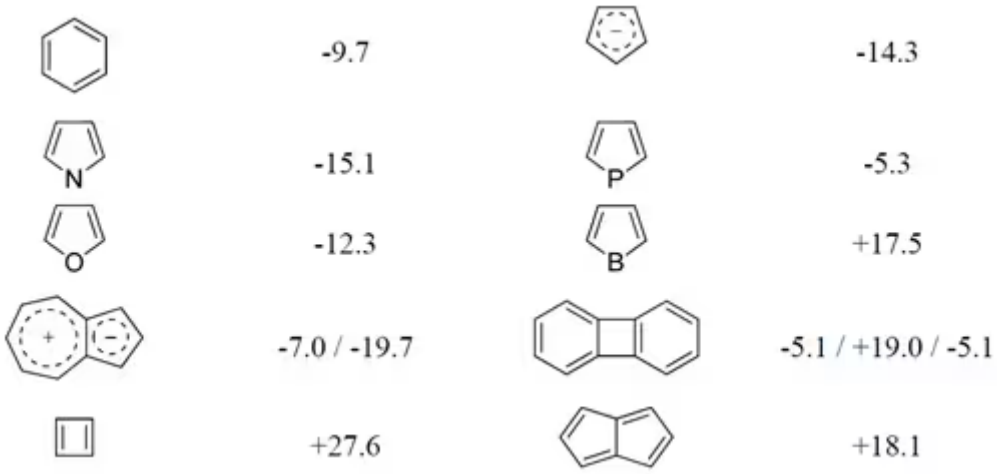


图 16　核独立化学位移[7.]

核独立化学位移是在环的中间“虚拟”一个氢原子，计算这个氢原子的化学位移，如果这个分子有芳香性，那么**由于上面讲的芳香环内部的氢原子的化学位移的原理，那么可以预计这个氢原子的化学位移会很低很低**，低到负值，所以通过计算出的 NICS 值，就可以用于判断这个分子的芳香性。

除以上的特点，芳香性还有许许多多的表现形式，包括对其连接的基团的影响等。随着有机化学的发展，也有越来越多的判据和特征被发现被提出。这些将在接下来的几个部分进行讲解。

直到现在，很多人还津津乐道凯库勒的那个梦，这个著名的“科学神话故事”不管有没有发生过，其背后一定有着多年的思考与实验基础。

芳香性，从最早的单纯表示香味，到现在与量子化学，与分子结构的紧密联系，它的发展贯穿了整个有机化学的发展史。Roald Hoffmann 说，芳香性是是体现化学终极美感的概念。或许凯库勒也没想到，他画出的六元环，能够包含着出如此深刻的内涵吧。

—— to be continued ——

参考文献

1. https://zh.wikipedia.org/wiki/%E8%8B%AF#/media/File:Benzene_both.svg

2. https://en.wikipedia.org/wiki/Aromaticity#/media/File:Ouroboros-benzene.svg

3. J. Clayden, N. Greeves, Organic Chemistry 2nd Edition.

4. https://en.wikipedia.org/wiki/Benzene#/media/File:Benzene_geometrie.svg

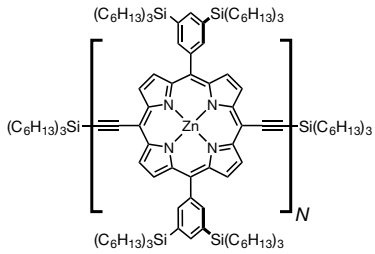
5. J. Am. Chem. Soc., 2016, 138 (10), pp. 3587–3595

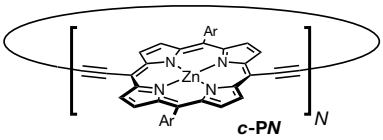
6. Org. Lett. 2002, 4(17): 2873-2876

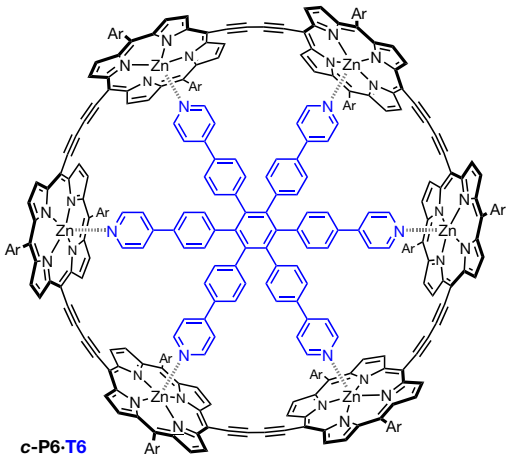
7. J. Am. Chem. Soc. 1996, 118(26):6317-6318.

以下图片（原封面图）来自

Nature, **2017**, 541, 200-203







简介计算机病毒分析与处理（第一期）

by 2511-今天校对V了吗

首先，计算机病毒是什么呢?

定义：计算机病毒（Computer Virus）在指编制者在计算机程序中插入的破坏计算机功能或者破坏数据，影响计算机使用并且能够自我复制的一组计算机指令或者程序代码。

广义：因计算机中病毒后，轻则影响机器运行速度，重则死机系统破坏，所以，病毒给用户带来很大的损失，通常在这种情况下，我们称这种具有破坏作用的程序为计算机病毒，也可就是任何以某种方式对用户、计算机网络造成破坏的恶意程序，例如：木马、蠕虫、Rootkit、广告软件、捆绑软件、间谍软件、勒索病毒等等。

总结：可以带来危害的程序。

那些著名的计算机病毒案例……

“CIH病毒” 爆发年限：1998年6月

损失估计：全球约5亿美元

“梅利莎（Melissa）” 爆发年限：1999年3月

损失估计：全球约3亿——6亿美元

“爱虫（Iloveyou）” 爆发年限：2000年

损失估计：全球超过100亿美元

“红色代码（CodeRed）” 爆发年限：2001年7月

损失估计：全球约26亿美元

“冲击波（Blaster）” 爆发年限：2003年夏季

损失估计：数百亿美元

“巨无霸（Sobig）” 爆发年限：2003年8月

损失估计：50亿——100亿美元

“MyDoom” 爆发年限：2004年1月

损失估计：百亿美元

“震荡波（Sasser）” 爆发年限：2004年4月

损失估计：5亿——10亿美元

“熊猫烧香（Nimaya）” 爆发年限：2006年

损失估计：上亿美元

“网游大盗” 爆发年限：2007年

损失估计：千万美元

“想哭”勒索病毒爆发年限：2017年

损失估计：暂时无法估价。但它袭击全球150多个国家和地区，影响领域包括政府部门、医疗服务、公共交通、邮政、通信和汽车制造业。

那么计算机病毒如何分类呢?

为了更好的了解计算机病毒，我们常从以下几点进行分类。

媒介分类

引导型病毒：指寄生在磁盘引导区或主引导区的计算机病毒。此种病毒利用系统引导时，不对主引导区的内容正确与否进行判别的缺点，在引导型系统的过程中侵入系统，驻留内存，监视系统运行，待机传染和破坏。

文件型病毒：主要通过感染计算机中的可执行文件（`.exe`）和命令文件（`.com`）。文件型病毒是对计算机的源文件进行修改，使其成为新的带毒文件。一旦计算机运行该文件就会被感染，从而达到传播的目的。

混合型病毒：指具有引导型病毒和文件型病毒寄生方式的计算机病毒

链接方式

源码型病毒：攻击高级语言编写的程序，病毒在高级语言编写的程序编译之前插入到源程序中，经编译成功后成为合法程序的一部分。

嵌入型病毒：指病毒是将自身嵌入到现有程序中，把计算机病毒的主体程序与其攻击的对象以插入的方式链接。

操作系统型病毒：用它自己的程序加入操作系统或者取代部分操作系统进行工作，具有很强的破坏力，会导致整个系统瘫痪。而且由于感染了操作系统，这种病毒在运行时，会用自己的程序片断取代操作系统的合法程序模块

攻击的系统

DOS病毒：只能在DOS环境下运行（引导型病毒不局限于DOS操作系统而存在，早期的某些单纯占用引导记录来作为病毒体的病毒，至今仍可破坏计算机硬盘引导记录），传染的计算机病毒，是最早出现的计算机病毒。

Windows病毒：指能感染Windows可执行程序并可在Windows下运行的一类病毒。

UNIX、Linux病毒：只感染Unix、Linux操作系统的病毒。

物联网病毒：针对硬件进行感染。

APP病毒：对移动APP进行感染。

功能分类

蠕虫：可以自我复制并感染其它计算机病毒。

后门：攻击者可以绕过安全认证，远程控制受感染的计算机

僵尸网络：由大量被感染的计算机组成的网络，可以发起大规模的网络攻击，例如DDOS攻击。

木马：通过特定的程序木马程序来控制另一台计算机。木马通常有两个可执行程序，一个是控制端，另一个是被控制端，与一般的病毒不同，它不会自我繁殖，也并不刻意地去感染其他文件，它通过将自身伪装吸引用户下载执行，向施种木马者提供打开被种主机的门户，使施种者可以任意毁坏、窃取被种者的文件，甚至远程操控被种主机。

下载器、启动器：用来下载和执行其它病毒的恶意代码

间谍软件：从受害者计算机上收集信息并发送给攻击者的恶意代码

广告软件：附带广告的电脑程序，以广告作为盈利来源的软件。此类软件往往会强制安装并无法卸载；在后台收集用户信息牟利，危及用户隐私；频繁弹出广告，消耗系统资源，使其运行变慢等。

捆绑软件：捆绑软件是指用户安装一个软件时，该软件会自动安装单个或多个软件。安装时静默安装，并没有告诉用户是否要安装这个软件。

Rootkit：获得计算机root权限的工具，可以用来隐藏其它计算机病毒。

勒索软件：通过加密受感染用户的文件或硬盘，索要赎金的病毒。

垃圾邮件发送病毒：感染用户计算机之后，使用系统与网络资源来发送大量的垃圾邮件，例如广告邮件、钓鱼邮件等。

攻击目标

大众型的计算机病毒：比如勒索软件，采用的是一种撒网捞鱼的方法，设计目标是感染尽可能多的机器。这类恶意代码比较普遍，恶意行为比较明显，容易被检测和防御

针对型的计算机病毒：比如特质的后门病毒，是针对特定组织而研制的。不是广泛传播的，样本收集难度大。代码非常复杂，病毒分析往往要借助于一些高分析技巧。例如震网病毒，定向攻击工业系统。

那什么是计算机病毒分析呢?

计算机病毒分析，通常是为一起计算机病毒事件的应急响应提供所需信息，包括以下的分析内容：

（系统到底发生了什么？）

定位被感染的计算机和可疑的程序；

对可疑文件进行分析；

提取出可以在系统和网络上检测病毒的特征码；

衡量并消除计算机病毒所带来的损害。

那计算机病毒分析的目标是什么?

- 分析出是如何感染的
- 分析出病毒运行流程

- 分析出病毒运行危害
- 分析出如何识别此病毒
- 分析出如何消除此病毒

那如何第一时间找到计算机病毒呢?

可以从计算机病毒特征码入手，可用于查杀病毒文件、阻止病毒传播，可以按照一下分类进行细分。

主机特征码

主机特征码关注的是病毒对系统做了什么，而不是病毒本身的文件特征，它对比反病毒软件特征码（备注：可通过技术手段修改文件信息，导致特征码失效，例如：程序加壳、使用花指令等）反而更加有效。

例如:

特定的文件创建、读取、修改行为;

特定的注册表创建、读取、修改行为;

特定的开机启动项创建、读取、修改行为;

特殊的行为，如自删除。

网络特征码

通过分析计算机病毒的网络通信数据提取出的特征码，其包括恶意的IP地址、URL、邮件、攻击数据包、计算机病毒间的通信协议等，也可以和主机特征码想结合，提供更高的检测率和更少的误报。

例如:

与公布出的钓鱼网站、放马平台进行数据交互。

说了这么多，那常见的计算机病毒分析技术都有哪些呢?

为了更好的了解，我们从以下方面进行分类讲解:

静态分析：静态分析方法是在没有计算机病毒时，对其进行分析的相关技术，其包括分析病毒可执行文件，但不查看具体CPU指令的分析技术。

常见的分析工具有CoBOT、Coverity、Klocwork、Checkmarx、Fortify、Testbed、PinPoint、C++test、VirusTotal、strings、PC-Lint、QAC、IDA pro、cutter、GDA等等。

优点：非常快速、简单、可深入分析计算机病毒。

缺点：难以分析复杂的计算机病毒，而且可能会漏掉一些重要的行为，并易受代码混淆技术的干扰，例如加壳、花指令等。

动态分析：动态分析方法则需要运行计算机病毒，使用调试器来分析一个病毒运行时刻的内部状态，动态执行每一条指令验证静态高级分析的结果。

常见的分析工具有RegShot、Process Mointor、ApateDNS、Netcat、Wireshark、INetSim、OllyDbg、x64dbg、dnspy、Sysinternals工具集等等。

优点：简单、可以缓解代码混淆的干扰。

缺点：可能会漏洞一些重要病毒功能，并只能覆盖一条计算机病毒的执行轨迹。

那有什么分析经验可以分享吗？

首先，先做好分析前的防护工作，避免病毒使用某些途径进行扩散传播在进行静态分析、在进行动态分析，互相弥补，多方面思考，也不要过于陷入细节，关注其主要的危害、在后期深入分析之前做好有一个概要性的理解。

其次，可以尝试从不同角度、不同工具、不同方法来分析计算机病毒，不要被局限。

最后，计算机实在与时俱进的，分析也要与时俱进，不要停止前行的步伐，也要走在病毒发布者的前沿，先发现危害，着手进行防御以及分析。

在上面我们了解了计算机病毒的实质、分类和一些分析病毒的方法。但是在实际的应用中，同样重要的是解决问题。当我们在计算机中发现了病毒文件或代码时，该怎样处理呢？下面展示了一些更实用的——

常见病毒处理办法！

一、感染型病毒、蠕虫病毒/X KB 快捷方式病毒 /X KB EXE病毒

清空信任区，进入带网络的安全模式使用安全软件进行全盘查杀。

在全盘查杀之前记得进入设置开启选项“发生所有类型操作都扫描”。

二、勒索病毒

（一）加密文件型勒索

通常暂时无法解密，请附上被加密文件的后缀和勒索信图片交给安全厂商以确认勒索病毒家族。

安全厂商解密的条件只有两个：

1. 病毒作者公开了密钥；
2. 病毒加密模块存在逻辑缺陷可以逻辑漏洞进行破解。

数据恢复公司可以恢复部分机械硬盘和数据库的数据。

部分数据恢复公司会联系黑客赚取差价帮你进行解密。

（二）MBRLock型勒索

请提供病毒文件以供破解。

若无法破解，你需要：

1. 另一台电脑
2. 一个空U盘

制作U盘Windows PE后，打开DiskGenius，同意用户协议，右击你的整块硬盘，选择“重建主引导记录”，选择“是”按钮。

若发现硬盘分区全没了，整块硬盘变为“未分配”状态，你还需要点击DiskGenius上面的“查找丢失分区”，弹出的窗口全部选择“保留”，随后右上角“保存”，然后“确认执行操作”。

随后，你还需要在Windows PE内修复系统引导BCD等。

（三）敲竹杠型勒索（用户锁勒索）

请提供病毒文件给安全厂商以供破解。

若无法破解，你需要：

1. 另一台电脑
2. 一个空U盘

制作U盘Windows PE后，双击桌面上面的“Windows密码破解工具”（NTPWEdit），

单机右上角的“打开”按钮，选择被锁的用户名，点击“解锁”，随后单机“保存更改”，然后重启计算机。

三、广告程序（ADWare/PUP/PUA）

用杀毒软件全盘扫描一遍（此处推荐火绒、Kaspersky Virus Remove Tool、Norton Power Eraser）

用分析工具打开以下目录：

```
C:\Users\(UserID)\Appdata\local

C:\Users\(UserID)\Appdata\Roming

C:\Users\(UserID)\Appdata\locallow
```

寻找随机英文文件夹，尤其是带有“Path”“Host”“Svc”“PDF”“Note”字样的，全部删除。

最后，再使用CnSoftKiller工具扫描处理一下即可。

四、驱动病毒（内核后门病毒/RootKit）

请尝试使用火绒恶性木马专杀工具、360急救箱强力模式扫描2遍（360急救箱请去360官网下载最新版本，是一个压缩包形式的，使用前请解压全部的文件，不建议使用360所属的产品内自带的360急救箱）。

如无法解决，请咨询对应的安全厂商安全员，并留下你的QQ，可为远程操作安全员准备好以下工具：

360急救箱

火绒剑

PCHunter/PYArk/PDArk

AutoRun

必要时可能需要Windows PE（由空U盘制作）。

如使用360急救箱尝试扫描过，请提供360急救箱同目录下的`SysRepair.log`，强制进行内核驱动对抗存在一定的风险，可能会存在有容易蓝屏的问题，请知悉。

五、U盘文件乱码问题等

多数非病毒所致。

请以管理员权限打开Cmd/Powershell/终端，

键入以下命令后回车：`Chkdsk:/ /F`

六、系统更新升级失败问题

多数非病毒所致。

请打开记事本，

内容填写：

```
Sfc /scannow

Dism /Online /Cleanup-Image /ScanHealth

Dism /Online /Cleanup-Image /CheckHealth

Dism /Online /Cleanup-image /RestoreHealth

SC config wuauserv start=auto

SC config bits start=auto

SC config cryptSvc start=auto

SC config trustedinstaller start=auto

SC config wuauserv type=share

net stop wuauserv

net stop cryptSvc

net stop bits

net stop msiserver

ren C:\Windows\SoftwareDistribution SoftwareDistribution.old

ren C:\Windows\System32\catroot2 catroot2.old
```

```
net start wuauserv

net start cryptSvc

net start bits

net start msiserver

netsh winsock reset
```

随后另存为到桌面，“文件类型”选择“全部”，文件名称后缀修改为“`.bat`”

联网状态下，右击该文件，选择“以管理员权限运行”。等待执行完成后，请重启电脑后再次尝试。

七、蓝屏问题

请找到

```
%SystemRoot%\Minidump

%SystemRoot%\Memory.dmp
```

（`%SystemRoot%`是系统当中的一个环境变量，实质`C:\Windows`（若C盘为系统盘））

随后，将这几份文件交给你的安全软件厂商或者微软社区进行分析。

如蓝屏代码为英文的Memory Management，且无法成功生成Dump文件，请优先考虑是否为内存条出现了硬件故障的问题。

八、其他系统问题

多数非病毒所致。

请以管理员权限打开Cmd或Powershell

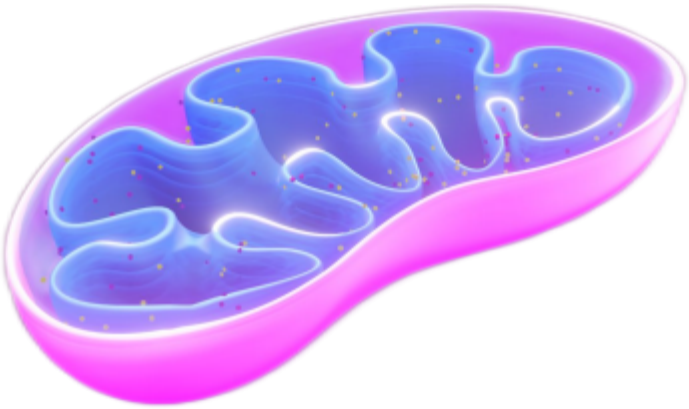
联网状态下，依次键入以下命令后回车：

```
Sfc /scannow
Dism /Online /Cleanup-Image /ScanHealth
Dism /Online /Cleanup-Image /CheckHealth
Dism /Online /Cleanup-Image /RestoreHealth
```

执行完毕后，重启电脑。

线粒体与细胞有氧呼吸

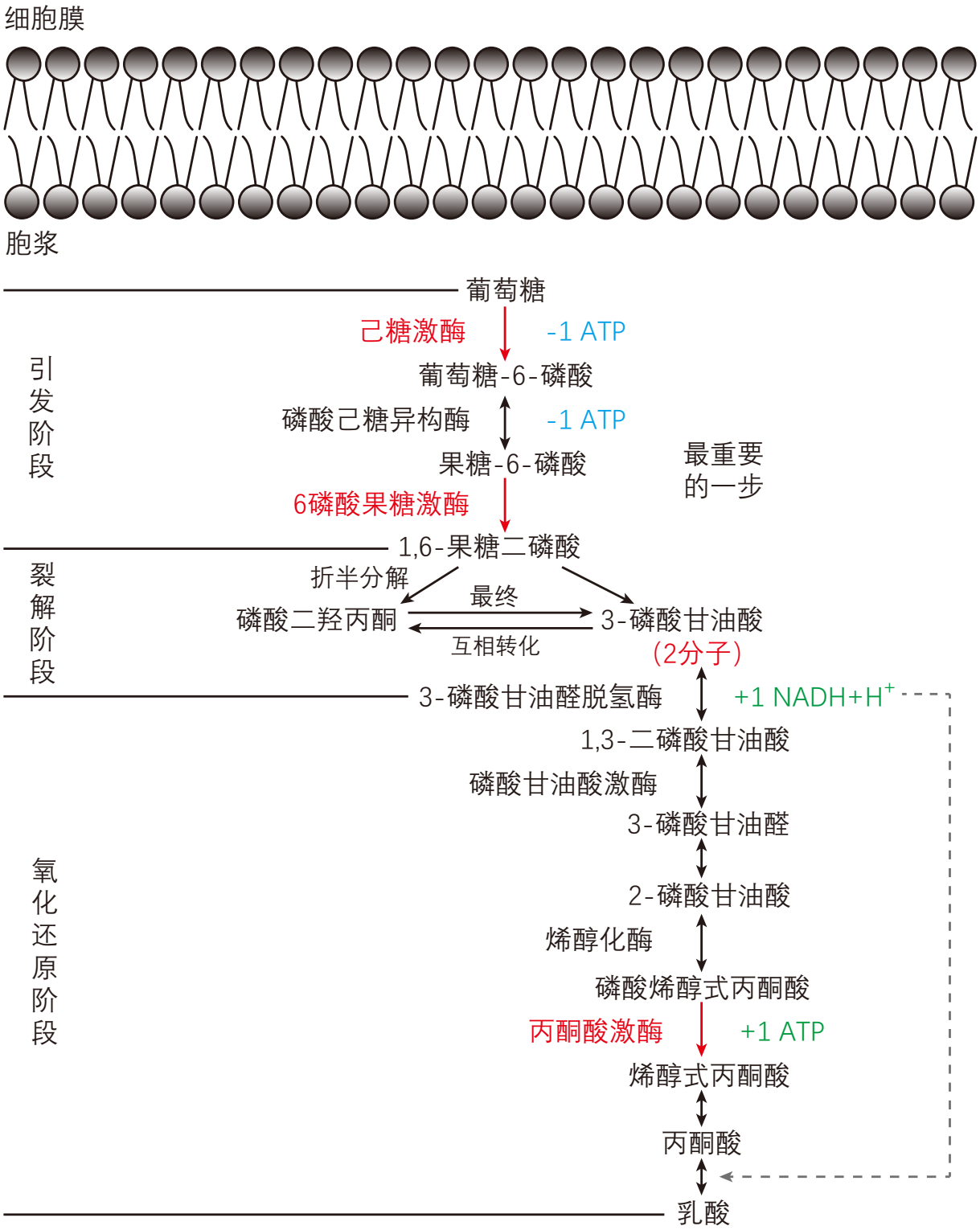
by 2516-toRefind



每当被问及“什么是生命”，我脑海中会浮现出两个答案“生命是物质、能量和信息的统一体”“生命是一个熵减奇迹”，前者体现了生命的组成，后者体现了生命存在的道理，这两句话在微观层面的细胞上体现得很清楚。下面就利用细胞的有氧呼吸让读者体会一下这种的观点。

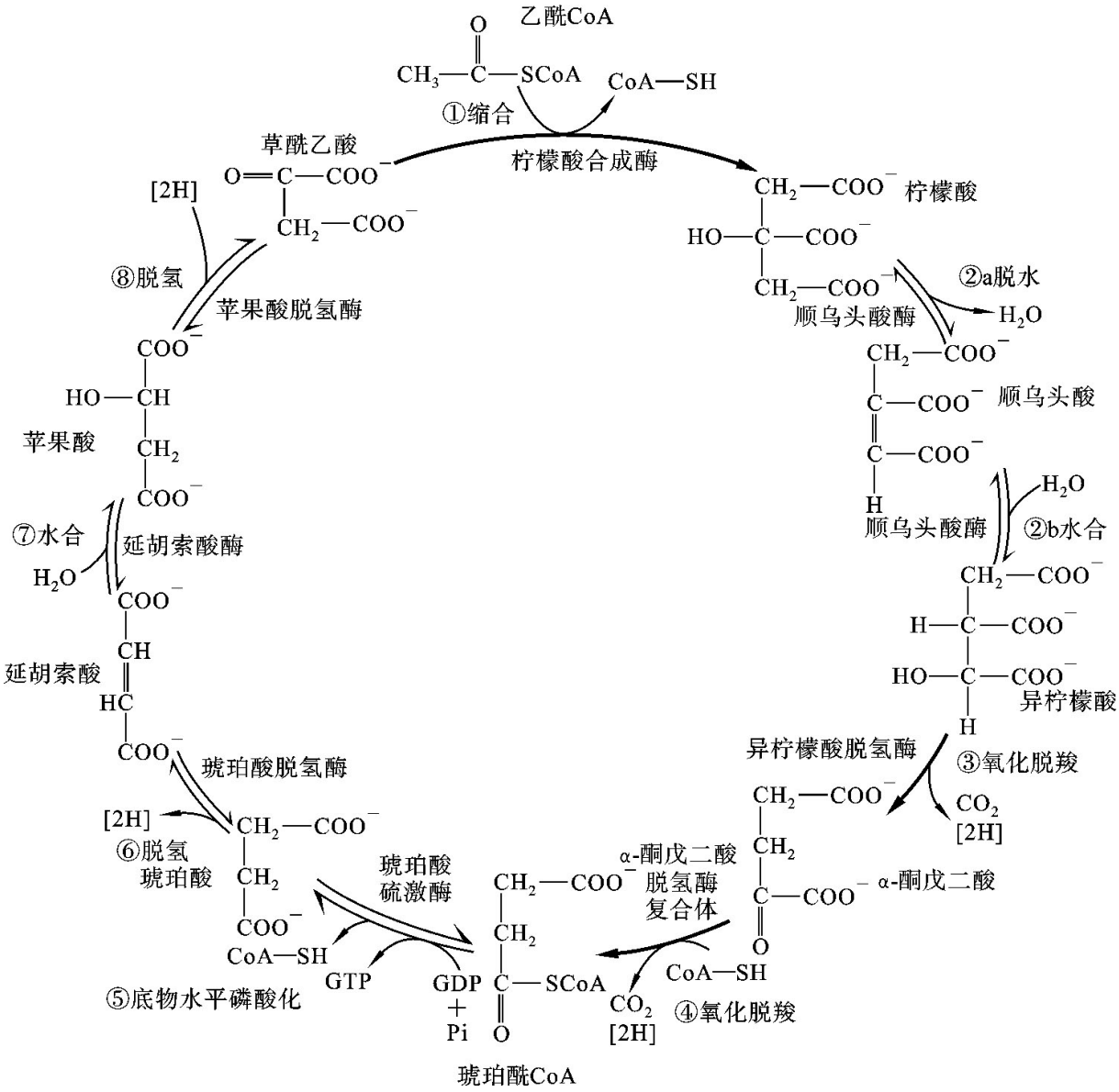
线粒体有两层膜。外膜通透性很高，所以膜间隙中的离子环境几乎与胞质相同。内膜的蛋白质/脂质质量比≥ 3：1，且具有不透性，限制了所有分子和离子的自由通过，是质子电化学梯度的建立及 ATP 合成所必需的。内膜向内折叠形成的嵴极大地增大了内膜的膜面积，如，肝细胞线粒体内膜的表面积相当于细胞质膜的 17 倍。能量需求多的细胞中的线粒体的嵴的数量比较多，如，心肌细胞或骨骼肌细胞线粒体的嵴是肝细胞的三倍。这体现了生物“结构与功能相适应”的特点，体现了物质本身性质对生物功能的重大作用——课桌的结构与功能也相适应，但这不能说明课桌是生命体。

众所周知，有氧呼吸分三阶段进行，分别是糖酵解，三羧酸循环，线粒体内膜上的电子传递。糖酵解把葡萄糖转化为丙酮酸，在细胞质基质进行。然后在线粒体基质中丙酮酸转化为乙酰 CoA，乙酰 CoA 转化为柠檬酸（含 3 个羧基 $-\mathrm{CO_2H}$ ）参与三羧酸循环。三羧酸循环产生的 NADH 在脱氢酶的作用下产生氢离子与高能电子，高能电子在线粒体内膜经过一系列化合物（电子传递链）的传递，把能量转化为膜两侧的氢离子势能差，进而通过 ATP 合酶生成 ATP。



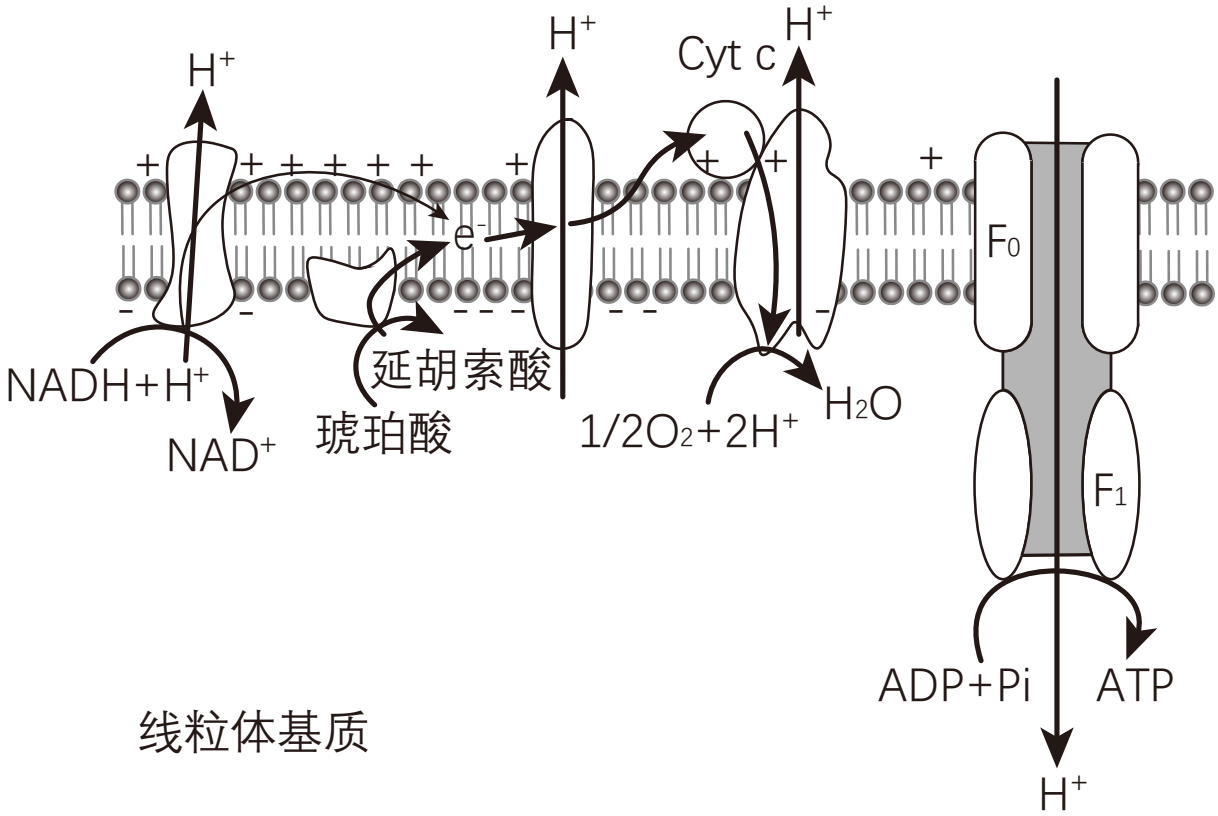
糖酵解。糖酵解在细胞质基质进行，完全不需要氧气，一分子葡萄糖在一系列磷酸化与异构化之后，生成一分子果糖-1，6-二磷酸，一分子果糖-1，6-二磷酸又可生成两分子 3-磷酸甘油醛， 3-磷

酸甘油醛经过一系列反应生成丙酮酸。全过程产生2个 ATP、2个 NADH。若为无氧呼吸，则丙酮酸在细胞质基质消耗 NADH 生成乳酸或乙醇和 CO_2。



三羧酸循环。关于三羧酸循环，朱子鹍同学曾给我们提供过一个有趣的顺口溜：“乙酰草酰成柠檬，柠檬又成 α-酮，虎酰虎酸延胡索，苹果落在草丛中。”一个葡萄糖分子在三羧酸循环中要产生 2 个 ATP、6 个 NADH、2 个 $FADH_2$（一种高能化合物）。

封面图片来自 Bilibili *@BioCraft* 上传的视频 BV1E64y1t7X5；
文中插图由作者提供，编者重新绘制 (1 & 3) 或使用图像处理工具提高分辨率 (2)。



电子传递链。呼吸链中的电子载体有严格的排列顺序和方向，按氧化还原电位从低到高排列。排在越前面的还原性越强，这样它在得到电子后可以传递给后一个电子载体，氧化还原反应放出的能量用于把 H^+ 泵出线粒体内膜，形成了 H^+ 浓度梯度， H^+ 通过 ATP 合酶进入线粒体基质，其减少的势能用于合成 ATP。

在体外氧化葡萄糖（如燃烧）所需条件较剧烈，且能量大部分都会作为热量散失。而在细胞内，葡萄糖的彻底氧化是分为许多步进行的，每一步都有专门的酶催化，使得原本葡萄糖氧化的反应活化能被拆成许多小截，而且酶本身就有降低活化能的本领，这使得细胞内的葡萄糖氧化反应条件十分温和，而且使葡萄糖的化学能得以逐渐储存入 ATP 中。这些生化过程应该就不用多说了：每一步反应都有信息的指导，包括反应物自带的信息，也包括调控分子携带的信息。

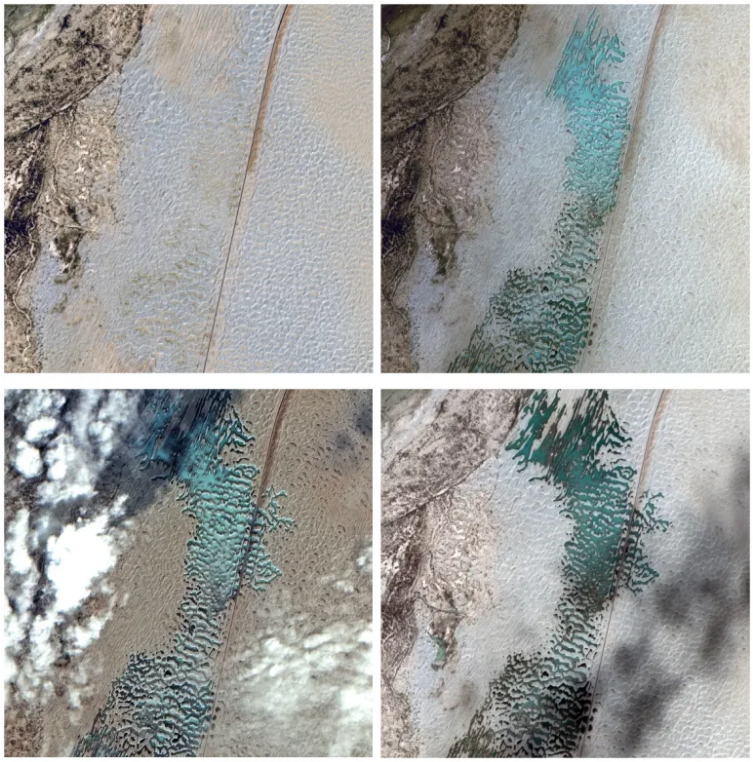
人类的线粒体基因组只编码 13 种蛋白质，约为线粒体生命活动所需蛋白质总数的 1%。但已知的线粒体病大多源于线粒体基因组编码的蛋白质缺失或缺陷，可见线粒体 DNA 突变的频率远高于细胞核 DNA。出现这种现象可能与呼吸链产生的自由基相关。正常情况下氧自由基可被线粒体中的 Mn^{2+}-SOD（超氧化物歧化酶）清除，但机体衰老及退行性疾病时 Mn^{2+}-SOD 活性降低，线粒体内氧自由基积累导致线粒体 DNA 损伤或突变。由此可见正确的（遗传）信息对于生物的正常生命活动有多重要，又可见适宜的的物质环境对生物的正常生命活动有多重要。

物质的结构决定了它具有怎么样的能量，决定了它携带什么信息。物质是能量与信息的载体。在信息的指导下，生命利用能量维持自身低熵状态，以此区别于非生命事物。这些观点应该是在任何生命现象中都有体现的，读者可以尝试着留心一下。

死亡之海变真海

——沙漠里会出现洪水?

by 2513-YaSin



起因

前一段时间，当我正愉快的在哔哩哔哩上学习，突然刷到塔克拉玛干大沙漠出现洪水了，这不由得引起了我的好奇心，“大沙漠里竟然出现洪水？”所以我上网搜集一些资料，向同学们介绍一下塔克拉玛干大沙漠，以及为什么大沙漠里会发生洪水。

气温与降水

塔克拉玛干沙漠系暖温带干旱沙漠，酷暑最高温度达 67.2℃，昼夜温差达 40℃ 以上；年最高气温为 39℃，沙面温度可达 70～80℃，而冬季最低气温低于 -20℃。

塔克拉玛干大沙漠平均年降水不超过 100 毫米，最低只有 4、5 毫米，降水量极低。从地域分布上看从西部的 38 毫米到东部的 10 毫米不等；而平均蒸发量高达 2500～3400 毫米。

塔克拉玛干沙漠地区的降水分布很不均衡，变率大，不稳定，降水集中于暖季，多雨月份的降雨又决定于一日最大降水量或某一次天气过程降水量的特点。

死亡之海

中国最古老的地理著作《禹贡》《山海经》，以及《大唐西域记》《西域水道记》都不乏对塔克拉玛干沙漠的记载，称其为流沙、大漠、瀚海。而塔克拉玛干这个名称要到20世纪才会出现。塔克拉玛干沙漠（تەكلىماكان قۇملۇقى）维吾尔传说中的被诅咒的、淹没在沙漠之下的城市，这也许与塔克拉玛干沙漠作为流动沙有千丝万缕的关系。

“塔克拉玛干”一词其实应该分做，“塔克拉”和“玛干”两部分，其中“塔克拉”意为“地下”，“下面”；“玛干”意为“家园”，“生活的地方”；因此“塔克拉玛干”意思是“地底下的城市”。

此外，塔克拉玛干还有一个威名显赫的外号：“死亡之海“，这个名字源于瑞典探险家斯文·赫定，他进入塔克拉玛干沙漠前曾说了一句自大妄为的话： “从没有哪个白人的脚触到大地的这个部分，我所到之处都是被人未曾到过的。“他俨然一副征服者的姿态，好像自己就是沙漠里的君王。但是在接下来的探险中，他的探险队几乎全军覆灭，仅剩他一人和两名助手狼狈地爬到和田河干涸的河道，那里的一泓泉水救了他们的命。从此以后，他就称塔克拉玛干为“死亡之海“。

位置

塔克拉玛干沙漠位于中国最大内陆盆地塔里木盆地的中部，介于北纬 36°50'~41°10'，东经 77°40'~88°20'。沙漠北为天山，西为帕米尔高原，南为昆仑山，东为罗布泊洼地。

整个沙漠东西长约 1000 余公里，南北宽约 400 多公里，总面积 337600 平方公里。是中国境内最大的沙漠，也是全世界第二大的流动沙漠，仅次于阿拉伯半岛的鲁卜哈利沙漠（65 万平方千米），流沙面积世界第一。

为什么塔克拉玛干会发生大洪水?

事实上，仅在 2017 年至 2021 年，塔克拉玛干沙漠便经历了 20 次洪水。其中 2021 年洪水淹水面积达到了 300 多平方公里（也就是 3 亿多平方米，约等于 1.4 万个开封高中东校区）这样看来，塔克拉玛干大沙漠好像一点也不“干”。那么，为什么大沙漠里面会出现洪水呢?

要研究塔克拉玛干的洪水，我们要从两方面来分析：水从哪里来？水到哪里去?

由此分析，洪水大概由以下原因构成:

1. **天山山脉的降雨**，由于塔克拉玛干大沙漠位于北纬 40 度左右，处于盛行西风带，西风环流携带大西洋水汽，掠过平坦的亚欧大陆，来到天山山脉，受到地形抬升作用形成地形雨（比如天山山脉上的塔里木湖就被誉为大西洋的最后一滴眼泪），然而到这里肯定会有人惊讶，大西洋到天山山脉距离六七千公里，大西洋的水气竟然可以一路保存?

事实上另一方面，当西风环流掠过亚欧大陆时也在不断的从亚欧大陆上携走水汽，随着全球变暖，原本冷干的西伯利亚冻土开始解冻，变成湿润的沼泽，为西风环流提供了大量水汽。与此同时，新疆南部的青藏高原海拔高气压低，低气压吸引了南部印度洋的水汽北上，同样带来了来自印度洋和青藏高原湖泊的水汽。

2. **冰川冰雪融水**，由于塔克拉玛干深居内陆，远离海洋，三面环山，其主要水源是雪山融水。每年的春季开始，天山，昆仑山等山脉的雪山融水流向塔里木盆地，也是塔里木河的主要水源来源。

同样地，随着全球变暖，以 2024 年为例，阿尔金山脉地区的气温较常年同期偏高 1~4 摄氏度，6 月份以来，新疆南疆地区降水量 60 毫米，较常年同期偏多三成，2024 年 8 月 17~27 日，塔里木河上游降水偏多了 4 倍以上，塔克拉玛干沙漠出现了洪水。

3.**副热带高压影响**，8 月上中旬，伊朗高压北伸东扩，脊顶伸到塔克拉玛干沙漠腹地，导致南疆出现大范围 35℃ 以上，部分地区 40℃ 以上的高温天气。塔克拉玛干沙漠周围天山、昆仑山等环绕，持续升温导致高山冰川和积雪融化，使得塔里木河流域支流来水量加大。8 月 19 日以来，南疆西部持续频繁降水，山区局地暴雨，雨水汇合流入塔里木河支流，导致塔里木河流域的多条支流形成混合型洪水。

4.**受到削弱的下渗**，塔克拉玛干沙漠还是著名的盐碱地（因此在沙漠里能够饲养咸水鱼），由于沙漠的蒸发作用强，底部的无机盐随水汽输送到表面与沙子混合，经过太阳暴晒结壳，减少了水的下渗。其次沙漠里面植株稀少，对水的拦截作用弱，因此下渗弱，地表径流得到增强。最后虽然塔克拉玛干的表面是疏松的沙子，但往下几米，由于压力的作用，沙子被挤压的十分紧密，最后犹如混凝土一般坚硬，同样不利于雨水下渗。

毫无疑问，每一条都与近年来的全球变暖有关，其中最重要的是冻土和雪山上的积雪积冰融化，为塔里木盆地带来了更多的降水，进而导致了洪水现象的发生。

然而，冻土雪山的积蓄是有限的，IPCC（政府间气候变化专门委员会）的《全球 1.5℃ 增暖特别报告》中提到，全球变暖导致冰川融水量在一定时期内增加，但在十年到百年尺度上越过“拐点”后将持续下降。这意味着，随着冰川和冰盖的进一步融化，它们对河流和湖泊的补给将减少，可能导致干旱和水资源短缺的问题，尤其是在依赖这些水源的干旱和半干旱地区，在未来，塔克拉玛干沙漠可能会变得更加干旱。

塔克拉玛干沙漠的地下水

事实上,从另一方面来说,塔克拉玛干也不“干”，在距离茫茫沙漠地表仅不到八米的地下，就存在着地下水，并且储量十分惊人，估计约有 80 多万亿立方米。

塔克拉玛干沙漠地下水的主要补给来源包括绿洲区地下水侧向补给、降水入渗以及河水入渗补给。其中，绿洲区的地下水通过侧向流动补给沙漠地下水，而降水和河流入渗也是重要的补给途径。

沙漠地下水的主要排泄方式是蒸散发，这是由于沙漠地区蒸发量巨大。此外，还有少量的人工开采。

塔克拉玛干沙漠的形成与演化

（**以下内容是从百度百科上复制的，对于地理感兴趣的同学可以仔细阅读**）

塔克拉玛干沙漠形成于 530 万年前的上新世初期，其形成演化与构造运动和古地理轮廓有关。

塔里木盆地原是古地中海的一个海湾—塔里木湾。新近纪以来，印度板块与欧亚大陆板块强烈碰撞，青藏高原和包括帕米尔高原、昆仑山、阿尔金山和天山在内的盆地周边山地隆起，塔里木海湾随同古地中海一起消失，一个深居内陆的盆地逐步形成。盆地形成后，四周的高山阻挡了外围湿润气流的进入同时，从周围山地下沉进入盆地的气流又具有焚风效应，使盆地中的空气变得十分干燥；四周高山风化侵蚀的碎屑通过河流；和洪沟向盆地中搬运堆积，为沙漠的发育提供了丰富的物质来源。

塔里木盆地的干旱环境可上溯到白垩纪，但具有现代干旱特征的沙漠环境始于新近纪末，成于第四纪初。早更新世，盆地中东部首先出现风沙堆积；中更新世晚期，开始出现大片沙漠；晚更新世，塔克拉玛干沙漠初具规模。全新世以来，青藏高原及周边山地继续抬升，加剧了塔里木盆地的干旱化。整个全新世都处于以干暖与干冷交替为特征的极端干旱荒漠气候环境，并表现为沙漠范围和面积上有盛衰的周期性变化。

结语

中国有个成语叫沧海桑田，茫茫的塔克拉玛干大沙漠，在几百万年前竟然是一片大海；中国同样有一句古诗“瀚海阑干百丈冰”不由得让人感叹奇妙的巧合。

创作本篇文章，数据来源主要来自于百度百科（塔克拉玛干大沙漠）、我看的一些哔哩哔哩视频（BV1cn4y1Z7Rk、BV1Hv411J7MW、BV1vTtsepEKf）、澎湃新闻（《新疆塔克拉玛干沙漠发洪水，会变绿洲吗？》）以及kimi智能助手。考虑到我们校刊的发行量和公益性质可能都够不成侵权，所以不发免责声明了，只将借鉴内容贴出，感兴趣的同学可自行补充了解。

《人类简史》的作者赫拉利说：“当今世界信息如此丰富，我们被信息淹没，而且其中很多是垃圾信息……我们需要信息节食。” 我所做的就是在纷杂的信息中寻找信息，提取信息，整理信息。

同时感谢地理老师唐梦玮的指导，唐老师是一个很热情，有耐心的地理老师，提出了很多宝贵的修改意见。

希望本文对同学们的地理学习有所帮助，受限于作者的知识文化水平，若有同学发现本文的数据内容，遣词造句等不甚严谨之处，请向校刊编辑部反映斧正，谢谢。