学校的理想装备

电子图书·学校专集

校园网上的最佳资源

现代科技与人文大观

数学在你身边

数学在你身边

走向数学
1.世界通用的语言
——数

数学，顾名思义，是研究数的科学。数自然而然成为数学最原始的、最基本的对象。实际上，正如一位大数学家克洛耐克所言，“ 上帝创造了自然数，其他一切都是人造的 ”。文明开始于计数，而数又使世界通向繁荣进步之路。数的概念是人类经过成千上万年才获得的抽象概念。但是同是 1、2、3、4……不同的民族却有不同的称呼和记号，长期以来并没有统一的迹象。这正如圣经的故事所说的，民众要建筑通天塔，上帝怕他们搞成，于是，搅乱他们的语言，使他们互相不能理解，从而不能同心协力建成巴贝尔塔。因此，当你跨过国界，又不通晓当地语言文字时，其困难处境可想而知。在这种情况下，你所能认识的，你所能依靠的，你所能求助的，恐怕只有 10 个印度—阿拉伯数码 1、2、3、4、5、6、7、8、9、0，以及由它们组成的千千万万种数字和符号，这是世界上唯一通用的语言，尽管数字的读法仍然还不能统一，但是电话号码和地址上的数码是一致的。

在我们走向信息社会的转折关头，数不仅是人和人交流的通用语言，而且更是人与机器交流的通用语言。至今计算机还很难懂得自然语言，你要让它工作，给它指令，那就必须运用它所能理解的机器语言，它也是由数字表示的一串符号，最后甚至转变成只有 0 和 1 的一串数列，机器的一切操作都是在它们的通用语言——数串指导下进行的。因此，把信息翻译成数码，是自动化的一个关键问题，信息社会的主要技术就是数字化。

借助于数字化，我们可以在计算机上实现各种自然界存在的和看不到的美丽的图形。例如，当前热门的浑沌理论，其千奇百怪的图形大都是在屏幕上显示的，这就是数字显示，声音虽然可以借助电流直接传送，但是数字化之后，更精确更抗干扰，这就是数字通信。当用计算机自动控制切削工具或量体裁衣，数字化帮助你更精密的控制，这就是数字控制。数是人和机器通用的语言，它不仅能够使大家相互理解，也使一切活动更为精密和准确，这就是数的威力。

2.数一数宇宙的沙粒
——大数

数的出现是靠数（shǔ）的操作。对于十几个、几十个甚至几百个上千个，人们还是能够对付，虽然在直观上不太明显。可是有许多事物是数不胜数的，最典型的例子是海中的沙粒。在《圣经》中，海中的沙粒被认为是不可数的，这也就是原始的无穷多的概念。

可是，早在公元前 3 世纪，古希腊大数学家、大科学家阿基米德就提出过异议，他专门写了一本书，书名称《计沙术》，其中写道：“有人认为沙粒是不可数的，我所说的沙粒不仅是叙拉古的和西西里岛其他地方的沙粒，而且所有地方的沙粒，不管这个地方有人还是没人居住。还有的人不认为沙粒是无穷多的，他不相信比沙粒数还大的数已经命名……但是我力图用几何的论据来证明，在我给宙希波的信中所命名的那些数里面，有的数不仅比地球上的沙粒数目还大，而且比全宇宙的沙粒数目还大。”

这样一来，人们必须来对付大数，而在位值制还没有很好建立的时代，就得给每个 10 的幂次一个特殊的名称。在这方面，印度走得最远，其中许多随佛教传到中国和日本，从个、十、百、千、万出发，又有表示大数的特殊词汇，亿、兆、京、核之外又有多种多样的表示，例如极=10^{48}（也就是 1 后面有 48 个 0），恒河沙=10^{52}，阿增祇=10^{56}，那么它=10^{60}，不可思议=10^{64}，无穷大数极=10^{58}，印度大数到此为止了。写完无穷大一方，还有无穷小一方，除了分、厘、毫、丝、忽、微之外一直到虚 $=10^{-20}$（也就是 $\frac{1}{10^{20}}$），空=10^{-21}，清=10^{-22}，净=10^{-23}，一个立方厘米只有一个分子当然够清净的。

这些对表示宇宙中数量大体上是够了。从弦论出发，宇宙的量级 80 级左右，从 10^{-40} 到 10^{40}。不过，数学家可以想象任何大的数。为此，又加上有名的两个超级大数：一个是古戈尔（googol），它等于 10^{100}，也就是 10 的后面还有 99 个零，另一个是古戈尔普莱克斯（googolplex），它等于 10^{10100}，它大得已经无法用语言来形容了，因为我们常说的天文数字比起它来真是小巫见大巫了。不过数字不管怎么大，总可以用 10 的幂来表示，因此现代人也不再操心为每个数取一个特殊的名字了。

3.驯服大数
——位值制

由于大数无穷无尽，我们必须用一种统一的办法来处理它，为此人们经历过千百年的努力。现在小学一年级学生都知道的位值制并不是一蹴而就的。例如古代希腊的大数学家如阿基米德，也没有想到“位值制”这种好方法来驯服大数，对此，18 世纪著名数学家拉普拉斯曾写道：“用十个记号表示所有的数，每个记号不但有绝对的值，而且有位置的值……这是一个深远而又重要的思想,今天看起来它如此简单,以至我们忽视了它的真正伟绩……而当我们想到它竟逃过了古代最伟大的数学家阿基米德和阿波隆尼乌斯的天才思想的关注时，我们更感到这个成就的伟大了。”显然，西方数学家在计算方面的落后正是由于他们那种复杂的记数方法及繁琐的演算过程，而当他们输入印度—阿拉伯数码以及位值制后，这一切都改变了。

但是，各国走向这条路的途经并不相同，大体上要经过如下的几个阶段：

（1）对不同的数给予不同的名称。

（2）建立统一的数基，原始人以 2、3、4 为数基，后来形成有 5 进制、10 进制、12 进制、20 进制、60 进制等，其中 60 进制来表示角度及分秒至今仍在应用。

（3）在用符号表示时，产生出基本的数码，例如，现在还在用的罗马数码，基本的数码为 I、V、X、L、C、D、M，根据加、减，重复可以组成所有数字。但是，一般人很难一下子看出，MCMXCIV 就是 1994，并且，这也不是位值制，可以说是（有序的）加法表示。

（4）建立位值制，特别是用印度—阿拉伯数码的 10 进位值制，从此，数真正成为统一的通用的语言。

有了位值制，也构成了简便表示数量级的方法以及数值计算的方法，而且还可以推广到二进制，构成计算机逻辑运算及数值运算的基础。

4.加减乘除出新数
——数系的扩大

现在我们认为分数、零、负数、无理数、虚数及复数的出现是自然的事，但在历史上每一次都经历极为激烈的斗争。因此，这些数的确不是自然的数，而是人造的数，而它们完全得到普遍的承认及使用只是近一、二百年的事。

由于数按其来源来讲是数（shǔ）出来的，所以上面的“数”本质上不是数，而是量，是出于测量以及计算的需要而引进的抽象的理想的元素，不过现在把它们都当成了数。

分数是最先出现的，从历史记载看，埃及人最早使用分数，不过“埃及分数”除$\frac{2}{3}$外都通过分子为一的分数，也就是$\frac{1}{2}$，$\frac{1}{3}$，$\frac{1}{4}$，$\frac{1}{5}$……来表示的，这样就出现一个问题，其他分数如$\frac{2}{5}$，$\frac{3}{7}$等是否可以用这些埃及单位分数来表示？原则上讲，这个问题是显然的，因为

$$\frac{2}{5}=\frac{1}{5}+\frac{1}{5}$$

可是如果要求单位分数不相同呢？这个问题就不简单了。埃及人通过列表得出$\frac{2}{n}$的单位分数表示，其中n是从5到101中的所有奇数，例如

$$\frac{2}{7}=\frac{1}{4}+\frac{1}{28}$$

$$\frac{2}{97}=\frac{1}{56}+\frac{1}{679}+\frac{1}{776}$$

$$\frac{2}{99}=\frac{1}{66}+\frac{1}{198}$$

一般的问题到13世纪由菲波那奇解决。不过一般的分数无须这么麻烦，它们成为现代的分数，即有理数。分数的一个最常用的表示是十进小数，它在很晚才得到普及。从测量及计算来看，有理数是最自然的量。

0在十进位中起着决定性作用，但0的出现有复杂的历史。长期以来，采用十进位时，只用9个数码，相当于0的位置空在那里。这样当然容易错，一直到很晚才用0填补这个空缺。但是0与其他数字的地位并不平等，特别是不能当分母，它在数学中有着独特的地位。

负数的出现各民族很不一样，东方民族很快由借贷等活动接受了负数，而西方人，甚至许多大数学家都不愿接受负数，这样减法就会遇到麻烦。因此，在数系扩大的过程中，只要使加、减、乘、除（除去0做分母）通行无阻，我们就得到了有理数的集合，在抽象代数中称为域。

在测量及计算过程中，我们碰到了另外两个量：一个是无理量，一个是虚量，只有到了100多年前，数学家才把它们严格地纳入“数”的范畴之中。

5.比比看
——等与不等

在数学发展的早期阶段，就已经抽象出数与量的观念。而许多实际问题首先涉及数量之间的关系，等与不等、大和小、宽窄、长短等问题，总之，涉及物体及数量的比较问题。

对于一个一个的人、动植物与物体，通过数（shǔ）得到数（shù），它们之间的比较可归结为数的大小的比较，而这在还不会描述大数时，已经可以通过一一对应来解决。古代有两个农民，要比一比他们的牛群中谁家的牛多，只要双方把自己的牛一头一头地赶过一个门，每次每家各过一头，如果甲方的牛和乙方的牛同时过完门，他们的牛群的牛数目相同。如一方过完了，而另一方还有，则牛的数目不同，这用数学表示，就是两家牛数 p、q 的不等式 $p<q$。

虽然正整数的比较不太难，但图形的比较就比较复杂。早期图形的研究涉及图形的形状及图形的量度，不同形状的图形较难比较，因此几何学的主要问题是它们何时恒等或全同。直观上讲，也就是它们何时可以重叠在一起。从实用上讲，更多地需要研究两个图形的长度、面积和体积是否相等。例如两块形状不同的土地是否面积相等，或者哪一块大，这就是几何量的比较问题。可通约的量可归结为有理数，它们的比较可以通过通分来解决。不可通约量的比较可以通过精密的或近似的方法来解决，与它对应的无理数理论经历了漫长的时间才严密地建立起来。

早在欧几里得时代，已经建立了数量关系的基本原理，这些在整个有限数学中都得到体现，此即欧几里得 5 个公理：

（1）与同一物相等的一些物，彼此也相等（这一条对任何等价关系都成立）。

（2）等量加等量，总量仍相等。

（3）等量减等量，余量仍相等。

（4）彼此重合的东西相等。

（5）整体大于部分。

从这些公理出发，欧几里得得出数和量的许多等式及不等式，它们构成数学运算的基础。

6.测量引起的危机
——不可通约量的出现

19 世纪德国数学家克洛耐克有句名言：“ 上帝创造了数，其他的都是人造的。 ”这种观点实际上代表了从古到今许多数学家的观点。毕达哥拉斯学派就声称“ 万物皆数 ”，但是，最终遇到了不可逾越的困难，主要问题在于是否能把几何量表示成数？

以测量物体长度为例，首先确定一个长度单位（例如尺），然后把尺放在物体上面，数一数放了多少次，在某些情况下正好放了整数次，比如说 8 次，长度自然是 8 尺，但是这种情况很少见，当测量不正好是整数个单位时，就必须把单位加以分割，以便用更小的单位来表示量。当把单位用整数分割到了一定程度，我们可以用整数个小单位来度量原来的长度，这样长度可以用分数来表示，这时我们说要求的线段与单位线段可以通约，也就是它们的长度之比是整数之比。按照毕达哥拉斯学派的观点，任何两个线段都是可通约的，这似乎是天经地义之事。

公元前 5 世纪，毕达哥拉斯学派成员希帕索斯或者本人发现不可通约量的存在，或者由于传播这一发现，被投进大海。这事件在历史上的冲击是非常大的。最早发现的不可通约量是正五边形的边长和它的对角线长之比，即 $1：\frac{\sqrt{5}-1}{2}$，但他们很容易用归谬法证明任何正方形的边长与其对角线长是不可通约量，它们的比是$1：\sqrt{2}$。

它们的证明大意如下：如果边长与对角线长度之比=a：c，而且 a 与 c 是最小整数，那么根据勾股定理

$$c^2=2a^2$$

由于 c^2 是偶数，c 也是偶数，但奇数的平方一定是奇数，且 a：c 最小，所以 a 一定是奇数

设 $c=2d$

则 $c^2=4d^2=2a^2$

$$a^2=2d^2$$

这样一来 a^2 是偶数，a 是偶数，同时 a 又是奇数，因此产生矛盾，从而 a、c 不可通约。

这样在几何量或连续量与数之间出现了一条鸿沟。后来的学者特别是公元前 4 世纪的欧多克斯引进了与数不同的量的概念，并且研究了任意量的比和比例理论，但是他不把量的比当成数。这样，他把数与几何学分离开，而把算术与几何的重点颠倒了过来，以前以算术及数为重点，而现在则认为只有几何才是严密的。而亚里士多德更进一步把数论归结成量论乃至几何的一部分，这样限制了西方数论的发展，却产生一个严密的几何学演绎体系。

7.丰富多彩的三角形
——三角形的恒等与相似

无论理论上还是实践上，三角形都是最基本的图形，平面几何学许多定理都是有关三角形的，至今还有许多新定理陆续发现。三角形各种各样，如等边三角形、等腰三角形、直角三角形、钝角三角形、锐角三角形等等。可是几何学研究的多是任意三角形的一般性质或共同性质，其中最典型的是：

（1）三角形三内角和等于 180°

（2）三角形两边之和大于第三边。

（3）三角形中大边对大角。

三角形几何学最基本的一些定理是关于两个三角形的比较，首先是两个三角形恒等，也称为全等、合同、全同，也就是两个三角形可以重叠在一起，也就是两个三角形的三边 a、b、c 以及对应的夹角 A、B、C 都能重合在一起。欧几里得几何学的一个重要成就是，我们判断两个三角形恒等而不必验证这两个三角形的三个边和三个角都对应相等，而只需知道这六个量中的三个对应相等就可以判断两三角形是否恒等，也就是两个三角形恒等的充分条件是：

（1）s、s、s（三边），两三角形三边 a、b、c 等于 a′、b′、c′。

（2）s、a、s（两边一夹角），两三角形 a、b 及其夹角 C 分别等于 a′、b′、C′。

（3）a、s、a（两角一夹边），即 A、C、B 分别等于 A′、c′、B′。

（4）a、s′、 a（两角一对边），即 A、a、B、分别等于 A′、a′、B′。

除了上面四种情况之外，还有两种情况有三个量对应相等，但并不能保证两三角形恒等。

（5）s、s、a′（两边一侧角）这时有两种情况：如果侧角对边是大边（S、S、A），也就是侧角是两角中较大者，这时两三角形仍恒等。如果侧角对边是小边（s、s、a），则这两三角形一般不恒等。

（6）a、a、a（三角对应相等），这时两三角形称为相似。显然，只要两三角形两个角相等时，第三个角一定相等，这样两角相等就是两三角形相似的充分条件。

两相似三角形也有许多共同的性质，各种对应线段的度量互成比例，其中最简单的是

a：b：c=a′：b′：c′。

8.周长与面积
——海仑公式

大家都知道，几何学来源于土地的测量。古埃及尼罗河一年泛滥一次，河水退后，土地的界限就不清楚了，需要重新丈量土地来确定哪块地是你的，哪块地是我的。因此实用几何学最主要问题是求出一块土地的面积。要知道求任意形状的一块土地面积决非易事，但是，直边形的土地最简单、最实际、也可以进行精密的测量，这是因为直边是容易画出来的，只要把一根绳绷紧就可以。因此，古代测量员被称为拉绳者。其次，任意直边形都可以分成若干三角形之和，虽然最容易求的图形面积是正方形和长方形。可是，并不是所有直边形都可以分解为正方形或长方形之和。这样一来，问题集中于任意三角形面积如何算？

小学生都知道，任意三角形的面积等于底乘高除以 2，可是，这个众所周知的公式并不实用，因为要想应用这个公式，首先得在一个很大很大的三角形中作一个直角，其次测量过程中难免要穿过土地，而这无疑有实际困难。显然，最好的方法是求出三角形三边的长度就能求出这个三角形的面积来。显然，只要知道三个点，就可以每两点拉一条绳，拉紧了之后，再用尺量一下，马上就可以得到三角形三边的长度，不管这个三角形多大，只要把绳拉直了就行，这样就得出三角形三边长度 a、b、c，三角形的周长 2s 当然是

$$2s=a+b+c$$

大约 2000 年前，希腊的数学家海仑得到一个美妙的公式，三角形的面积△

$$\triangle = \sqrt{s(s-a)(s-b)(s-c)}$$

海仑写了很多书，公式出现在他的《测地术》中。因此这个公式后来称为海仑公式，这是一个精密的公式，实际上阿基米德早已知道这个公式并给出证明。

但是，海仑并不完全是位数学家，他更像现在的工程师。数学得出公式就到此为止，但工程师还需要具体计算，不过当时开方的运算还没有精密的方法。海仑又由这个公式得出一些近似公式供实用计算之用，他的测地术一直沿用了很长时期。

中国的秦九韶（约 1202－1261）在《数学九章》（1247）中也独立给出与海仑公式等价的公式。

628 年印度婆罗摩笈多，把海仑公式推广到任意四边形，如其边长为 a、b、c、d，则其面积 A

$$A = \sqrt{(s-a)(s-b)(s-c)(s-d)}$$

其中s也是半周长，即$s = \frac{1}{2}(a+b+c+d)$，显然，当$d=0$时，四边形成为三角形，这个公式变成海仑公式。

9.经 1 周 3
——π的故事

平面几何中最重要的图形除了直边形之外就是圆了。所有的圆都相似，它们的差别在于面积 A 不同或周长 1 不同，而这归根结底是由于半径 r 或直径 d 不同。数学上一个最重要的发现是：不管大圆还是小圆，

A：r^2=常数

l：d=常数

这个常数就是圆周率π。π是几何中最重要的常数，例如三角形的常数是 180°，而 180° 用弧度表示也是π，因此在几何学乃至整个数学中离不开π这个数。

π可不是一个非常好的数，它不是一个整数也不是一个分数，即有理数，可是证明它很困难，一直到 1767 年才由兰伯特证出来。因此，它是一个无限不循环小数，这样要算出它的数值来是不可能的。因此，我们只能满足于求出它的近似值。而π的近似值的计算大大推动各国计算技术和技巧，在这方面中国古代居于领先地位。众所周知，祖冲之早在公元 5 世纪已知它的近似值精确到小数点后 6 位：

$$\frac{22}{7}<\pi<\frac{355}{113}$$

而欧洲人到 1000 多年之后才达到同样的精确度。16 世纪后期起，对于π的计算像马拉松的竞赛，纪录一个接着一个。他们的办法大体相似也就是求圆的面积，用圆内接正多边形或外切正多边形两头来逼近圆的面积，当正多边形的边数越来越多时，算出来的π值也越来越精确。

1579 年韦达用 6×2^{16}=393216 边形，π值准确到小数点后第 9 位。

1593 年罗曼用 2^{30} 边形，算准到第 15 位。

1610 年范·柯仑用 2^{62} 边形，算准到第 35 位，由于他花了毕生精力来计算，因此直到今天，在德国数学文献中还常称π为鲁道夫数，而鲁道夫是范·柯仑的名字。

另一方面，在没有计算机的情况下，这么算下去，穷数学家的一生也改进不了多少。因此，数学家开始从方法上进行改进，特别是找出π的无穷表达式：

1579 年韦达证明

$$\frac{2}{\pi}=\frac{\sqrt{2}}{2}\cdot\frac{\sqrt{2+\sqrt{2}}}{2}\cdot\frac{\sqrt{2+\sqrt{2}}}{2}\cdots\cdots$$

1650 年沃利斯证明

$$\frac{\pi}{2}=\frac{2\cdot2\cdot4\cdot4\cdot6\cdot6\cdot8\cdot8\cdots\cdots}{1\cdot3\cdot3\cdot5\cdot5\cdot7\cdot7\cdot9\cdots\cdots}$$

而在后来的计算中起作用的是用无穷级数

$$\arctan x=x-\frac{x^3}{5}+\frac{x^5}{5}-\frac{x^7}{7}+\cdots\cdots,$$

而最常用的公式是马钦公式

$$\frac{\pi}{4}=4\arctan\frac{1}{5}-\arctan\frac{1}{239}$$

他在 1706 年达到 100 位大关。而 1844 年达塞利用公式

$$\frac{\pi}{4} = \arctan\frac{1}{2} + \arctan\frac{1}{5} + \arctan\frac{1}{8}$$

算精确到 200 位。1873 年香克斯用马钦公式把π算到 707 位，这项纪录一直保持到计算机的问世。1946 年发现该π值从 528 位就错了，从此开始用计算机计算π值新一轮的竞赛。1947 年一开始就达到 808 位，1949 年超过 2000 位。

10.圆规直尺够用吗
——几何作图

从古到今，作图都是数学的重要问题，无论中外，主要的作图工具是直尺和圆规。直尺的用处是，过任意两点可以作一条直线，而且可以使这条直线在两边无限延长。圆规的用处是，过任意一点，可以作一圆通过另一点。虽然在作图过程中也用其他工具和方法,但是在 2500 多年间几何学往往只限于讨论尺规作图。

它们能做什么呢？它们的能力比初看起来要大得多，它能解决一系列难题，例如阿波隆尼斯问题，也就是作一个圆与已知三个圆相切。还有马尔法替问题，对已知三角形，求三个圆分别与三角形两边相切，而且它们彼此相切。这类问题对圆的大小及相对位置不加限制，讨论起来很复杂，完整的解答到 19 世纪才得出，它们是能用尺规作图的范例。

但是，许多问题，即便只涉及直线及圆的作图，单用圆规直尺也不能解决，尺规的这种局限性给数学开辟一个新领域，也就是判定哪些问题可用尺规解，哪些不可解，还有求可解的条件。著名的不可解问题有几何三大问题以及作正多边形问题。高斯证明，绝大多数正 n 边形不能用尺规作图，能作的正 n 边形只有当 n 满足下列条件时才行：

$n=2^{\lambda}p1\cdots pk$

其中 $\lambda \geq 0$，p1，…pk 是不同的费尔马素数，现在已知的费尔马素数只有 3、5、17、257 和 65537，可证正 7 边形，正 9 边形，正 11 边形等等不能用圆规、直尺作出。

到 19 世纪，人们又考虑单用直尺及圆规的作图问题，结果令人惊异。一方面单用圆规可以解决一切可用尺规能作的图，当然这时作直线就变成找到直线上两点即可。另一方面，单用直尺也可以解决一切可用尺规能作的图，只不过这时需要给定一圆及其圆心，这时作圆只需作出圆心及圆上一点即可。

由 19 世纪至今，几何学家也考虑其他的作图工具来解几何作图问题，其中特别值得一提的是用直角尺和圆规可解决单用直尺圆规不能解决的三等分任意角和倍立方体问题。

11.神坛加倍引出的问题
——几何三大问题

在几何学历史上，最著名的问题是所谓希腊几何三大作图问题。现在看来，这三个问题的重要性并不那么大，但在 2000 多年的发展历程中，却给数学带来许多新发现，最古老的问题是倍立方体问题，即造一个立方体，使它的体积是已知立方体的一倍，这问题又称德利安问题，这源于德利安人为了摆脱瘟疫，遵照神谕必须把他们的守护神阿波罗的立方形神坛加大一倍，这个问题受到柏拉图的重视，许多人都提出自己的解法。

另一个问题是三等分任意角，这个问题直到现在还得到许多业余爱好者的研究。第三个问题是化圆为方，也就是作一个圆，使它的面积等于给定的正方形的面积。

一谈到几何三大问题，我们必须强调，解决这些问题允许用什么工具，因为通常讲几何三大作图问题时，往往加上只许用圆规直尺的限制条件。在这种情况下，这三大问题是不可能解决的，证明这点到 19 世纪才完成。但是，如果不限制作图工具，古希腊人就已经想出许多方法。

为了解倍立方体问题，梅纳科莫斯在公元前 350 年发现了圆锥曲线，后来还有许多人用其他曲线如蔓叶线来解决，另外还有人用机械工具。

同样解三等分任意角问题，也可以用高次曲线，特别是蚌线，后来还发现圆锥曲线也可以使用。希皮阿斯的割圆曲线和阿基米德的螺线甚至可以把角任意等分。同样，也有许多机械装置可以三等分任意角。这些当然还不能令人完全满意。也有人从另外角度来考虑这个问题，也就是求近似解。1525 年大画家丢勒发明一种方法用圆规直尺近似地三等分任意角，误差不超过 1°。

三大问题之中化圆为方的问题最有兴趣也最为困难。古希腊人也用割圆曲线和阿基米德螺线解决化圆为方问题，而近似解法无非是求 π 的问题，而最令人吃惊的进展发生在几年之前。1987 年匈牙利数学家真的把圆分解成有限多块并拼成方形，当然这不是用尺规做的。

12.曲线的新家族
——圆锥曲线

欧几里得的《几何原本》只讨论了直线和圆。但是古代希腊还有一项最重要的发现直接与近代科学的发展密切相关，这就是圆锥曲线的发现。由于行星运动大都沿着椭圆轨道，对椭圆及其他圆锥曲线的性质的了解推动了牛顿力学的诞生，这是古希腊数学对人类文明的又一大贡献。

圆锥曲线在西方称为圆锥截线，这是由于生成圆锥曲线的方法是靠用平面截圆锥面而得到的。最早是用垂直于一条母线的平面来截正圆锥，如果正圆锥的顶角小于、等于或大于 90°，就分别得到椭圆、抛物线和双曲线的一支。而到公元前 3 世纪，希腊一位伟大的几何学家阿波隆尼乌斯写了《圆锥曲线论》8 卷，是古希腊研究圆锥曲线的顶峰。在书中他第一次得出圆锥曲线的一般作法、现在的命名以及许多重要性质。

首先阿波隆尼乌斯给出圆锥曲线最一般的生成方法，他采用的是一般的斜圆锥，而且可以是对顶的。同时平面也可以任意截取，这样不仅可以得两支的双曲线，也可以得出圆锥曲线所有退化情形。

其次，阿波隆尼乌斯把圆锥曲线移到平面上加以统一考虑和命名。他得出抛物线的基本性质，用现代解析几何的考虑，即

$y^2=px$

而椭圆满足

$y^2<px$

双曲线满足

$y^2>px$

其中 p 为常数。从这些方程出发，阿波隆尼乌斯命名的椭圆（ellipse）含有不足之意思双曲线（hyperbola）含有超出之意思。而抛物线（parabola）则表示不多不少正好的意思，而当把两个不等式化成等式时，就可以得到熟知的椭圆及双曲线方程

$$y^2 = px - \frac{px^2}{\alpha}$$

$$y^2 = px + \frac{px^2}{\alpha}$$

虽然古希腊人知道圆锥曲线的许多性质，但为了应用，许多最基本概念及性质到 17 世纪才明确提出。例如焦点是 1704 年由开普勒首先使用，另外还有离心率的概念。牛顿在他的《自然哲学的数学原理》中，第一次证明，过平面上五点可作出唯一一条圆锥曲线，而圆锥曲线几何性质的研究则导致射影几何学的产生。圆锥曲线的系统研究最终纳入解析几何学二次曲线论的范畴。

13.神秘的柏拉图立体
——正多面体

我们生活在三维空间，具有多姿多彩的立体无时无刻不在吸引着我们。但是，从数学上研究这就太多了。在平面几何中，我们研究基本的图形：三角形、矩形等直线形和圆等曲线。同样立体几何，我们也从直线形开始，相当于三角形的是四面体，相当于正方形的是立方体，这都是常见的，除此之外呢？

为了简单起见，我们先考虑正多面体，也就是所有的面都是全等的正多边形。由天然的晶体我们已经认识到，有正四面体、正六面体（正立方体）、正八面体。而十二面体及二十面体的准晶直到不久前才在实验室中产生出来。但是，通过人类的丰富想象，它们很早就在人类的思维中诞生。1885 年在意大利娄法山中发掘出公元前 500 年左右埃特卢斯卡人的正 12 面体玩具，而希腊人泰阿泰德也知道正 12 面体和正 20 面体。

不管怎样，首次对正多面体的系统描述是古希腊大哲学家柏拉图，他在《蒂迈欧编》把正多面体与宇宙的结构联系在一起。当时希腊人认为宇宙元素由火、气、水、土四个元素构成，于是柏拉图做出下列对应：

火——正四面体

气——正八面体

土——正六面体

水——正二十面体

开普勒发展了这种神秘思想，他直观地认为正四面体具有某种极小性质，即一定表面积，体积极小，因此代表最干的土。同样，20 面体体积极大，因此代表最湿的水。正六面体最稳定，代表地，正八面体最不稳定，因此代表易流易动的气。而正十二面体恰巧对应黄道十二宫，从而它代表宇宙普遍元素。

柏拉图和开普勒的神秘主义当然是无稽之谈，但从数学上看，有三点值得注意：一是他论述，这些多面体都可由等边三角形的一半、正方形的一半以及正五边形生成，这实际上预示着后来的组合多面体论乃至组合拓扑学。二是体积、表面积关系，这是数学上一个极重要的问题。一方面涉及体积、表面积计算问题，另一方面涉及极值问题及密堆积问题。三是正多面体只有五个，一个不多，一个不少。严格讲，这并不正确，这个定理只对正凸多面体成立，而如果不要求凸性成立，则还有四种正多面体，二种是开普勒发现的，另外两种一直到 19 世纪初才由普安索得到。

14.美丽的图形
——对称性

不论是天然的事物还是人工的产品，一个最突出、最直观的几何性质就是对称性。现实世界中，最明显的是人体的左右对称性。正如大数学家外尔所说“对称性和美紧密相连”。但是对称性在数学给出严密的定义之前，多少还是模糊的。由于对称性与美联系在一起，往往与匀称也就是比例均匀、一个整体各部分配置平衡以及和谐、优美、适中及不走极端。这些与空间的几何图形的对称性有一定距离。从数学上看真正的对称性是与某种变换或某种操作下的不变性连系在一起的。

例如一个图形具有左右对称性，那它在反射的操作下仍然重合到它本身。一个圆、一个球，在转动之下，我们仍然得到同一的圆和球，这就是转动对称性。这两种抽象对称性，我们看起来还是颇为直观的。另一种抽象对称性看来就不那么直观，这就是平移对称性。三角形的几乎所有性质都是与它在空间中的位置无关的，也就是无论我们把它移到哪里，它 的性质都保持。这么看，平移不变性是一种抽象的对称性，但它也有一种具体的背景，即三维的结晶体和二维的壁纸，显然具有直观的对称性，而这种对称性在装饰图案上到处可见。

对于这种具有对称性的几何图形，数学的问题是什么呢？

（1）两种对称的图形，它们的对称性本质上是相同还是不同的？

（2）对于各种对称性加以分类。

（3）证明这种分类是完备的，也就是具有某种对称性的图形必定属于其中之一种。

这些问题解决起来并不容易。我们考虑的出发点，就是对称操作。任何一个对称图形都有使它不变的对称操作，我们要认识它的对称性，就是把它的所有对称操作找出来。两个图形对称性相同，如果它们的所有对称操作都相同，如果能找到它们之间的不同之处，那它们显然不同。经过长期的研究，我们知道，所有的对称操作构成群。因此，刻划对称性的问题就变成求它们对称性的群的问题。而研究对称性就是找出所有的不同构的群。对于结晶体，1890 年已经找到所有 230 种空间群，而且证明它是完备的，也就是任何一种晶体必定属于其中一种。对于壁纸，不同的壁纸群的数目就要少得多，它只有 17 种，它也是完备的，也就是任何图案的群必定属于其中一种。人类历史上的建筑及装饰不计其数，装饰图形也千变万化，但其对称性不超过这 17 种。当然设计这些图案的匠人及艺术家不知道 17 种，可是他们却设计过所有的图案。在西班牙的狮子宫中，人们在装饰图案中，的确找到所有这 17 种对称性，这就是数学的威力！

15.数学的圣经
——欧几里得《几何原本》

西方数学中最有影响的著作是欧几里得的《几何原本》。自从 1482 年第一个印刷本问世以来，已经出版了 1000 多种不同版本，翻译成各种文字。除了《圣经》以外，没有任何其他著作有这么多人阅读、学习和研究，更重要的是它对数学发展的影响无与伦比。

《几何原本》原始的抄本已不复存在。现在的版本都是后人编订的，它共有 13 卷，465 个命题，讨论的主要内容大致是：

（1）平面几何 1 至 4 卷主要讨论直线形及圆的平面几何学，它们分别有 48，14，39，16 个命题。

（2）5 及 10 卷，分别讨论比例理论及无理量理论。

（3）数论，7，8，9 卷共有 102 个命题。

（4）立体几何，11，12，13 卷。

《几何原本》的重要意义在于它把以前的零散的数学结果组织成为一个伟大的体系，通过定义、公理、公设、定理、证明，把数学变成演绎与证明的科学，正是《几何原本》把数学变成脱离实际的纯粹数学。

从公元 3 世纪以来，欧几里得的《几何原本》一直就是数学家工作的楷模和典范。实际上它的影响远不止数学家。大科学家牛顿在撰写他的经典名著《自然哲学的数学原理》，大哲学家斯宾诺莎写他的《伦理学》也都是精确地依照《几何原本》那种“定义—公理—公设—定理—证明”这种严格的格式来阐述的。应该说，欧几里得的模式包含着三层内容：

（1）欧几里得的几何学内容是我们现实世界中的真理，也就是说我们生活在三维欧几里得空间当中，这个空间中的几何都可以由欧几里得几何公理及公设推出来。

（2）数学的对象仅限于服从《几何原本》中定义的并服从其公理的事物，并且，似乎这些事物的一切关系都可以从中推导出来。

（3）数学家的主要任务就是证明定理，即从公理、公设和已知的定理演绎地推出定理来。数学家与自然科学家不同，他们不做观察实验，也不用归纳得到自己的定理。

虽然欧几里得《几何原本》模式不断被批判、被突破，例如，非欧几何的引进，无穷禁戒的解除，新的公理化，但是，不能不承认《几何原本》仍然是我们考虑问题的起点，而这正是它的历史意义之所在。

16.画出来的数
——图形数

毕达哥拉斯学派常把数描绘成点或小石子，然后按点和小石子所排列的形状把数进行分类。这样，他们得到三角形数，1，3，6，10……，因为它们能用点排成正三角形。同样，他们得到正方形数，1，4，9，16……，正五边形数 1，5，12，22，……，正六边形数，1，6，15，28，……以及其他的正多角形数。这些是最简单的图形数。

由于图形具有明显的直观性，很容易得出图形的许多性质，例如

（1）三角形数，$\triangle_n$ 是自然数列的前 n 数的和，$\triangle_n$ 表示第 n 个三角形数，则 $\triangle_1=1$，$\triangle_2=1+2=3$，$\triangle_3=1+2+3=6$，$\triangle_4=1+2+3+4=10$，……，这些都由图形立刻看出来，其中 $\triangle_4=10$ 特别受到青睐，因为它的正三角形每边都是 4 个点，他们由此得出

$$\triangle_n = 1+2+\cdots\cdots+n=\frac{n(n+1)}{2}$$

（2）两个相继的三角形数的和是个正方形数，即

$\triangle_n+\triangle_n+1=\square_n+1$ 用现代的记法

$$\frac{n(n+1)}{2}+\frac{(n+1)(n+2)}{2}=(n+1)^2$$

（3）正方形数还可以如图进行曲 R 状分析，而且每个正方形数 $\square_{(n+1)}$ 都是前一个正方形数 $\square_n$ 加上一个曲 RL_{n+1} 得到。用现在写法

$$\square_{n+1}=(n+1)^2=n^2+(2n+1)=\square_n+L_{n+1}$$

由此得出一普遍规律

$$\square_n=n^2=1+3+5+\cdots\cdots+(2n-1)$$

（4）正五边形数同样可表示为等差级数的和，其首项为 1，公差为 3，即 1，$1+4$，$1+4+7$，$1+4+7+10$……，而一般项为 $\frac{1}{2}(3n^2-n)$（$3n^2-n$），而且第n个五边形数 $\frac{1}{2}$（$3n^2-n$）$=n+3\triangle_{n-1}$。

（5）正六边形数可表示为公差为 4 的等差级数和，一般项为 $2n^2-n$。

17.乘法的原子
——素数

从化学我们知道，所有物质都由原子组成，在化学变化的水平上，原子有 90 种左右，它们是构成我们物质世界的基石。对于正整数来说，我们也可以用基本的数来表示它，这些基本的数从某种意义上来讲是数的原子。我们把正整数结合在一起的运算有两种，一种是加法，一种是乘法。加法表示数的单位是 1，每一个数都可以表示成为若干个 1 的和，我们也可以说 1 是加法的原子。但是，乘法的情况较为复杂，每个正整数都可以分解为若干个因子的乘积，这些因子又能再分解成更小的因子，如此下去，有的因子再也不能分解为更小的因子了，这个数就称为素数，也称为质数。由于 1 是所有整数的因子，每个数乘上多少个 1 与不乘 1 完全一样，因此现在不把 1 算作素数，虽然 20 世纪之前，也把 1 当成素数。

这样一来，所有正整数分成三类：

1.单位数 1。

2.素数。

3.复合数，不是 1 也不是素数的正整数，因此，它至少有二个素因子。

最小的素数我们很容易定出来，

2，3，5，7，11，13，17，19，23，29，

31，37，41，43，47，53，59，61，67，71，

73，79，83，89，97……

但是，较大的数是不是素数，就不简单了。古希腊人发明了筛法，把整数中所有的 2 的倍数，3 的倍数，5 的倍数 7 的倍数……都筛掉，剩下的数就是素数了，看下面的表：

除了 2 之外，把第 2 列，第 4 列，第 6 列，第 8 列，第 10 列都去掉，它们都是 2 的倍数，然后，除 5 以外把第 5 列和第 10 列的数去掉，再把 3 的倍数去掉，如此下去，剩下的就是素数了。

原则上，一个数 n 是不是素数是可以判定的，表面上只要用比它小的素数去除它，看看是否除得尽。实际上，只要用小于$\sqrt{2}$的素数去除就行了。这样，素数就少了一大半。例如，要判断 53 是否素数，不必用比 53 小的 15 个素数一个一个去除它，而只要用比$\sqrt{53}$小的4个素数2，3，5，7去除它就行了。由于这 4 个素数都除不尽 53，因此 53 是素数。

但是，当一个数很大时，完成了这些除法也需要很多时间。因此，即使在当今有巨型计算机时代验证大素数也是极难的问题。

18 .正整数找到了归宿
——素因子唯一分解定理

除 1 以外，任何正整数不是素数就是复合数。对于素数 p，除了 1 和本身 p，没有其他的因子或因数，它没有分解因子的问题。但复合数的因子分解就比较复杂，它可以有各种各样的因子分解或乘法表示的方法，例如：

12=2×6=3×4

100=4×25×=10×10=5×20

因此，并没有唯一的因子表示的方法。但是，如果把每个因子继续分解，使得每个因子都是素数，在这种情况下，我们可以看出一种相同的因子分解或乘法表示式。例如上面的例子

12=2×6=3×4

2 和 3 是素数不能再分解了，而 6 和 4 是复合数可以再进一步分解，这样就得出

12=2×6=2×3×2

12=3×4=3×2×2

这样都不能进一步分解了，而且最后的答案除了次序之外是一样的。由于乘法满足交换律，所以乘积是相等的。同样

100=4×25=4×5×5=2×2×5×5

100=10×10=2×5×2×5

100=5×20=5×2×10=5×2×2×5

最后结果也是一样的。每一个复合数都可以按照这种方式表示成素因子的乘积，而且可以证明，这种标准分解式是唯一的。这个定理是算术或数论中最基本的定理，常常称为“算术基本定理”，如果我们把素因子按照大小顺序排列，我们就可以得出，任何正整数 n 可以唯一地表示为

$$n=p_1^{a1}\cdot p_2^{a2}\cdot p_3a_3\cdots\cdots p_k^{ak}$$

例如

$$12=2^2\cdot 3$$

$$100=2^2\cdot 5^2$$

这里也可说明为什么我们不把 1 看成素数的道理,因为要把 1 看成素数，那么在 n 的任何标准分解式前，可以乘上 1 的任何次幂。这样一来唯一性就完全被破坏了，所以 1 不被看成素数。除 1 以外，任何整数不管是素数还是复合数，都有唯一的乘法表示，这个素因子唯一分解定理成为数论的其他定理的出发点。

19.求同存异
——最大公因数和最小公倍数

任何两个正整数 a，b，如果它们不相等，那从它们的因子看，就会出现两种情况：

（1）a，b 没有公共因子，也就是除了 1 以外，a 的任何因子除不尽 b，b 的任何因子也除不尽 a，在这种情况下，我们称 a，b 互素，记作（a，b）=1。如果 a，b 都是素数，当然 a，b 互素，但是反过来可不对。例如 12 和 25 都是复合数，可是它们也互素，因为它们除 1 外没有公共因子，显然 12 的因子是 1，2，3，6，12，而 25 的因子是 1，5，25。

（2）a、b 除 1 以外有公共因子，公共因子数目有时不止一个，那么其中最大的称为最大公因数，例如 12 与 18 有公因数 2，3，6，6 最大，所以 6 是 12，18 的最大公因数，记做（12，18）=6。

求最大公因数的方法很早就有，它就是欧几里得辗转相除法。中国古代也发现类似的方法，例如求 52 与 117 的最大公因数，先用小数除大数得

117=2×52+13

再用余数除小数

52=4×13

结果除尽，因此 13 就是 117 和 52 的最大公因数。

最大公因数的一个重要性质是：

（a，b）=ax+by

总有整数解 x、y，而且对任何 x、y

（a，b）|ax+by

其中|表示（a，b）可整除 ax+by。

对于两数 a，b，存在无穷多数，它们既是 a 的倍数，也是 b 的倍数，它们称为 a，b 的公倍数，一个最典型的公倍数是 ab，在 a，b 的所有公倍数之中，存在最小的数，称为最小的公倍数，记作[a，b]。例如，12 的倍数有 24，36，48，60，72……，18 的倍数有 36，54，72……，显然，36，72……是 12 和 18 的公倍数，其中 36 是最小公倍数即

[12，18]=36

对于任意 a，b，有下面重要关系式

[a，b]（a，b）=ab

对于 3 个或多个正整数，也可以定义它们的最大公因数和最小公倍数，如

（a，b，c）=（（a，b），c）

[a，b，c]=[[a，b]，c]

然后用归纳法定义即可。

20.素数知多少
——素数定理

欧几里得已经证明，素数有无穷多，其后有不少新证明其中欧拉及库默尔的证明均极为简单巧妙,但是某个数 x 以下的素数个数π(x)等于多少呢？

通过人手及计算机的计算得出：

π（10）=4

π（10^2）=25

π（10^3）=168

π（10^4）=1229

π（10^5）=9592

π（10^6）=78498

π（10^7）=664579

π（10^8）=5761455

π（10^9）=50847534

π（10^{10}）=455052512

π（10^{11}）=41180548130

π（10^{12}）=37607912018

π（10^{13}）=346065536839

π（10^{14}）=3204941750802

π（10^{15}）=29844570422669

π（10^{16}）=279238341033925

18 世纪末，勒让德及高斯根据自己的计算以及当时的素数表，估计不大于某实数 x 的素数的数目即π（x）的大小。他们估计

π（x）～$\frac{x}{\log x}$，这通常称为素数定理。这里的log是以e为底的自然对数，～表示当 x→∞两端渐近相等。素数定理被认为是解析数论中最重要的定理，由它出发发展了解析数论一系列重要工具和问题。首先证明素数定理的是法国数学家阿达马和比利时数学家瓦莱—布桑在 1896 年独立得出的,他们用的都是黎曼引进的函数 S（x）。其后素数定理又有许多证明，用的都是分析工具。不用分析令人惊叹的初等证明一直到 1949 年才由挪威数学家塞尔伯格以及匈牙利数学家埃多什各自独立得到的。其后又有许多改进的证明。

1837 年狄利克雷还证明欧拉的猜想，任等差级数 an+q 中（a，q 互素，n=1，2，3……）存在无穷多素数，而且也有类似的分布。

21.素数中的孪生兄弟
——双生素数

哥德巴赫猜想现在已经是尽人皆知的了。但是，哥德巴赫猜想还有一个姐妹问题，它和哥德巴赫猜想一样困难，一样重要，这就是双生素数问题。双生素数也叫孪生素数，是说两个素数靠得很近，好像一对双胞胎一样。素数中间，只有第一个素数 2 是偶数，从 3 起的素数都是奇素数，除了 2 和 3 相差为 1 之外，剩下的素数中，两个素数相差都是偶数，也就是说最少差 2，所以我们管两个相差为 2 的一对素数叫作双生素数。例如（3，5），（5，7），（11，13）……都是双生素数。双生素数问题就是问：双生素数有多少对，是有限多对还是无穷多对？这问题讲起来很容易，但它确实是现代数论中还未解决的著名难题之一。

一开始我们就看见好几对双生素数，比如（3，5），（5，7），（11，13），（17，19），（29，31），（41，43），（59，61），（71，73），100 以内的素数就有这么 8 对，而 500 以内只有 23 对，1000 以内有 34 对，2000 以内有 60 对，3000 以内有 80 对，4000 以内有 102 对，5000 以内才到 124 对，所以增长不规则而且缓慢。

要想证明双生素数有限多还是无穷多，欧拉证明素数无穷多的方法对我们有所启发，如果素数有限多，则$\sum_{p}\frac{1}{p}$有限（Σ表示对所有素数p的倒数求和），反之，当p过所有素数，无穷级数$\sum_{p}\frac{1}{p}$发散，则素数有无穷多，欧拉的确证明了后者。欧拉的证明启发我们去验证$\sum_{p^*}\frac{1}{p^*}$是否收敛或发散。这里p^*表示一对双生素数中较大的一个。如果这级数发散，就可以知道双生素数是无穷多对。遗憾的是，布隆应用他的筛法证明了$\sum_{p^*}\frac{1}{p^*}$收敛，所以双生素数是有限多还是无穷多仍然没有解决，不过这也是数论上的一大成就，更重要是他引进的筛法对其他数学问题，比如说哥德巴赫问题，也有极大的推动。

布隆定理并没有告诉我们双生素数是有限还是无穷，但是说明双生素数非常稀少，所以要想证明双生素数有无穷多对，还是十分困难的事，它困难的程度不下于证明哥德巴赫猜想，也须要数论的方法有本质的改进才行。

与双生素数问题有关的是，相邻素数之差的分布问题。有意思的是，相邻素数的差的分布并不很规则，时大时小，特别是相差的偶数并不是按顺序出现的，出现了 2，4，6，8 之后，跟着并没有出现 10，12，而是突然出现了 14，后来才出现 10。后来的出现 18，20，22，24 之后，16 才姗姗来迟。具体我们看下表：

表中的 n 表示最小的自然数使得 $d_n=2k$

2k	n	Pn	Pn+1
2	2	3	5
4	4	7	11
6	9	23	29
8	24	89	97

10	34	139	149
12	46	199	211
14	30	113	127
16	282	1831	1847
18	99	523	541
20	154	887	907
22	189	1129	1151
24	263	1669	1693

那么，对于每个偶数 2k，是否都存在一对相邻素数 P_n、P_{n+1}，使得 $2k=P_{n+1}-P_n=d_n$ 呢？这也是一大难题。只有小的偶数，我们还的确能找出相邻素数，使它们的差为给定的偶数，比如相差为 100 的一对素数是 396733 和 396833。

22.素数的近亲
——伪素数

数论的定理与大素数的判定都要求对素数的性质进行深入研究。素数最基本的性质是费尔马小定理，费尔马小定理给出素数的必要条件，也就是如果 p 是素数，a 是与 p 互素的任何自然数，则

$$a^{p-1}\equiv 1 \pmod{p}$$。

它的意思是说 $a^{p-1}-1$ 可以被 p 整除，或者 $a^{p}-a$ 可被 p 整除。

长期以来，人们似乎认为，费尔马小定理的逆定理也成立，甚至错误地认为

n 是素数当且仅当 $2^{n}\equiv 2 \pmod{n}$

欧洲的一些人甚至把这个错误归诸于中国，这是完全荒谬的，因为中国古算中并没有素数的概念。首先举出费尔马小定理的逆定理具有反例的是萨吕斯。他在 1819 年证明

$$2^{341}\equiv 2 \pmod{341}$$

可是 341=11×31 是复合数，这一下子引起许多人的兴趣，接着又发现另外一些数具有这种性质，如 561，645，1105，1387，1729，1905 等，后来，我们就把满足同余式

$$2^{n-1}\equiv 1 \pmod{n}$$

的复合数称为 2 基伪素数，之所以说它是伪，是因为它是复合数，当然素数也满足这个条件。我们特别声明 2 基，是因为我们还没有考虑其他的基，而且 2 在计算方面是最简单的。有时我们省去 2 基，简称为伪素数。

显然，每个（2 基）伪素数 n 是奇数，反之，每奇复合数满足

$$2^{n-1}\equiv 1 \pmod{n}$$

是伪素数，而且由于奇素数也满足上述同余式，因此，如果对于 n

$$2^{n-1}\not\equiv 1 \pmod{n}$$

则 n 是复合数，这是我们素数判定的第一步。

我们仿照素数研究伪素数的类似的问题,我们问,伪素数是否有无穷多？1903 年马洛首先证明，答案是肯定的。此外，还可以定义偶伪素数，即满足

$$2_{n}\equiv 2 \pmod{n}$$

的偶数。1950 年莱默尔证明，最小的偶伪素数是 161038=2×73×1103，次年贝格证明，偶伪素数也无穷多。

我们很容易把伪素数推广成 a 基伪素数，复合数 n，对于 1＜a＜n，满足

$$a^{n-1}\equiv 1 \pmod{n}$$

例如最小的 3 基伪素数是 91=7×13，但它不是 2 基伪素数，也不是 5 基伪素数。最小的 5 基伪素数是 217=7×31。。基伪素数也是无穷多个。不同 a 基伪素数的集合都有交集，也都不重合，也就是存在复合奇数，

1105=5×13×17

它是 2 基伪素数，又是 3 基伪素数，但不是 5 基伪素数，如此等等。

最后就产生一个自然的问题，它真正满足费尔马小定理的逆定理，即这样的复合数存在不存在 n，对于每个 a，1＜a＜n 且 a 与 n 互素，它都是 a 基伪素数。这个问题一直到 1912 年才解决。

美国数学家卡迈克尔在 1912 年举出第一个这样的数

561=3×17×11

由于条件苛刻，卡迈克尔数极为稀少，到 80 年代中发现最大的卡迈克尔数只有 1000 多位，因此，自然出现这样一个问题：

卡迈克尔数的数目是有限的还是无穷多？这问题是如此之难，以至于 1990 年，连一个极弱的问题都没有解决：

同时是 2 基伪素数和 3 基伪素数的复合数的数目是否无穷多？

数学经常以不可思议的步伐前进，它往往走了几步之后，就一下子完全解决了，1994 年三位数学家证明：卡迈克尔数有无穷多！

23.整数的乘法分析
——费尔马数的因子分解

虽然我们原则上对任何正整数可以判定它是不是素数，而且如果不是素数，可以把它的因子写出来，但是实行起来却决不简单，对于几百位的数，用手算算一辈子也算不完，用最快的计算机也不行。因此，我们现在所知道的最大素数和最大因子分解也都是特殊形状的数，其中最有趣的数是费尔马数。

费尔马数 Fn 是形如

$2^{2n}+1$ n=0，1，2，3……

的数，当 n 小时，

$F_0=3$ $F_1=5$ $F_2=17$ $F_3=257$ $F_4=65537$

均为素数。费尔马在 1640 年给帕斯卡的信中，要求他证明，所有费尔马数都是素数，这个错误的猜想很快就被欧拉在 1732 年推翻，他成功地把 F_5 分解成因子

$F_5=2^{32}+1=641\times6700417$

其后不但没有证明任何一个 F_n（n≥5）是素数，而且对许多 n，具体的把因子甚至素因子求出来。时至今日，这个问题还在继续用最快的电子计算机进行研究，得出各种不同程度的结果：

最好的结果是给出 F_n 的完整的素因子分解

最次是找出 Fn 一个或几个素因子

再次是证明 Fn 是个复合数，但还找不出它的素因子。

到 1995 年初，已经完成的费尔马数的素因子分解的有 F_5，F_6，F_7，F8，F9 和 F11。F10 成为尚未分解的最小的费尔马数，但已知 F_8 有 8 位和 10 位的素因子，余下的 291 位的数现在还不知是素数或复合数。已经证明，F_{14} 是复合数，但现在一个因子尚未找到。另外已经知道有 130 个费尔马素数的 160 多个素因子，大部分 n 为二位或三位数，也有 10 几个 n 为 4 位数以及 3 个 n 为 5 位数，即 F_{18749}，F_{23288}，F_{23471}。

因此，至今围绕着费尔马数仍有许多极为困难的问题有待解决：

（1）费尔马数中只有有限多个素数还是有无穷多个？

（2）费尔马数是否不存在平方因子？

费尔马数在数学中有重要意义，首先它在用尺规作图中不可缺少，而且通过数论变换对于信号处理有重要应用。

24.十全十美的数
——完全数和马桑素数

完全数概念也来自毕达哥拉斯，它是指该数 N 等于其真因子之和的数，或者说等于其所有因子之和 σ（N）的 2 倍的数，即 N=2 σ（N）。如 6=1+2+3，28=1+2+4+7+14，是最早知道的两个完全数。关于完全数古时只知道下述定理，欧几里得《几何原本》第九篇中证明，如 2^n-1 是素数，则 $2^{n-1}(2^n-1)$ 是完全数，这种完全数自然是偶数。笛卡儿声称逆定理也成立，但 1849 年欧拉的一篇遗著发表，才首次披露其证明，即偶完全数都可以表为 $2^{n-1}(2^n-1)$ 型的数，其中 2^n-1 为素数。2^n-1 型的素数被称为马桑素数，因他首先特别注意研究这类素数（如 2^n-1 为素数，则 n 必为素数 q，该马桑数记作 Mq）。1640 年马桑已知 M_{11} 复合数（$M_{11}=23\times89$）并指出当 q=13，17，19，31，67，127，257 时，Mq 是素数。但 M_{67}，M_{257} 不是素数（前者在 1876 年证明，后者在 1927 年证明），而且在 257 以内漏了 61，89，107。

对大部分素数 q，Mq 是复合数。欧拉在 1750 年宣布，拉格朗日在 1755 年证明下列定理：

如 q 是 4n+3 型（n≥1）素数，则 2q+1 整除 Mq 当且仅当 2q+1 是素数，从而 Mq 是复合数。不过有趣的是，现在也不知道马桑复合数是否无穷多？因此搜寻完全数的问题归结成下面两个至今还未解决的问题；

①马桑素数是否有无穷多？

②奇完全数是否存在？

一般猜想奇完全数不存在，这激起许多数学家进行研究，到现在已知奇完全数许多必要条件，其中有：

（1）奇完全数的形态：德国数学家史特恩 1886 年证明，4k+3 型完全数不存在，从而要有奇完全数必是 4k+1 型。1953 年有人更进而证明奇完全数具有 12k+1 或 36k+9 型。

（2）奇完全数 N 的因子数目：1975 年有人宣布奇完全数不同素因于数目≥8，1983 年又证明如 3 不是奇完全数的素因子，则不同素因子数目≥11。

（3）奇完全数的最大素因子。1975 年证明最大素因子＞10^6，实际上，1955 年宣称最大素因子＞10^8 但未发表。由这些必要条件可以看出奇完全数是难得存在的。

通过计算机的搜索，发现奇完全数 n 要是存在一定相当大，下面是历年的记录：

1957	n＞10 20
1973	n＞10 50
1980	n＞10 100
1988	n＞10 160
1989	n＞10 200
1990	n＞10 300。

25.免掉麻烦
——同余理论

长期以来，我们已习惯在星期日休息了，最近星期六也是休息日，只是星期一到星期五是工作日。如果我们把日子按自然数顺序排列，那么我们很难确定第 19579 日或 378 日是工作日还是休息日,甚至也难知道 9 月 30 日是星期几。因此 7 天一轮的制度给我们减少许多麻烦，而且每月如果 3 日是星期日，那么 10 日、17 日、24 日、31 日也是星期日，显然它们的差是 7 或 7 的倍数。

为此我们引进了数的同余概念。高斯的定义是，如果两个整数 a 和 b 的差可被非零整数 m 整除，那么 a 和 b 就对模 m 同余。换句话说，两个整数模 m 同余，如果它们除以 m 之后所得的余数相同。高斯用三条短平行线来表示同余，这个记号到今天仍然在使用：$a\equiv b\ (\mathrm{mod}\ m)$。例如，17 和 52 是模 7 同余的，因为它们除以 7 之后所得的余数都是 3，即 $17\equiv 52\ (\mathrm{mod}\ 7)$。高斯还指出了用同余数进行运算的可能性。对于同一模 m 同余，我们有下列规则：

1.如两数同余于第三数，则这两数也彼此同余

2.两同余式可以相加、相减、相乘

即如 $a\equiv b\ (\mathrm{mod}\ m)$　$c\equiv d\ (\mathrm{mod}\ m)$

则 $a\pm c\equiv b\pm d\ (\mathrm{mod}\ m)$

$ac\equiv bd\ (\mathrm{mod}\ n)$

定义同余以后，就要问一次同余方程 $ax\equiv b\ (\mathrm{mod}\ p)$ 有没有解？实际上这就是解一次不定方程 ax+py=b，这个问题不难解决。但是二次同余方程 $x^2\equiv q\ (\mathrm{mod}\ p)$ 有没有解可就不简单了，比如 $x^2\equiv 13\ (\mathrm{mod}17)$ 有没有解呢？你当然可以一个一个试，但是有没有一般的规律呢？当然欧洲几位最大的数学家都想解决这个问题，但是只取得部分的成功，而二十岁刚出头的高斯一举证明一个一般的规律——二次互反律，其中一个情形是说：如果 p，q 都是 4n+1 型的奇素数，那么如 $x^2\equiv p\ (\mathrm{mod}\ q)$ 有解，则 $x^2\equiv q\ (\mathrm{mod}\ p)$ 也有解，如果 $x^2\equiv p\ (\mathrm{mod}\ q)$ 无解。则 $x^2\equiv q\ (\mathrm{mod}\ p)$ 也无解，反过来也一样。这个规律太重要了，所以高斯称它为黄金规律。

二次互反律涉及的一些特殊情形，欧拉也早就知道。欧拉证明，当 p 为 4n+1 型的奇素数时，二次同余式

$$x^2+1\equiv o\ (\mathrm{mod}\ p)$$

有整数解 x，而当 p 为 4n+3 型的素数时，

$$x^2+1\equiv o\ (\mathrm{mod}\ p)$$

没有整数解 x。用二次剩余的语言来讲，对于 4n+1 型素数，-1 为二次剩余，而对于 4n+3 型素数，-1 为二次非剩余。

同余理论是近代数论的基础，它直接导向代数数论，利用它还可以证明许多重要定理，其中包括，如 p 为 4n+1 型素数，则它唯一地可表为两个数的平方和。

26.整数的加法分析
——加法表示

每个正整数都可以唯一表示成为素因子的乘积，所以说，素数是正整数乘法表示的原子，如果把乘法换成加法，问题就变得比较复杂。每一个整数有多种多样的加法表示，以 5 为例：

5=5，5=4+1，5=3+2

5=3+1+1，5=2+2+1

5=2+1+1+1 5=1+1+1+1+1

其中两个极端是 5=5 和 5=1+1+1+1+1，实际上每一个正整数都有这两个极端的表示，但 5=5 不是一个标准的加法表示，而 5=1+1+1+1+1 是平凡的加法表示，没什么意思。因此，正整数的加法表示需要具体分析。

首先，我们必需知道一个正整数到底有多少不同的分析，这个数目我们称为分析数，用 p（n）来表示，例如 p（5）=7。n 小时，p（n）的数值如下：

p（1）=1

p（2）=2

p（3）=3

p（4）=5

p（5）=7

p（6）=11

p（7）=15

p（8）=22

p（9）=30

p（10）=42

显然，p（n）是增长很快的函数，实际上，可以算出

p（100）=190，569，292

p（200）=3，972，999，029，388

即接近 4 万亿。因此，要想通过手算找出 p（n）的精确值是很困难的事，即使靠计算机，算出 p（1000）也并不容易。数学家于是去求当 n→∞，p（n）的渐近公式。1918 年英国著名数学家哈代和印度数学奇才拉马努詹证明，当 n→∞时，

$$p(n) \sim \frac{1}{4n\sqrt{3}} 3A\sqrt{n}$$

其中$A = \sqrt{\frac{2}{3}}$，这公式是说p（t）是随n指数增长的。

为了研究正整数的加法分析的规律，我们转而研究这些加法表示中有什么合乎规律的东西。在这方面，我们去寻找在这些加法表示中，是否每个数都能用一些特殊的数来进行加法表示，如果可能需要用多少个。特殊的数有平方数、立方数、素数以及图形数等等。而证明每一个正整数都能表为若干个这类特殊的数之和则是十分困难的问题，常常是使数学家几十年几百年研究还不能解决的问题，这些问题构成加法数论的主题。

27.Eureka！（我发现了）

——Num=△+△+△

阿基米德在洗澡时，发现了浮力定律，他从浴盆里一跃而出，大喊：Eureka !Eureka！从此 Eureka 成为表达“ 我发现了 ”的惊喜心情的标准用语。高斯在他 1796 年 7 月 10 日的日记中写道：Eureka，Num=△+△+△，这是加法数论第二条一般定理，这里的 N 是任一正整数，而△表示三角形数，N=△+△+△是说任一个正整数可以表为不超过三个三角数之和。

三角形数是最简单的图形数。顾名思义，它可以用三角形来表示，它的一般表达式是：

$$\frac{n(n+1)}{2} \qquad n = 2,3,4\cdots\cdots$$

最小的三角数是，3，6，10，15，21，28……为了完整起见，把 0 和 1 也加入其中。我们看看定理对不对，例如

19=3+6+10

20=0+10+10

21=0+6+15=0+0+21=1+10+10

22=1+6+15=0+1+21

23=3+10+10=1+1+21

但是，列举充其量只能证明有限多种情形，而无穷多个正整数则需要严格的证明。高斯在日记中只是给出断言，而第一个发表的证明是勒让德在 1798 年给出的。

实际上 1636 年左右，费尔马在他的书上提出重要定理：每个正整数或者本身是个三角形数，或者是两个或三个三角数之和；每个正整数或者本身是个正方形数或者是两个、三个或四个正方形数之和，每个正整数或者本身是个五边形数，或者是两个、三个、四个或五个五边形数之和，如此等等。费尔马声称他用无穷递降法证明了它们，实际上很可能言过其实。

其中正方形数猜想即四平方和问题首先由拉格朗日在 1770 年证明,而高斯在 1796 年对三角形数得到证明。而一般的多角形数猜想是柯西在 1813 年证明了的。

28.用平方数能表示所有整数吗
——四平方和定理

四平方和问题是加法数论中第一个一般性的猜想。巴谢在他出版的丢番图的《算术》中加上评注，他说，丢番图多次做出猜想：每个正整数都能表为四个整数的平方之和，这就是著名的四平方和问题。他补充道，他从 1 到 325 验证过这个猜想，但是不能给出一个一般的证明。笛卡儿也研究过这个问题，但是他后来承认，这个问题太难了，他想不出什么办法来解决它。费尔马在他那本巴谢的译本中，费尔马说他已经证明一个一般的定理，由此可推出四平方和的猜想。他 多次许诺发表证明细节，不过同往常一样没能实现。

后来在费尔马 1654 年写给帕斯卡的一封信中,他谈到为了证明这个一般性的定理，必须证明下面几个命题：

（1）每个形如 4n+1 的素数等于两个整数的平方之和。

（2）给定形如 4n+1 的素数，能找到一个普遍的方法求出两个整数，使其平方和等于该素数。

（3）每个形如 3n+1 的素数均具有 x^2+3y^2 的形式。

（4）每个形如 8n+1 或 8n+3 的素数均具有 x^2+2y^2 的形式。

（5）不存在边长为有理数的直角三角形，使得其面积为完全平方。

在这些命题中，1 曾在 1640 年底给马桑的信中提到过，后来也说过他用无限递降法证明，实际上没有发现他的证明。2、3、4 也有类似情况。他真正作出证明的只有 5，这是用无限递降法做的。

其后，欧拉化了许多力气但是没能完全证明四平方和定理，第一个给出完全证明的是法国大数学家拉格朗日，他的证明用到欧拉的一个基本公式：

$$\begin{cases} (a^2+b^2+c^2+d^2)\ (p^2+q^2+r^2+s^2) = x^2+y^2+z^2+v^2 \\ x = ap+bq+cr+ds,\ y = aq-bp\pm cs\pm dr \\ z = ar\pm bs-cp\pm dq,\ v = as\pm br\pm cq-dp \end{cases}$$

（其中±或∓，或者都选上方符号，或者都选下方符号，所得两套公式都对）。这个公式告诉我们，如果两个数都能表为四个平方数之和，那么它们的乘积也可以表为四个平方数之和，因此，剩下的问题只要证明任一素数都可以表示为四个平方数之和即可。

除了 2 之外，所有的素数可以分成 4n+1 和 4n+3（或 4n-1）两类。对于 4n+1 型的素数很早人们就知道它能表为两个平方数之和。这定理第一个证明是欧拉给出的，对于 4n+1 的复合数，因为有公式（a^2+b^2）（c^2+d^2）=（ac ± bd）2+（ad ∓ bc）2 所以也可以表为二个平方数之和。对于 4n+2 的数，总可以写为两个 4n+1 型的数之和，因此可以表为四个平方数之和。对于 4n+3 的情形，拉格朗日给出证明。而 4n 的情形，因为 4 是平方数，又归结为 n 是上三种情形。这样就得出任一正整数都可以表示为四个平方数之和的结论。这是数论历史上第一个不太简单的一般定理。

29.立方数呢
——华林问题

由平方数人们自然考虑到立方数、四次方数……但是方幂越高，这样的方幂数就越稀少，因此是否能够用有限个来表示每一个正整数就很成问题。1770 年华林发表了《代数沉思录》。其中说，每一个整数至多是 9 个立方数之和，至多是 19 个四次方数之和。他还猜测，每一个正整数都可以表示成至多 r 个 k 次幂之和，其中 r 依赖于 k，这就称为华林问题。

这个问题发表以后，许多人进行了大量验算，他们发现把一个正整数表为立方数的和时，至少需要 9 个立方数，头一个需要 9 个立方数是 29，第二个是 239，但是对于较大的正整数往往只需要 7 个甚至 6 个立方数，这就促使人们把所有正整数和充分大的正整数两种情形区分开。我们通常用 g（k）表示每一个正整数都至多可以表为 g（k）个 k 次方数之和，而用 G（k）表示每一个充分大的正整数都至多可以表为 G（k）个 k 次方数之和，显然，G（k）≤g（k）。

由 g（k）的定义可知，华林的问题就是 g（2）=4，g（3）=9，g（4）=19，g（5）=37，……由上面讲的可以看出，平方数问题已经难倒大数家欧拉，而立方数、四次方数问题，19 世纪许多数学家都进行过研究，分别得到比 9 和 19 大的结果。第一个证明超越数存在的数学家刘维尔就曾经证明过 g（4）≤53。一直到 20 世纪初期，53 才改进到 38。

1908 年，希尔伯特一举证明了对于每个方次 k，g（k）都存在并且是有限值，证明过程开始用到 25 重积分，一下子就把华林问题从原则上解决了。尽管希尔伯特得出 g（k）有限的结论，人们还是希望给出 g（k）的精确值，至少给出一个上界。20 世纪两位解析数论的大家哈代和李特渥德创造了圆法，得出 g（k）＜k2k-1+1 的结论。比如 k 等于 4，g（k）≤33。这比以前的结果改进很多，而且是个一般的定量结果了。在这之前也证明 g（3）=9。对于 k≥6，可证明一般有

$$g(k)=2^k+[(\frac{3}{2})^k]-2$$

其中[x]表示 x 的整数部分。g（5）=37 已由陈景润在文化大革命前得到了证明。1986 年 g（4）=19 才最后获得证明。这样，原来的华林问题基本上得到解决。

数学不断发展，老问题解决了又产生许多新问题有待解决，其中最主要是 G（k）问题。现在只已知 G（2）=4，G（4）=16，而首当其冲的 G（3），现在只知 4≤G（3）≤7。

30.到 1+1 的艰苦攀登
——哥德巴赫猜想

哥德巴赫是德国数学家，他喜欢研究数字的规律。他发现，把偶数拆成两个奇数的和时，其中至少有一组是两个奇素数，而把一个奇数拆成三个奇数的和时，其中至少有一组是三个都是奇素数。例如：

（1）6=3+3

10=3+7=5+5

12=3+9=5+7

16=3+13=5+11=7+9

（2）31=3+11+17=5+11+15=7+9+15

45=5+17+23=13+15+17=7+13+25

=9+15+21=15+15+15

他对列在最前面的自然数逐一验算，发现都具有这种规律。可是，当数目一大，分析成的组数越来越多，验算也就麻烦了。究竟这是一条普遍的规律，还是有例外？他没有把握。于是他写信（1742 年）去请教当时首屈一指的数学家欧拉。他问欧拉，是不是每个偶数都是两个素数的和，每个奇数都是三个素数的和？欧拉给他的回信中说：他验算到 100 多，发现是对的。他也相信这是对的，但是不能给出一般的证明。

其实，解析几何的创始者笛卡儿，早在一百多年前就已说过：任何偶数是一个素数或两个素数或三个素数的和。只是他说得不确切，要去掉“ 偶 ”字才更全面。后来到 1770 年，华林首次把这个问题以猜想的形式写在书中，并公诸于世。不过，当时都把 1 当成素数，所以问题提得含糊。

现在，“ 哥德巴赫猜想 ”的标准提法是：

（1）每个≥6 的偶数都可表成两个奇素数的和；

（2）每个≥9 的奇数都可表成三个奇素数的和。

容易看出，后一命题是前一命题的推论。

1921 年英国数学家哈代和李特渥德密切合作，运用分析方法，对哥德巴赫问题取得了第一个突破。他们证明：在广义黎曼猜想成立的前提下，每个大奇数都可表为三个奇素数的和，并且导出表法的渐近公式。1937 年，维诺格拉朵夫在“ 圆法 ”的基础上，利用他所创造的“ 三角和方法 ”，去掉了那个前提，证明任何大奇数，都可表示为三个素数的和。

1920 年，布隆改进了原始的筛法，创造了所谓‘ 布隆筛法 ”，得到了（9+9）的结果。后来，别人又陆续证明了（7+7）（1924 年）；（6+6）（1932 年），（5+5）（1938 年）和（4+4）（1940 年）。1947 年，赛尔贝尔格宣布，用他的方法可以证明（2+3）。

1965 年，罗斯，特别是朋比利，大大改进大筛法，顺手得到（1+3）。1966 年，陈景润宣布了（1+2），1973 年全文发表，创造了世界纪录。王元、潘承洞也曾经是世界记录的保持者。1956 年，王元证明了（3+4）；1957 年，王元又证明了（2+3）；1962 年，潘承洞证明了（1+5）；1963 年，潘承洞、王元又都证明了（1+4）。不过现有方法还不能达到顶峰 1+1。

31.埃及的分数
——单位分数

古代埃及使用的分数除了$\frac{2}{3}$以外，往往表示为单位分数之和，这样给数学提出了许多有趣的问题，有些至今仍没有解决。

一个问题是 1 的表示，如果用不同的奇数分母来表示，则至少需要 9 项，例如

$$1=\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{11}+\frac{1}{15}+\frac{1}{21}+\frac{1}{135}+\frac{1}{10395}$$

$$1=\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{11}+\frac{1}{15}+\frac{1}{21}+\frac{1}{165}+\frac{1}{693}$$

$$1=\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{11}+\frac{1}{15}+\frac{1}{21}+\frac{1}{231}+\frac{1}{315}$$

$$1=\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{11}+\frac{1}{15}+\frac{1}{21}+\frac{1}{33}+\frac{1}{45}+\frac{1}{385}$$

$$1=\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{11}+\frac{1}{15}+\frac{1}{21}+\frac{1}{35}+\frac{1}{45}+\frac{1}{231}$$

它们的前 6 项都是相同的。1977 年，李奇又提出一个 11 项表示法

$$1=\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{11}+\frac{1}{15}+\frac{1}{21}+\frac{1}{27}+\frac{1}{35}+\frac{1}{63}+\frac{1}{105}+\frac{1}{135}$$

他还提出不管用多少项来表示，最大的分母至少是 105。关于 1 的单位分数表示有一个一般的定理。当 n＞77 时，存在 n 的不等分析

$n=x_1+x_2+\cdots+x_t$

其中$x_1<x_2\cdots<x_t$，使$\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_t}=1$。

另一个著名问题是表示的最小项数。原则上讲，$\frac{2}{n}$可以表示为两项单位分数之和，$\frac{3}{n}$可以表示成为三项单位分数之和，如果要求各项不同，也不难把$\frac{2}{n}$表示为不同的三项单位分数之和，例如：

$$\frac{2}{3}=\frac{1}{2}+\frac{1}{6}\qquad\frac{1}{6}=\frac{1}{7}+\frac{1}{42}$$

$$\frac{2}{3}=\frac{1}{2}+\frac{1}{7}+\frac{1}{42}$$

但是对整数 n＞1，是否

$$\frac{4}{n}=\frac{1}{x}+\frac{1}{y}+\frac{1}{z}$$

都有正整数解，也就是$\frac{4}{n}$都有三项表示，这个问题还没有完全解决，用计算机进行数值计算 $n\le10^8$，这个猜想是肯定的，当 n 为一些特殊的素数时，这个猜想还没有完全证明。

相应有人猜想

$$\frac{5}{n}=\frac{1}{x}+\frac{1}{y}+\frac{1}{z}$$

也都有正整数解，但只有 $n\leq 10^6$ 时得到验证。

32.毕达哥拉斯的直角三角形
——勾股定理

勾股定理，在西方称为毕达哥拉斯定理，是数学历史上第一个最重要的定理。它是一个几何的定理，即任何直角三角形中，斜边（弦）上的正方形等于两直角边（勾、股）上的正方形的和。由于毕达哥拉斯相信万物皆数，所以他的学派只致力于求满足不定方程

$$x^2+y^2=z^2$$

的整数解，这样的解称为毕达哥拉斯三数组。但是，也正是因为这个定理，发现了不可通约量，因此，算术的内容不能完全概括几何的内容，它的算术内容成了求解不定方程的出发点和范例。

勾股定理的解在许多民族中都独立发现过，其中最著名的是在巴比仑的泥板文书中的确明显列出多组解，从（3，4，5）、（5，12，13），（8，15，17）一直到（13500，12709，18541）埃及、中国和印度都有不同的解法。毕达哥拉斯的确知道生成三组数的一般公式，也就是 $2n+1$，$2n^2+2n$ 分别是二个直角边，$2n^2+2n+1$ 是斜边，则它们构成一个三数组，这是由于

$（k+1）^2=k^2+（2k+1）=1+3+5+\cdots+（2k-1）（2k+1）$

取 $2k+1=（2n+1）^2$

则 $k=2n^2+2n$

因此 $（2n^2+2n）2+（2n+1）^2=（2n2+2n+1）^2$

显然这三数组并非勾股方程的全部解，也不是互素的解。其后得出互素的三数组

$x=p^2-q^2$

$y=2pq$

$z=p2+q2$

而一般解可表为

$x=r（p^2-q^2）$

$y=2pqr$

$z=r（p^2+q^2）$

在欧几里得《几何原本》第 10 卷中有一个一般解的等价形式，而且给出它是一般解的第一个证明。

33.丢番图的年纪
——由确定方程到不定方程

丢番图是位古希腊数学家，他生活在公元 3 世纪，距欧几里得活动的公元前 3 世纪相差 500 多年，他的最有名的著作是《算术》，对于后世的数论及代数影响巨大。

丢番图的生平不很清楚，不过他的墓志铭却泄露一点情况：

坟中安葬着丢番图
多么令人惊讶
它忠实地记录了所经历的道路
上帝给予的童年占六分之一，
又过十二分之一，两颊长胡
再过七分之一，点燃起结婚的蜡烛
五年之后天赐贵子
可怜迟到的宁馨儿
享年仅及其父之半，便进入冰冷的坟墓，
悲伤只有用算术的研究去弥补
又过四年，他也走完了人生的旅途。

这段列成方程即是

$$\frac{1}{6}x+\frac{1}{12}x+\frac{1}{7}x+5+\frac{1}{2}x+4=x$$

x=84，这个方程是确定方程。在他著的《算术》中既有一、二、三次的代数方程（确定方程），也有不定方程，但他更主要以不定方程方面的贡献而知名。因此，现在不定方程被称为丢番图方程，研究不定方程求解问题的方法也称为丢番图分析、丢番图逼近、丢番图几何等等。

所谓不定方程或不定方程组，是指方程中的未知数的数目大于方程的数目。解方程的难易程度大体上同方程的次数有关。一次、二次不定方程已有完整的结果，三次及三次以上的方程只有零星的结果。

丢番图主要研究的是一、二、三次的不定方程，他的不定方程比较特殊，而且运用特殊的方法，这些方法没有什么普遍性，因此，数论之父是费尔马。费尔马在两个方面与丢番图不同：第一，费尔马关注于一般问题和一般方法，第二，费尔马首先强调整数解而不是丢番图的有理数解。由于有理数解比整数解更难，费尔马的确开辟一条数论发展的正确道路。而有理数的问题，到 20 世纪又开始活跃，藉助代数几何方法解不定方程，发展出算术代数几何的前沿领域。

34.物不知其数
——中国剩余定理

在我国古代数学著作《孙子算经》中有一著名问题：“今有物不知其数，三三数之剩二，五五数之剩三，七七数之剩二，问物几何？”这个问题可写成同余方程

$$x\equiv 2\ (\mathrm{mod}\ 3)$$
$$x\equiv 3\ (\mathrm{mod}\ 5)$$
$$x\equiv 2\ (\mathrm{mod}\ 7)$$

也可写成不定方程

$x=3y+2=5z+3=7w+2$

我国古代对这个问题已有实际解法，有的是以歌诀表出的：

“三人同行七十稀，
五树梅花廿一枝
七子团圆正半月
除百零五便得知。”

它的意思是说用 70 乘以模 3 的余数 2，加上 21 乘以模 5 的余数 3，再加上 15 乘以模 7 的余数 2，即 70×2+21×3+15×2=233，再减去 105 的 2 倍得 23，即所求的答案。

由这个问题得出一个定理，称为孙子定理或中国剩余定理：

n 个整数 m_1，……，m_n 两两互素，则同余方程组

$x\equiv a\ i\ (\mathrm{mod}\ mi)\ 1\leq i\leq n$

有唯一解 $\mathrm{mod} m_1\cdots m_n$。

35.无限繁殖的兔子
——菲波那奇数

许多动物繁殖能力极强，很少时间内呈现惊人的增长，到底它们繁殖的数目有多大呢？

假如每一对兔子每个月都生下一对小兔子，但每对兔子至少要在出生一个月之后才能下崽，那么一对兔子在一年内能繁殖出多少对兔子？

开始有一对兔子，一个月后仍是一对，这时它们能够繁殖，因此两个月就有两对兔子，一对是原来的能繁殖的兔子，另一对不能繁殖。第三个月后就有三对兔子，其中两对能繁殖。到第四个月后就有五对兔子，这样下去，我们第 n 个月后就有 Fn 对兔子，因此：

F_0=1，F_1=1，F_2=2，F_3=3

F_4=5，F_5=8，F_6=13，F_7=21

F_8=34，F_9=55，F_{10}=89，F_{11}=144，F_{12}＝233……

而 233 就是一年内繁殖的兔子数。这个序列是 13 世纪初意大利数学家菲波那奇首先得到的，因此称为菲波那奇序列。

由上面的序列可以看出菲波那奇序列有一个最突出的性质，也就是每一项等于它前面两项之和，即：

$F_{n+1}=F_n+F_{n-1}$

在自然界中有许多数字同菲波那奇数有关。例如，在各种植物中，一朵花的花瓣数有 3，5，8，13，21，34，55，89 的都发现过，最后这三数，在雏菊中就有。

在电路设计中也有菲波那奇数

如每个元件____1 欧姆，最右端的元件中电流为 1 安培，则由右向左每个元件上的电压的伏特数恰好都是菲波那奇数 1，1，2，3，5，8……

知道菲波那奇数的性质之后，只要换一下前两位的数字，我们就可以得到各种各样的菲波那奇数列，例如

4，7，11，18，29，47……

5，1，6，7，13，20，33……

这里面最简单而且最有趣的是前面两项是 1 和 2，如果是 1，2，它就是原来的菲波那奇序列，但如果是 2，1，就得出

2，1，3，4，7，11，18……

这个序列完全不同于原来的序列，被称为吕卡斯序列。

菲波那奇数列和吕卡斯数列在数学及科学技术中有许多重要应用，至今仍有大量研究，近年来甚至有专门的数学期刊《菲波那奇季刊》，表明它在理论及应用中的重大作用。

36.张冠李戴的方程
——佩尔方程

1732 年欧拉把 $x^2-Dy^2=1$（其中 D 是一个正整数，一般假定没有平方因子）这种类型的方程错误地叫做佩尔方程，这大概是因为他错误地以为英国数学家约翰·佩尔是一个老解法的发明者，而实际上佩尔与这个方程毫无关系。

这个方程有着悠久的历史，公元前四、五世纪时，印度人在求$\sqrt{2}$的近似值前，曾得出$x^2-2y^2=1$的解（17，12）和（577，408）。同时，毕达哥拉斯学派也得出 $x^2-2y^2=\pm1$ 一个递推公式。同样阿基米德也得出 $x^2-3y^2=1$ 的解（1351，780）。丢番图也研究过这个方程许多特殊情形。

关于佩尔方程，我们已经有了非常完整的结果：如果 D 是一个正整数并且不是完全平方数，则佩尔方程 $x^2-Dy^2=1$ 存在无穷多组整数解（x，y），且任何一组解都可以由某一特殊的解（称为基本解）（x，y）生成出来。具体来讲，任何解（x，y）与基本解（x_0，y_0）的关系为

$$x+y\sqrt{D}=\pm(x_0+y_0\sqrt{D})^n,$$

其中 n 是任意整数。

这样一来问题就变成求（x_0，y_0）的问题，1759年欧拉通过把$\sqrt{D}$展开成连分式而给出解佩尔方程的方法，他的想法是，如果 x，y 满足方程，则x/y是$\sqrt{D}$的很好的近似值。但是，他不能证明他的方法总能求出解，并且所有解都能由$\sqrt{D}$的连分式展开给出来。一直到1766年拉格朗日才完全解决这个问题。

举例来讲，$\sqrt{2}$的连分式展开为

$$\sqrt{2}=1+\cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{2+\cdots}}}$$

由它的近似值$\frac{x}{y}=1,\frac{3}{2},\frac{7}{5},\frac{17}{12}$，……，（x，y）都是$x^2-2y^2=\pm1$ 的解，这正好跟毕达哥拉斯学派所得到的不谋而合。

对于不同的 D，最小解的大小差别很大，这从表中的数据可以看出。

表 $x^2-Dy^2=1$ 的最小解

D	x	y
8	3	1
10	19	6
11	10	3
12	7	9
13	649	180
14	15	4
15	4	1
……		
60	31	4
61	1766319049	226153980
62	63	8

……		
108	1351	130
109	1580706719 86249	15140424455100
110	21	2
……		

对于更一般的方程 $x^2-Dy^2=k$，费尔马说他自己知道这个方程什么时候有解，并且说如果有解的话，他也能够解出来，但他照例没有具体给出证明。最后结果是拉格朗日在 1766 到 1770 年间给出的。

37.永远除下去
——连分数

在表示实数时，不管是有理数还是无理数，通常用十进小数表示。但是在许多情况下，表示分数已很麻烦，更不用说无理数了。在数学上，还常用另一种表示法，这就是连分数。它是从欧几里得辗转相除得来。不过可以化为下面简单的形式，其中 c_1，c_2……都是正实数，而 a_1，a_2，a_3……都是正整数。

$$c_1 = a_1 + \frac{1}{c_2},$$

$$c_2 = a_2 + \frac{1}{c_3},$$

$$c_3 = a_3 + \frac{1}{c_4},\cdots\cdots$$

如果除到某一步，比如说

$$c_n = a_n + \frac{1}{c_{n+1}}$$

c_{n+1} 是一个正整数，这个过程就停止了，这样，我们得到一个有限连分数。不难证明，有限连分数是一个普通分数，反过来，任何分数也可以表示成有限连分数，由这也可看出连分数比小数的优越之处。而无理数则可永远除下去，成为一个无限连分数。我们可以用[a_0；a_1，a_2……]来表示上述连分数。它的第二个好处是二次无理数，都可以表示成为无限循环连分数，正如有理数可以表示为无限循环小数一样，也就是

[a_o，a_1，a_2，……a_n，a_1，a_2，……a_n，a_1……]这样我们可以简记为

$$[a_o; \overline{a_1, a_2\cdots\cdots a_n}]$$

例如 $\sqrt{2} = [1;2,2,2,\cdots\cdots] = [1;\dot{2}]$

实际上更一般的连分数就是从开方得到的。1572 年邦贝利提出一个求平方根的方法，完全等价于把它展成为无穷连分数，即若求

$$\sqrt{A} = \sqrt{a^2 + r}$$

则

$$A - a^2 = (\sqrt{A} + a)(\sqrt{A} - a) = r$$

因此 $\sqrt{A} = a + \dfrac{r}{a + \sqrt{A}}$

这样得到一个递推公式，这就是无穷连分数。不过邦贝利只是把分数照平时办法一步一步做下去，并没有提到连分数，特别是无穷连分数。因此，无穷连分数的发明者的桂冠落在卡塔尔迪身上，他不仅明确给出一个记号，而且首先得出近似分数的概念，用现在的记号来表示：

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots\cdots}}}$$

可记作

$$b0+\frac{a_1}{b_1}+\frac{a_2}{b_2}+\cdots\cdots$$

把$\frac{P_n}{q_n}$记作第n次近似分数。而连分数的一个最重要用处是求无理数的近似值，而这个近似值从某种意义上来讲是最佳的，例如

π=[3；7，15，1，292，1，1……]

其渐近分数依次为

$\frac{3}{1}$，$\frac{22}{7}$，$\frac{233}{106}$，$\frac{355}{113}$，$\frac{103993}{33102}$……

而$\frac{22}{7}$和$\frac{355}{113}$正是祖冲之的疏率和密率。

38.更高一次

——三次不定方程

一次、二次不定方程已有比较完整的理论，但三次及三次以上的方程只有零散的结果，最简单的三次不定方程是莫德尔方程

$y^2=x^3+k$　　k 为一个整数

即使这个方程也颇费周折。早在 17 世纪，费尔马宣布

$y^2=x^3-2$

只有一组整数解 x=3 y=5，但到 1875 年才得到证明。同样也只有一组整数解 x=1，x=0。第一个普遍性结果是莫德尔在 1918 年证明，对任何整数 k，莫德尔方程只有有限多解，显然这反映出三次及高次方程与一、二次方程明显不同，但是对于具体的 k，究竟有多少组整数解，有时并不简单，例如，k=17 有 8 组解

（x，y）=（-2，3），（-1，4），（2，5），（4，9）

（8，23），（43，282），（52，375），（5234，378661）到 1968 年，贝克尔证明莫德尔方程的整数解满足

$\max(|x|,|y|) < \exp(10^{10}|k|10^4)$

其中 max 表示|x|，|y|两者之中的最大值，exp 表示以 e 为底的指数函数，这个数是个天文数字，但原则上给出一个界，可以在这个界内，试验 x 是否满足莫德尔方程。

另一种两变元三次方程是具有常数项的齐次方程。实际上，更一般的 n≥3，

$f(z)=a_nz^n+a_{n-1}x^{n-1}y+\cdots+a_1xy^{n-1}+a_oy^n=c$

（c 为非零整数）只有有限多组整数解，例如

$x^3+dy^3=1$

最多只有一组 x，y 均非零的整数解，对于具体问题还需具体分析。

更一般的定理是西格尔在 1929 年得到的，设 $fi(x_1,\cdots,x_n)=o$（$1\leq i\leq m$）是 n 维空间的代数曲线，如曲线方格＞0（大致是说不是直线，则该不定方程组的整数解的组数有限。

如果我们考虑有理数解，由于有理数比整数多得多，因此，求有理数解的组数就更为困难。这个以莫德尔猜想命名的难题最终在 1983 年由德国数学家法尔廷斯解决，他证明以有理数为系数的代数曲线如其方格＞1，则其上的有理点只有有限多组。

39.数学女王的桂冠

——费尔马大定理

近代数论之父费尔马大约在 1637 年在丢番图《算术》的拉丁文译本上做了大量的旁注。他看到原书讲求解 $x^2+y^2=z^2$ 的问题时，忽然出现灵感，于是在这页的边上写下：“另一方面，不可能把一个立方数表为两个立方数之和或把一个四次方数表为两个四次方数之和，或者，一般来说，一个次数大于 2 的方幂不可能是两个同次方幂之和。”接着他又写道：“我已经发现一个确实非常奇妙的证明，但是书的页边太窄了，写不下。”也就是说，丢番图方程 $u^2+v^2=w^2$，当 $n>2$ 时，没有正整数解，这就是所谓费尔马大定理，也称为费尔马最后定理。

费尔马大定理的证明可归结为 n=4 与 n=奇素数的情形。n=4 时，费尔马用自己的方法的确可以给出证明，但当 n=奇素数时，就很难给出一个统一的证明了。在这种情况下，只好对每一个奇素数来进行证明，但进展是非常缓慢的。1770 年欧拉证明了 n=3，1825 年勒让德及狄利克雷证明了 n=5，1839 年拉梅证明了 n=7。在库默尔取得第一次突破之前，只有这三、四个特殊情形获得证明。库默尔在 1847 年宣布，他对于“无穷多个”素数，证明了费尔马大定理，为此，他发展了一系列理论，特别是理想数以及分圆数理论，他把所有素数分成两部分：正则素数和非正则素数，他证明，对于正则素数费尔马大定理成立。据理论估计，正则素数约占全体素数 60%，非正则素数占 40%，有了电子计算机，就可以借助它对非正则素数进行证明。1978 年完成了对 12.5 万以下非正则素数的证明，1987 年达到 15 万，1992 年达到 100 万，当然这样是不能彻底解决问题的。

许多人从几何方面进行研究，通过解析几何学知道，$u^n+v^n=w^n$ 可以看成 $x^n+y^n=1$（其中 x=u/w，y=v/w，它代表一条平面曲线）。如果费尔马问题有解，那么解就是这条曲线上的有理点，也就是 x，y 坐标都是有理数的点。因此，费尔马大定理也就变成为代数曲线上的有理点数的估计问题。1983 年德国数学家法尔廷斯动用代数几何学的先进工具，一举证明了莫德尔猜想，从而使费尔马定理取得第二次突破，也就是，如果丢番图方程 $u^n+v^n=w^n$ 有互素的正整数解，那么解的个数最多只有有限多个。

1986 年一位德国的年轻数学家符莱出人意料地把费尔马大定理和一种特殊类型的代数曲线——椭圆曲线联系起来。符莱的工作就是把方程 $u^p+v^p+w^p=0$，（其中 p 是素数且 $P\geq5$）与椭圆曲线 $y^2=x(x-u^p)(x+v^p)$ 联系起来，但要求 v 是偶数，u 是 4m+3 型的奇数，这是一种所谓半稳定的椭圆曲线，接着他猜想并经后来多人证明，由谷山猜想可以推出费尔马大定理，而猜想很简单：所有有理数域上的椭圆曲线均为模曲线。最后英国数学家维尔斯花了 6—7 年的时间，利用了包括自己在内十几位大家的方法和技巧，终于漂亮地证明半稳定曲线的谷山猜想，这对于费尔马定理就足够了。

40.无穷无尽的推广

——欧拉的猜想

最原始的不定方程或丢番图方程是

$x^2+y^2=z^2$ (1)

许多不定方程可以看成它的推广，费尔马大定理是它第一个推广。100年后，欧拉又做了第二次推广，他指出

$x^3+y^3+z^3=a^3$ (2)

有解，因为

$3^3+4^3+5^3=6^3$

仿照三次情形，他在 1778 年猜想

$x^4+y^4+z^4+t^4=u^4$ (3)

有解，但

$x^4+y^4+z^4=t^4$ (4)

没有解。注意到（5）和（6）的左边项数都等于方程的次数，还可以作类似的推广：不定方程

$x^5+y^5+z^5+t^5+u^5=v^5$ (5)

有正整数解，而

$x^4+y^5+z^4+t^5=u^5$ (6)

没有正整数解，如此等等。

大量不定方程就是这么推广来的，这种推广虽然不费事，可是却够几代数学家干一辈子还不一定有什么结果。费尔马大定理已经有 350 年历史，直到最近才得到解决。但欧拉猜想却有了结果，1911 年得出（6）的一组解 $30^4+120^4+274^4+315^4=353^4$ 但一直到 1987 年夏天才用椭圆曲线理论及计算机推翻（4）无解的猜想，给出了一个反例：

$$95800^4+217519^4+414560^4=422481^4$$

而五次方的欧拉猜想（6）无解，在 1966 年也给出了反例

$$27^5+84^5+110^5+133^5=144^5$$

但（5）的解至今尚未得出。这个问题还可以从不同方向推广，一是向高次幂推广，一是左边项数增多，一是右边项数增多。当（3）的左边再加上一次，100 年前就找到一组解：

$$4^4+6^4+8^4+9^4+14^4=15^4$$

（5）的左边再加上一项也有解

$$4^5+5^5+6^5+7^5+9^5+11^5=12^5$$

可以看出单是等幂次的不定方程就可以无尽无休地推广下去。

41.唯一因子分解定理不成立之时

——代数数论

1831 年高斯把原来的有理整数的概念加以推广，而建立一种全新的整数概念，即形如

$$a+bi$$

的“数”，其中 a，b 是通常的有理整数，i 是 $x^2+1=0$ 的根，他称这种数为复数。为了与通常复数区别，我们这里称为复整数。高斯对于复整数引进加法、减法及乘法，发现它们也具有通常有理整数的性质。只是有理整数的“单位”+1 及-1，现在变成 4 个，即+1，-1，+i，-i。如果两数相差一个单位因子，称为相伴，在分解因子时可不加区别。高斯的重要贡献在于他把通常整数的唯一因子分解定理推广到复整数上，从而开辟了代数数论的新领域。然后他定义素数，他称一个复整数为素数，如果它不是单位，也不能表为两个非单位复整数的乘积。由这个意义，显然通常的复合数仍然是复合数，但通常的素数却不一定仍是素数，例如

$$2=(1+i)\ (1-i)$$

$$5=(1+2i)\ (1-2i)$$

更一般讲，4n+1 型素数 p 都不再是素数，因为

$$p=m^2+n^2=(m+ni)\ (m-ni)$$

另一方面，高斯证明，4n+3 型的素数在复整数中仍是素数，接着他对复整数引进欧几里得算法并证明唯一因子分解定理。

但是并不是其他的代数整数都像高斯整数一样，满足素因子唯一分解定理，例如形如

$$a+b\sqrt{-5}$$

的所有代数整数，其中 a，b 是通常的整数，它们相加、相减、相乘仍是这种形式的代数整数，它们构成一个环。在这些代数整数中，2，3，$1+\sqrt{5}$，$1-\sqrt{5}$都不能再分解因子了，它们也不是单位，可是6却可以分解成两种形式

$$6=2\times 3=(1+\sqrt{5i})(1-\sqrt{5i})$$

因此，对这些代数整数，素因子唯一分解定理不成立。现在知道，在诸多的代数整数环中，能唯一分解的只是极少数。数学家引进一个数称为该代表整数环的类数（h）。环的类数反映它与素因子唯一分解定理的偏差，只有当 h=1 时，素因子唯一分解定理才成立。因此计算类数是非常困难的数学问题，对于虚二次域，也就是形如

$$a+b\sqrt{-D} \qquad D>O$$

的代数整数环。到 1966 年才证明 h=1 的只有 9 个，即 D=-1，-2，-3，-7，-11，-19，-43，-67，-163，对于虚高次域，到 1994 年才得出全部的 h=1 的域，共 172 个。但实域，甚至实二次域还没有最后的结果。至今还不知高斯猜想是否成立：有无穷多个实二次域，其 h=1。

42.有理数与无理数的差别

——丢番图逼近

丢番图逼近是求无理数用有理数来逼近，而这个近似值往往需要解丢番图方程，因此这个领域称为丢番图逼近。

古希腊人就知道$\sqrt{2}$是一个无理数，也就是不能表示为分数，用丢番图方程的术语来讲，$x^2-2y^2=0$ 是没有非零整数解。因此 x^2 和 $2y^2$ 作为两个整数相差至少为 1，也就是我们如果求出丢番图方程

$$x^2-2y^2=\pm 1$$

的解（x，y），则$\frac{x}{y}$是$\sqrt{2}$的“最佳近似值”。到17世纪已经知道佩尔方程具有无穷多解，由小到大排列如下：

x	y	x/y
1	1	1.0
3	2	1.5
7	5	1.4
17	12	1.416 ……
41	29	1.4137
99	70	1.41428 ……
239	169	1.414201 ……
577	408	1.414215 ……
1393	985	1.4142132 ……
3363	2376	1.4142136 ……
…	…	…
…	…	…

同样，在一般的情形下，利用一个无理数α的连分数展开，可以得到α的一系列最佳近似值。而且α的两个相继的渐近分数中，一定有一个满足

$$|\alpha-\frac{p}{q}|<\frac{1}{2q^2}$$

而三个相继的渐近分数中，必有一个满足

$$|\alpha-\frac{p}{q}|<\frac{1}{\sqrt{5}q^2}$$

因此，任一个无理数有无穷多有理数$\frac{p}{q}$使

$$|\alpha-\frac{p}{q}|<\frac{1}{\sqrt{5}q^2}$$

而且$\sqrt{5}$是最佳的，也就是如果$A>\sqrt{5}$，则存在实数α使

$$|\alpha-\frac{p}{q}|<\frac{1}{Aq^2}$$

没有无穷多解。

另一方面，对于代数数α，只有有穷多个 p/q，满足

$$|\alpha-\frac{p}{q}|<\frac{1}{q^{2+\varepsilon}}$$

其中 ε 为任意小正数，这是著名的罗斯定理，是英国数学家罗斯于 1955 年证明的，这也是最终的、不能再改进的结果。

43.不满足代数方程的数

——超越数

除了把实数分为有理数和无理数的分法之外，还有一种更为重要的分法，即分为代数数和超越数。凡是满足整系数代数方程的数称为代数数，而不是代数数的数就称为超越数。

由于超越数的概念比较精致，所以直到欧拉的著作中才首次出现。他在1748年出版的《无穷分析引论》第一卷第六章论指数及对数量中，他未加证明地断言：“如果数b不是底a的幂，其对数就不再是一个无理数。事实上，如有$\log ab=\sqrt{n}$，则有$a^{\sqrt{n}}=b$。假如a，b都是有理数，这等式不能成立，因因而对于这种不是底a的幂的数b，其对数应当恰如其分地命名为超越数”。一直到100年后法国数学家刘维尔才第一次在1844年证明超越数的确存在。他认识到代数数和超越数的区别是有理数和无理数区别的深化，也就是说，超越数是无理数中更为“无理”的数，也就是性质和有理数差异更大的数。

刘维尔的想法是一个有理数用另外的有理数来逼近，差别总是较大的，对于代数数来说，只有有限多对有理数能够很好地逼近它，也就是，如果ξ是一个代数数，满足n次整系数代数方程，则对任一ε＞0，A＞0，满足不等式

$$|\varepsilon-\frac{q}{p}|<\frac{A}{q^{n+\varepsilon}}$$

的（p，q）只有有限多对。因此，要是对于任何 n，有无穷多对（p，q）都满足上述不等式，那肯定就是超越数了。也就是说，超越数能够被有理数极为精确地逼近。

于是，刘维尔本着这样的原则，造出来下面的超越数：

$$\varepsilon=\frac{1}{10}+\frac{1}{10^{2!}}+\frac{1}{10^{3!}}+\cdots\cdots$$

其中n！是n的阶乘，n！=1×2×3……n

这样一来，超越数总算存在了，不过这究竟是人为造出来的。又过了近30年，1873年埃尔米特第一次证明一个微积分上常见的数e是超越数。1900年国际数学家大会上提出了23个重要的问题，其中第七问题就是关于超越数的问题。他做了一些猜想，其中他推测像$2^{\sqrt{2}}$和e^{π}这样的数是超越数。这个问题在1929年开始取得突破，有人证明e^{π}是超越数，1930年就证明了$2^{\sqrt{2}}$的超越性。1882年德国数学家林德曼基本上用埃尔米特的方法证明了π是超越数。单单这个结果看来还不那么了不起，可是由此使希腊几何三大难题最后一个化圆为方也彻底得到解决了。

虽然在实数中，超越数要比代数数多得多，但是，证明某一个实数是超越数是非常困难的。例如现在还不知道e+π和欧拉常数r是否超越数。

44.圆内有多少格子点

——格点问题

圆内整点问题是求以原点 O 为圆心，半经为 R 的圆内整点的数目有多少个。这里所谓整点就是坐标为整数的点。对于任意的 R，一般并不能得出精确的点数，而解析数论的威力在于给出一个比较精密的上下界，一般表示为一个或几个主项加上一个误差项，主项一般较为简单，解析数论的功夫都在估计误差项上，而其主要应用也在于此。以圆内整点问题为例，圆内整点数 N（R）接近于圆的面积 πR^2，这个不难证明，可是真正的问题是差多少？高斯已知

$$N(R)=\pi R^2+O(R)$$

20 世纪最伟大的解析数论专家哈代在 1913 年证明

$$N(R)=\pi R^2+O(R^{2/3})$$

也就是有一个常数 M，使

$$\pi R^2-MR^{\frac{2}{3}}<N(R)<\pi R^2+MR^{\frac{2}{3}}$$

显然为了使估计的数更加精密，必须把误差项的指数 $\frac{2}{3}$ 降下来。表面上 M 减小也起作用，但这根本不是主要问题。一般的解析数论问题讨论最佳的误差项的指数，也就是指数降到什么程度就不能再降了。对于圆内整点问题最佳结果是

$$N(R)=\pi R^2+O(R^{\frac{1}{2}+\varepsilon})$$

其中 ε 为任意小常数。不过时至今日，这还只是一个猜想。

虽然近百年来，指数有一点改进，但离最佳结果仍相差甚远。圆内整体问题的一个姐妹问题是狄利克雷除数问题，这个问题是求和

$$\sum_{1\le n\le x}d(n)$$

d（n）表示 n 的因子数目。这个数值恰巧等于区域 $1\le u\le x$，$1\le v\le x$，$uv\le x$ 上的格点的个数，最佳结果为

$$\sum_{1\le n\le x}d(n)=x\log x+(2r-1)x+O(x^{\frac{1}{4}+\varepsilon})$$

现在所得指数也离 $\frac{1}{4}$ 尚远。

格点问题可推广到球内整点及 k 维椭球内整点乃至一般的区域。它不仅在数论、二次型理论、数之几何等方面有用，而且在物理理论中至关重要。

45.数学的女王

——数　论

高斯曾说过：“数学是科学的女王，数论是数学的女王。”从这时起，数论在数学中占有崇高的地位，是纯粹数学的主要代表。

数，特别是自然数，是数学的主要研究对象。数的科学在古希腊就已区别开算术与数论，前者是计算技术，后者是数的性质及数之间的关系。严格从字面的意义上讲，数论应该是关于数的生成、数的构造、如何产生新数以及数系的结构，不过，这些到 19 世纪才成熟的理论现在已经不属于数论的范围了。

从历史上来看，随着一个一个数论问题的提出和解决，越来越多的数论问题又提了出来，数论的历史可以说是数论的问题提出和解决的历史，在这个历史的长河中，产生出解决问题的技术及方法——算术方法（或说初等方法）、代数方法、解析方法、几何方法、概率方法等，另一方面，也形成一些理论，如代数数论，特别是类域论之类的漂亮理论。但是，与抽象代数和拓朴学不同，数论，特别是整数论还是个以问题为主的学科。

涉及整数的问题有四大类：

（1）乘法表示的数论问题：如整数的可除性、因子分解、素数、同余理论等问题，其中最主要的问题是代数数论中的互反律问题以及涉及素数分布的黎曼猜想。

（2）加法表示的数论问题。最基本问题是四平方和问题，即任一正整数均可表为最多四个平方数之和。这问题最终由拉格朗日在 1770 年完全证明，可以说是加法数论的奠基性定理。作为这个定理的推广，加法数论的基本问题是华林问题，原来的华林问题到 1986 年已彻底解决，但其各种推广则远未完结。另一个加法数论的问题是哥德巴赫猜想问题，它也可以说是加法数论及乘法数论的混合问题。由四平方和定理还衍生出二次型理论，作为它的推广，产生出一般型论、一般代数数论与椭圆模函数理论。

（3）丢番图方程求解问题。从某种意义上来讲，几乎所有数论问题，甚至数学问题都可以归结为丢番图方程求解问题。所谓丢番图方程，就是不定方程，也就是方程或方程组，其未知数的数目大于方程的数目。这个方向还形成丢番图分析、丢番图逼近、丢番图几何等当前热门。

（4）其他数论问题，如数论函数、整数序列、组合数论以及种种杂题。

时至今日，不仅数论的理论成果引人注目，而且近 10 多年来在科学技术，特别是物理学及信息技术方面获得越来越重要的应用。数论在编码及解码方面有不可忽视的作用。

组合数学与图论

46.无处不在的表示

——阶　乘

在组合理论中，n 的阶乘是一种非常关键的数字。随着 n 的增大，n！以指数的方式迅速增加。前面几个阶乘可以用手算出来：

1！=1

2！=2

3！=6

4！=24

5！=120

6！=720

7！=5040

8！=40320

9！=362880

10！=3628800

再算下去就是非常吃力的活了。16 世纪到 17 世纪，许多人进行阶乘的计算并且列出表来。1669 年基尔舍出版《组合学大法》，列出到 64！的数值。其中

50！=12737268388154203998513430837670055152977494547954734080000000000000≈1.27×10^{66}

这比全宇宙中的原子数还要多上亿亿亿倍，真是名副其实的天文数字。因此，从实用角度看，我们没有必要精确计算出 n！的值，而只需要求出近似公式，这样能够算出前几位准确值以及它的位数就够了。1730 年左右，英国数学家斯特灵和其他人独立地得出这个公式，这就是著名的斯特灵公式：

$$n!\approx(\frac{n}{e})^n\sqrt{2\pi n}$$

其中≈表示近似等于，这个公式中包含数学中两个最重要的常数：一个是圆周率π，一个是自然对数底 e，e=2.71828…。这种近似公式往往 n 越大，近似值越精确，用这个公式我们得出

$$70!\approx1.2\times10^{100}$$

也就是阶乘中首次突破 100 位大关的数。从这里也可以看出计算机出现之前手算的艰辛，同时也表明，数学中近似估计的无比重要性。

47.有多少种安排

——排列与组合

无论在日常生活中，还是在科学技术工作里，人们都大量遇到排列组合问题。例如，把三件东西排成一行，有多少种排法？对于这种简单的问题，我们只需把所有情形都列举出来，然后再数一下就行了，不妨用 A，B，C 代表这三件东西，那么我们可以有下面 6 种排法：ABC，ACB，BAC，BCA，CAB，CBA。想想看，三件东西排成一行只有这 6 种可能性。四件东西的不同排列方式只有 24 种。注意一下

$$6=3\times2\times1$$

$$24=4\times3\times2\times1$$

我们可以猜想五件东西，共有

$$120=5\times4\times3\times2\times1$$

种排列方法，不难把这个结果推广到 n 件东西，n 件东西排成一行共有

$$n\times(n-1)\times(n-2)\times\cdots\times3\times2\times1$$

种排法，为了方便起见，可以把上面的乘积记做 n！，称为 n 的阶乘，这样我们就可以得到一个一般的定理，n 件东西排成一行，共有 n！种排法。

下面我们问：如果从 n 件东西，选出 r 件排成一行，共有多少种排法？这个问题稍稍复杂一些，但也不太难。我们可以这样想：第一个位置，我们有 n 种选法，当第一个位置选定后，第二个位置就不能选已经占有第一个位置的东西了，所有第二个位置只有 n-1 种选法，同理第三个位置有 n-2 种选法，如此下去，第 r 个位置有 n-（r-1）种即 n-r+1 种选法，因此从 n 件东西中选出r件东西的不同排法数P_r^n是

$$n\times(n-1)\times(n-2)\times\cdots\times(n-r+1)。$$

现在我们考虑从 n 件东西中，选出 r 件东西而不管顺序，这个数目称为组合数，记作C_r^n，组合数比排列数少多少倍呢？前面我们知道，3件东西排列数为 3！，4 件东西排列数为 4！，r 件东西排列数当然为 r！，这 r！的排列现在不计顺序只相当于一种组合，因此

$$C_r^n=\frac{P_r^n}{r!}=\frac{n(n-1)(n-2)\cdots(n-r+1)}{r(r-1)(r-2)\cdots1}$$

这后一式常常记作$\binom{n}{r}$。这个公式在各国的典籍中都以具体的形式出现过，例如在 6 世纪印度经典中提到从 16 种原料中每次选 4 种配成香料，共有 1820 种配方，实际上，配方数就是

$$C_4^{16}=\binom{16}{4}=\frac{16\times15\times14\times13}{4\times3\times2\times1}=1820$$

注意在所有组合数中，两个数最特别，一个是

$C_1^n=n$，

另一个是

C_n^n永远是1。

48.奇异的三角形

——杨辉三角

组合数$\binom{n}{r}$有着许多奇妙的性质，我们先看一下当n，r都小时$\binom{n}{r}$的数值，不难算出

$\binom{2}{1}=2$，$\binom{2}{2}=1$

$\binom{3}{1}=3$，$\binom{3}{2}=3$，$\binom{3}{3}=1$

$\binom{4}{1}=4$，$\binom{4}{2}=6$，$\binom{4}{3}=4$，$\binom{4}{4}=1$

$\binom{5}{1}=5$，$\binom{5}{2}=10$，$\binom{5}{3}=10$，$\binom{5}{4}=5$，$\binom{5}{5}=1$

$\binom{6}{1}=6$，$\binom{6}{2}=15$，$\binom{6}{3}=20$，$\binom{6}{4}=15$，$\binom{6}{5}=6$，$\binom{6}{6}=1$

然后在每一行的左侧都加上1，我们可以把这个1解释为C_0^n，也就是从n件东西中1件都不选只有1种可能性。这样我们就用C_0^n，C_1^n，C_2^n，…C_{n-1}^n，C_{n-1}^n为第n行组成一个三角形

Cnn-2，Cnn-1，Cnn 为第 n 行组成一个三角形

```
                1
              1   1
            1   2   1
          1   3   3   1
        1   4   6   4   1
      1   5   10  10  5   1
    1   6   15  20  15  6   1
      ……
```

为了完整我们在顶尖上加上一个 1（第0行），不妨解释为C_0^0，这个三角形就是著名的杨辉三角，在西方称为帕斯卡三角形，这是因为帕斯卡在 1665 年写了《论算术三角形》中发表了这个三角形。实际上，中国及印度早都发现了这个三角形。这个三角形反映出组合数两个最重要的规律，一个是左右对称性，也就是每一行中

$$C_r^n=\binom{n}{r}=\left(\frac{n}{n-r}\right)=C_{n-r}^n$$

另一个是上下两行之间的关系，即每一组合数等于上面一行两组合数之和

$$C_r^n=(C_r^{n-1}+C_{r-1}^{n-1})$$

这是杨辉三角的绝妙之处。

49.奇妙的关系

——二项式定理

初等代数中最基本的定理就是二项式定理。当 n 为正整数时，二项式定理就是求

$(a+b)^n=a^n+Aa^{n-1}b+Ba^{n-2}b^2+\cdots+Lab^{n-1}+b^n$ 中各项的系数。当 n=2，3 时，很早就知道

$$(a+b)^2=a^2+2ab+b^2$$

$(a+b)^3=a^3+3a^2b+3ab^2+b^3$

大约公元 1100 年阿拉伯数学家已经得到 n=4，5，6 的公式，并且知道一般的公式。中国元代数学家朱世杰也知道到 n=8 的公式。他们的结果都不谋而合，$a^{n-r}b^r$的系数恰巧就是$C_r^n=\binom{n}{r}$。因此，后者也恰当地称为二项系数。帕斯卡在他 1654 年写的《论算术三角形》中证明了这个二项式定理

$$(a+b)^n=a^n+\binom{n}{1}a^{n-1}b+\binom{n}{2}a^{n-2}b^2+\cdots+\binom{n}{n-1}ab^{n-1}+b^n$$

而且明确递归关系

$$\binom{n}{r}=\frac{n-r+1}{r}\binom{n}{r-1}$$

并且给出二项系数两种不同的表示法。至此，正整数的二项式定理可以说大功告成。

有意思的是，后来人们谈到二项式定理往往说是牛顿的创造，很少提帕斯卡。实际上，牛顿在 1665 年 22 岁时首先考虑把二项式定理推广到分数指数的情形，这样，二项式公式就不是一个有限项公式而是一个无穷级数了。从这个广义的二项式定理出发，人们开始微积分的运算，已经不属于代数的范围了。同时一般的二项式定理的严密证明，一直到 1826 年才算基本完成。从这时起，指数又进一步推广到实数和复数上。

二项式定理另一方向上的推广是三项式、四项式、……k 项式定理，即求

$$(x_1+x_2+\cdots x_k)^n$$

展开式中各项的系数。

由二项式定理可以推出最有趣的关系，令 a=b=1，则

$$(a+b)^n=2^n=\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n-1}+\binom{n}{n}$$

50.至关重要的数

——伯努利数

数学上最重要的一个问题是求

$$Sk=1^k+2^k+\cdots+n^k$$

这就是所谓等幂和问题。很早人们就知道

$$S_1=1+2+3+\cdots+n=\frac{1}{2}n(n+1)$$

$$S_2=1^2+2^2+3^2+\cdots+n^2=\frac{1}{6}n(n+1)(2n+1)$$

$$S_3=1^3+2^3+3^3+\cdots+n^3=[\frac{1}{2}n(n+1)]^2=\frac{1}{4}n^4+\frac{1}{3}n^3+\frac{1}{4}n^2$$

$$S_4=1^4+2^4+3^4+\cdots+n^4=\frac{1}{30}(n+1)(2n+1)(3n^2+3n-1)$$

$$=\frac{1}{5}n^5+\frac{1}{2}n^4+\frac{1}{3}n^3-\frac{1}{30}n$$

到 17 世纪已经求到 S_{17}。费尔马等人从这些结果看出 S_k 可用 S_{k-1}，S_{k-2}……的代数关系式表出，而且当 k 为奇数时

$$Sk=n（n+1）\times（n\text{ 的多项式}）$$

当 k 为偶数时

$$S_k=n（n+1）（2n+1）\times（n\text{ 的多项式}）$$

最后，S_k 是 n 的 k+1 次多项式

$$S_k=a_1n+a_2n^2+\cdots+a_{k+1}n^{k+1}$$

问题是如何把这些系数求出来，它们之间有没有一般的规律性。这个普遍的结果一直到雅各·伯努利才得出来。在他去世后于 1713 年出版的《猜测术》中，他总结出 S_k 的一般规律，而且得出系数的具体表示，其中关键的数列 B_k 称为伯努利数，以纪念他的不朽功绩。

实际上，他给出的公式是一个形式公式，只用到二项式定理：

$$S_k=\frac{(n+B)^{k+1}-B^{k+1}}{K+1}$$

这里需要解释一下，B^{k+1} 不是一个数的 k+1 次幂，也就是 $B^{k+1}\equiv B_{k+1}$，它满足形式等式

$$(B+1)^k-B_k=\sigma_{k1}$$

σ_{k1} 除 k=1 时等于 1 外，对其他 k 都等于 0。这样我们由这个式子求出伯努利数列 B_0，B_1，B_2……的值，当

n=0 时，$B_0=1$

n=1 时我们通过其他办法求出

$$B_1 = \frac{1}{2}$$

n≥2 时 $B_k = \sum_{r=0}^{n} (\binom{k}{r} B_k$

这样可以逐步求出 B_k，而除了 k 为 1 外，对于所有奇数 k B_k=0 而对于偶然 k，B_k 都是正负相间的分数

$$B_2 = \frac{1}{6}, B_4 = -\frac{1}{30}, B_6 = \quad B_8 = \quad B_{10} =$$

有了伯努利数，很容易求出等幂和公式，例如

$$S_1 = 1 + 2 + 3 + \cdots + n = \frac{(n+b)^2 - B^2}{2}$$

$$= \frac{n^2 + 2nB^1}{2} = \frac{n^2 + n}{2} = \frac{1}{2} n(n+1)$$

$$S_2 = 1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{(n+B)^3 - B^3}{3}$$

$$= \frac{n^3 + 3n^2 B^1 + 3nB^2}{3} = \frac{1}{6} n(n+1)(2n+1)$$

$$S_3 = 1^3 + 2^3 + 3^3 + \cdots\cdots + n^3 = \frac{(n+B)^4 - B^4}{4}$$

$$= \frac{n^4 + 4n^3 B^1 + 6n^2 B^2 + 4nB^3}{4} = \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4}$$

$$= [\frac{n}{2}(n+1)]^2$$

$$S_4 = 1^4 + 2^4 + 3^4 + \cdots + n^4 = \frac{(n+B)^5 - B^5}{5}$$

$$= \frac{n^5 + 5n^{4B1} + 10n^3 B^2 + 10n^2 B^3 + 5nB^4}{5}$$

$$= \frac{n^5}{5} + \frac{n^4}{10} + \frac{n^3}{3} - \frac{n}{30} = \frac{1}{30} n(n+1)(2n+1)$$

[3n(n+1)-1]

既然伯努利数很容易计算出来，S_k 不难依法算出。另外等幂和还可以通过差分公式算出，道理是一样的。伯努利数不仅求等幂和必不可少，它在数论、拓扑、统计乃至物理中有着重要作用，是数学中至关重要的数。

51.龟背上的纵横图

——幻方

最古老的组合问题是幻方，它至少有 2000 多年的历史。中国的洛书上有如下的幻方：

4 9 2

3 5 7

8 1 6

这是最原始的三阶幻方，即把 1，2，3，4，5，6，7，8，9 九个数填在 3×3 的九个格子的正方形中，使得每行、每列和对角线上的三个数字之和均

相等。由于它具有很浓厚的神秘色彩，幻方成为古代及中世纪许多民族的研究对象。公元 990 年左右，巴格达的纯净兄弟会的穆斯林学者对幻方极感兴趣，他们曾构造出 3，4，5，6 阶幻方。

到 13 世纪，中国和伊斯兰国家有广泛的文化交流。在西安出土一块铁板，上面用东阿拉伯数码记载一个 6 阶幻方：

28	4	3	31	35	10
36	18	21	24	11	1
7	23	12	17	22	30
8	13	26	19	16	29
5	20	15	14	25	32
27	33	34	6	2	9

杨辉在 1275 年写了一本幻方的书，他把幻方称为纵横图，他举出从 3 阶到 10 阶的幻方，其中用 9 个 3×3 的幻方造出 9×9 的幻方。阿拉伯人的幻方后来传到欧洲，也引起许多人极大兴趣，其中一位是著名德国大画家丢勒（A.Dürer，1471—1528），他也是位数学家，在他一幅称为《忧郁》的画上就有一个幻方：

16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

丢勒的方法可以用来造任何 4n 阶幻方。虽然人们可以造出新的幻方来，但幻方的数学仍有非常多的难题没有得到解决。其中首要的问题是不同的 n 阶幻方到底有多少种？现在只知道 3 阶幻方只有一种，也就是所有的三阶幻方都是洛书上的幻方经过行变换、列变换或绕中心 5 转动而得到，总之中心 5 是不变的，但 4 阶幻方有 880 种。17 世纪初数学家已经列举出来所有的 4 阶幻方。长期以来，5 阶及 5 阶以上幻方的种数仍然不知道，最近用计算机算出 5 阶幻方共有

275305224 种。除了典型的幻方之外，还可以把其他数列排成纵横图，它们有着各种各样的性质，真是丰富多彩，美不胜收。

52.36 军官问题

——正交拉丁方

18 世纪中叶，普鲁士国王腓特烈二世当政，由于他文治武功都很显赫，历史上尊称为腓特烈大王。在他的柏林科学院中，曾邀请欧拉和拉格朗日等大科学家工作，使柏林成为当时的科学及文化中心。

传说腓特烈有一次要阅兵，他从他的六支部队中每个部队选出六名不同级别的军官，他要求这 36 名军官组成一个 6×6 的方阵，使得每行每列都有 6 个部队，6 个级别的军官。但是排来排去总也达不到腓特烈大王的要求。

欧拉开始研究这个问题，他先从简单的开始，假定 n 个部队用大写拉丁字母 A，B，C，D……为代表，n 个级别用小写拉丁字母 a，b，c，d 为代表，这个问题称为 n 阶正交拉丁方问题。当 n=2 时，这问题无解，n=3 时，他排出来了

Aa	Bc	Cb

Bb	Ca	Ac
Cc	Ab	Ba

对于 n=4，欧拉造出两个拉丁方

$$\begin{bmatrix} A & D & B & C \\ D & A & C & B \\ B & C & A & D \\ C & B & D & A \end{bmatrix} \qquad \begin{bmatrix} d & a & c & b \\ b & c & a & d \\ a & d & b & c \\ c & b & d & a \end{bmatrix}$$

也就是每行每列都由 A，B，C，D 或 a，b，c，d 组成，而且没有重复。同时这两个拉丁方放在一起，满足腓特烈大王的要求，它们构成一对正交拉丁方。

n=5 时，欧拉也造出了一对正交的拉丁方，但是 n=6 时，欧拉始终也造不出来。于是在他去世前一年 1782 年，他发表论方，根据 n=2，n=6 时正交拉丁方不存在，猜想 n=4m+2，m=0，1，2……时，不存在正交的 n 阶拉丁方。长期以来，数学家企图证明这个猜想，未获成功。但到 1959 年印度数学家玻色等人出乎人们意料地证明除了 n=2，6 之外，对于其他 n=4m+2，正交拉丁方都存在，也就是欧拉猜想并不成立。他们还先后具体造出 10 阶，14 阶，18 阶，22 阶，26 阶等正交拉丁方，不过 26 阶以上的正交拉丁方，都要靠计算机来构造。至此正交拉丁方的存在问题完全解决，即除了 n=2，n=6 之外，对其他每个 n，都存在正交拉丁方。

当今最重要的一个问题是，对于 n，有多少个互相正交的拉丁方，这个最大的数目记作 N（n），且

$N(n) \leq n-1$

只有 n 是素数幂 p^k 时，可以证明

$N(p^k) = p^k - 1$

这时称该拉丁方组是完全的，而对于 $n \equiv 1 \pmod 4$，$n \equiv 2 \pmod 4$ 时，则肯定 $N(n) < n-1$

对于 N（n）的精确数值，我们所知甚少，例如对小 n 只知道上界和下界

$3 \leq N(14) \leq 10$

$4 \leq N(15) \leq 13$

看来离完全解决相距甚远。正交拉丁方不仅是个数学游戏，在统计学中特别是试验设计中有重要作用。

53.散步增友谊
——寇克曼 15 女生问题

1850 年，英国数学家寇克曼提出一个著名问题，这个问题后来发展成为组合数学中一个大分支——史坦纳三元系理论，在各方面有着重要应用。

他的问题是这样的：在一所寄宿学校某班中有 15 名女生，她们经常每天 3 人一组出去散步。为了增进每个学生同其他学生的友谊，使得每个学生与另外任何学生一起散步的机会都相同，问应该怎样安排，使得每一星期中，每个女生都恰好有一天跟其他任何一位女生一起散步？

从一个女生出发，她星期一同两位女生散步，星期二换两位，这样下去，她在一星期之内同其他 14 位女生都一起散过步。对她来说这样安排并不困难，可是将每位女生都得这样轮一遍，就不太好安排了。在历史上，许多数学家都给出一种安排，但是任何安排都是等价的，也就是说本质上是一种。这样如何有系统地尽快地排出散步表就是一种最佳选择。下面是本杰明·皮尔斯的安排，我们很容易看出轮换的规律：

我们用*代表一位女生，其余 14 位分别用 1，2，3，4，5，6，7 和Ⅰ，Ⅱ，Ⅲ，Ⅳ，Ⅴ，Ⅵ，Ⅶ代表，那么星期日的排列是

1 2 Ⅰ
3 5 Ⅱ
4 7 Ⅲ
6 * Ⅳ
Ⅳ Ⅴ Ⅶ从星期一到星期六每天排法

54.台面上的礼貌
——夫妻围坐问题

按西方的礼仪，n 对夫妇围桌而坐，要求男女必须相邻而坐，而且一对夫妇不能相邻，也就是每位男士两边是两位女士但不能是他太太；每位女士两边是两位男士但不是她丈夫。按照这样的规则，问不同的入座数 M_n 是多少？

这个著名问题由卢卡斯在 1891 年提出来，在此之前，英国物理学家泰特研究纽结问题时出提出一个等价的问题，可见这个问题不是简单的数学游戏，许多数学家研究过这个问题。1934 年，图夏尔给出 M_n 的一个具体表达式，但没有给出证明。最后美国著名数学家卡普兰斯基证明了这个公式。由此可见，这个问题是个不简单的问题。

使许多数学家感到困难的是，由于“女士优先”（Ladiesfirst），也就是先把女士们的座位安排好，再把男士安排在其间，这样就形成有限制的排列问题，这问题要比原问题还难。卡普兰斯基的想法是打破这种礼貌上的限制，因为这对数学问题没什么必要，而直接考虑求解，从而轻易解决该问题。得到的公式是

$$M_n = \sum_{k=o}^{n} (-1)\binom{n}{k}.\,2.\,2n(2n-k-1)!\frac{(n-k)!^2}{(2n-2k)!}$$

这个数目是很大的，但它把本质上一样的排列法重复计算，例如 3 对夫妇围坐，M^3=12 种，而本质上只有一种，也就是每位丈夫对着自己的妻子坐，即

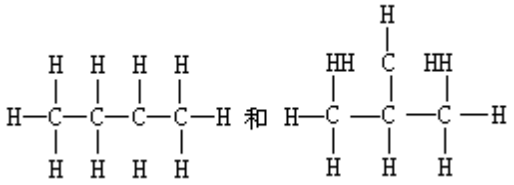
这样 $M_n/2$（n！）更反映不同类型，下面表给出 n 小时 M_n 及 $M_n/2$（n！）的数目：

n	Mn	Mn/2（n！）
2	0	0
3	12	1
4	96	2
5	3120	13
6	115200	80
7	5836320	579
8	382072320	4738
9	31488549120	43287
10	3191834419200	439792

由此可见，n 大时 Mn 及 Mn/2（n！）都增长很快。

55.有多少种同分异构体
—波利亚定理

19 世纪 60 年代，经过许多化学家的努力，知道碳是四价，碳的四个价键指向一个四面体的顶点，从此形成了有机化学的结构理论。正是由于这种特殊的结构，许多化合物，特别是有机化合物，出现了同分异构的现象，也就是分子式相同（同分）而结构式不同（异构）的情况。最简单的有机化合物是碳氢化合物，其中饱和链式碳氢化合物称为烷，分子式为 C_nH_{2n+2}。当 n=1，2，3 时，甲烷、乙烷、丙烷只有一种结构式，而从 n=4 也就是丁烷起，就出现了 2 种乃至多种异构体。例如丁烷 C_4H_{10} 有正丁烷和异丁烷，它们的结构式分别为



同样，戊烷（C_5H_{12}）有三种异构体，而且从实验室也的确分离出来性质各异的异构体，一个不多，一个不少。到底 C_nH_{2n+2} 每种应该有多少种同分异构体呢？60 年代到 70 年代，化学家绞尽脑汁也没有得出一个定论。没想到英国数学家凯雷早在 1857 年已经得出正确的数值。当然，他不是从化学出发，而是从组合数学问题——关于“树”的计数来进行研究的。他的树就是没有回路的点——线图，这样他得到如下的结果：

n 1 2 3 4 5 6 7 8 9 10

异构体数 1 1 1 2 3 65 9 18 35 75

饱和烃的构形只是多种多样构形中的一种。对于一般的构形，计数定理显然是不易得到的。一直到 1937 年，匈牙利著名数学家波尔亚得出一个非常一般的计数定理，几乎包括以前所有的定理。这个定理用到组合理论中的生成函数（也称母函数，发生函数）以及群论，特别是置换群以及表示理论的知识。利用波尔亚计数定理可以求出许多图论的定理，例如，p 点图的数目。所谓 p 点图，也就是由 p 个点以及两个点之间的可能连线构成的图。例如 5 点图，用波尔亚定理可以求出具有 s 条连线的不同 5 点图的数目，它是用生成函数来表示的：

$g_5(x)=1+x+2x^2+4x^3+6x^4+6x^5+6x^6+4x^7+2x^8+x^9+x^{10}$

这样有 5 个点，s 条连线的不同的图的数目就是 x^s 的系数。例如 5 个点 6 条连线的图共有 6 个。实际，我们可以画出来：

由这公式还可以看出，5 点图最多有 10 条连线，而且这种图只有一种。因为这时每对点之间最多只有一条连线，它的数目为这种图在图论中称为完全图。

$$\binom{5}{2}=\frac{5\cdot 4}{4}=10$$

56.环游世界

——哈密尔顿游戏

哈密尔顿是英国伟大的数学家和物理学家，在数学上他发明了四元数，这是复数的推广。在组合学上，他发明了环游世界的游戏。

当然，环游世界也不可能是世界上每个地方都走到，而是选若干个地方，比如说，大城市，如伦敦、柏林、巴黎、莫斯科、纽约等等，为了节约时间和金钱，当然希望每个城市都走到，而且只走过一次，不要来回兜圈子，最后回到原来出发的城市。这样的走法是否能够实现呢？数学家处理问题的方法，不是任意摸索，而是把问题抽象成数学问题。为了确切起见，他找到一个模型，这个模型就是一个正 12 面体。正 12 面体有 12 个面 20 个顶点，30 条棱，每个面都是相同的正五边形。在这个具体模型上，问题就变成找一条通过所有 20 个顶点的一个闭回路。

这样的闭回路的确存在，但是一个点一个点去试就太麻烦了。实际上我们只需在 12 面中选六个面，它们两两各有一棱相连接，这样六个面，共有 30 个顶点，30 条棱，但其中 10 条棱及其 20 个顶点重合，把重合的棱不计，重合的顶点只计算一次，就得到 20 个顶点，20 个棱的一个多边形，而这就是我们要求的封闭回路。

实际上，五种正多面体都可以这样通过顶点环游一周，最后回到原点。为此，我们把这些顶点摊在平面上，然后按照顺序经过所有的点即可。

57.比赛出冠军

——比赛图

数学上一个图是指一些顶点以及顶点间的一些连线（弧）。但在实际应用过程中，这样的图是不够的。在道路网络中有些道路是单行线，生产过程中工序先后是确定的，要加工的器件不能倒流，在比赛过程中如果不允许平局，两队一个胜，一个负，在表示这些关系时，图中顶点关系不是对称的，图中连接顶点的弧也是有方向的，这样的图称为定向图。定向图中如果任两点之间都有一条确定方向的连线，而且没有自己对自己的环路，则这个定向图称为完备图，也称为比赛图，因为它精确表示循环比赛时各队的局势。也就是如甲队胜乙队，我们从甲队画一条弧指向乙队。这样，所有比赛完成之后，这个完备的定向图就精确反映出比赛的结果。

可以想象，随着队数（在图中由顶点表示）的增加，可以出现越来越多不同的局面。n 个队有 2^n 种可能的局势，但是如果不考虑具体的队谁赢谁输，它的图的类型可以大大减少，但数目仍相当可观。

顶点数	比赛图数
1	1
2	1
3	2
4	4
5	12

但 10 个顶点的竞赛图数已超过 900 万。

对于比赛图，人们最感兴趣的是能不能产生一个冠军，更进一步说，能不能各队排成一个顺序，排名在前的队比后面的队更有优势。从经验上看，一般总可以办到，但实际上，并非如此。例如三个队，甲胜乙，乙胜丙，丙胜甲，各得一分。这样如不附加其他条件，就无法决定名次。如果把一个顶点指出的箭头数减去进入的箭头数称为得分，那么每一个顶点都有一个分数，什么情况下出现冠军呢？图论中有一个著名的定理：

如 v 是比赛中得分最多的顶点（冠军），则任何其他顶点 u 与 v 的距离 d（v，u）≤2。这也就是说，对于其他每个队 u，冠军不是直接胜它，就是胜了一个战胜 u 的队。但是这个定理并不排除几个队并列冠军，但却可证明不能只有两个队并列冠军。而且还有这样的情况，除了 p=2，4，两种情况，任何 p 个队比赛，都有可能出现 p 个队并列冠军的情况。

比赛图的一个基本定理是：

每个比赛图都具有一个哈密尔顿道路，即沿一个方向通过每个顶点的有头有尾的道路，因此只要这条道路不是封闭的，则这条道路就决定了各队的名次，无一遗漏。

58.画地图用多少颜色

——四色问题

在绘制地图时，为了区别一个国家与它的邻国，一个省区与它邻近的省区，我们给不同的国（省区）与它的邻近的国（省区）着上不同的颜色。为了节约起见，不相邻的国家不必非得涂上不同的颜色，那么为了绘制地图，至少得用多少种颜色？很容易画出地图，说明一种、二种和三种颜色是不够的。四种颜色够不够呢？这就是四色问题。更确切说，在平面上或球面上绘制地图只需用四种颜色，这也可称为四色猜想。1840 年，德国数学家莫比乌斯首先提出四色猜想，1850 年，英国的一个学生葛斯瑞，经过考虑后认为，四种颜色足够。其后不久，他寄给他弟弟一个证明，但这个证明已经遗失，看来这个证明是不对的。他的弟弟把这个问题告诉英国的几位大数学家，其中凯雷对这问题很感兴趣，但他直到 1879 年宣布，他还不能给出一个证明。

在凯雷宣布后不久,一位英国律师肯普在数学学术杂志上发表了一个“ 证明 ”，但一直到 11 年后，希伍德在 1890 年才指出，肯普的证明中有一个漏洞。实际上正是这个漏洞，使得四色问题又拖了近 100 年之久未能解决。

这 100 年的历史包含着不少点点滴滴进步，一个是希伍德发表一个严密的证明，证明五色足够，从这个方面再改进就是证明四色猜想。因此，数学家选择另外的方向：在国家的数目上加以限制。首先是弗兰克林在 1920 年证明，当国家数目≤25 时，四色定理成立。其后在 1926 年国家数提高到 27，1936 年提高到 31，1943 年提高到 35，最后 1968 年提高到 40。国家数增加得很慢，原因是国家每增一、二个，不同国家之间的边界关系类型就变得复杂得多，而证明的关键就是必须把地图的所有类型都考虑到，不能有所遗漏。

1976 年，阿佩尔和哈肯终于通过计算机的帮助最终证明四色猜想，这可以说是靠计算机证明的唯一的大定理。他们的证明实际上分成两步，首先是

把所有可能的地图类型归结为有限多个不同的类型，具体说是 1936 个。这一步花了六个月时间，第二步是一一证明它们四色足够，这个用了一个月，最终，他们完成这个伟业。

20 年来尽管对于这项计算机证明有着这样那样的非议，人们还是没有找到不依赖计算机的证明。不管怎样，这件事标志着计算机开始成为数学家不可少的工具。

59.有 400 人就有不少人同一天过生日

——抽屉原理

在研究组合问题时，我们有一些最基本的原理，其中最重要的是抽屉原理。抽屉原理也称为鸽洞或鸽笼原理，它非常简单，但是却很有用处。

鸽笼原理最简单的形式是这样的，如果要把 n+1 只或更多的鸽子放进 n 个鸽笼子里，那么至少有一个鸽笼中有两只或两只以上鸽子。值得注意的是，把 n+1 只鸽子放进 n 个笼子里的办法有许多，我们可以把 n+1 只鸽子都放在一个笼子中，也可以每个笼子放一只，最后一个笼子放两只。因此，我们的兴趣不在于怎样放法，而是用来证明一些情况的存在性。例如，一个生日宴会上有 400 人参加，那么一定有不少人同一天过生日，因为一年有 365 或 366 天，但是究竟这 400 人都是一天过生日，还是没有一个人当天过生日（假定过生日的主人不在 400 人之内），鸽洞原理并不能判断。这个看来用处不大的原理能解决不少问题。举例讲，一个边长为 1 的正方形内最多能找到几个点，使这些点之间的距离大于 1/2。碰到这类问题，虽然可以凑出来答案，但是要严格证明，却需要鸽洞原理。方法是把正三角形三边中点作一个小的正三角形，这样正三角形分成四个小三角形，这四个小三角形中点的关系有两方面：

（1）如果两个点在不同的小三角形中，则这两点的距离可能大于 1/2，当然也可能等于或小于 1/2。

（2）如果两个点在同一个小三角形中，则这两点的距离一定小于 1/2。

现在我们有 4 个小三角形，如果有 5 个点，那至少有一个小三角形中有两个点，它们之间距离小于 1/2。因此，我们在正三角形之内最多只能找到 4 个点它们彼此之间的距离大于 1/2。

同样在这个正三角形中.最多找到 9 个点，它们彼此之间距离大于 1/2，类似我们还可以解决更一般的问题。

60.都认识或都不认识

——拉姆赛理论

拉姆赛是位英国的天才科学家，他只活了 26 岁，却对数理逻辑和经济学做出不可磨灭的贡献。而作为研究逻辑的副产品，以他命名的拉姆赛理论已成为组合理论乃至数学的一个重要分支，在各个领域特别是计算机科学中有着重要应用。

拉姆赛的出发点非常奇特，看起来同数学没什么关系。任何一个集会、

聚会或者宴会，参加者都是四面八方来的人，两人可能相互认识或相互不认识。拉姆赛的定理是讲，如果集会的总人数等于或超过 6 个人，那么其中至少有 3 人，这 3 个人互相都认识或者都不认识。但是如果人数少于 6 人，则这种情况不一定出现。

有数学训练的人与没有数学训练的人之间的不同在于前者能把这样一个说起来模糊的问题变成为一个非常清楚的数学问题。6 个人我们可以用 6 个点来代表，而每两人之间的关系只有两种可能。两人相互认识或相互不认识。如果两人认识，我们在代表他们的这两点之间连上一条红线，如果两人不认识，则连上一条蓝线。这样对任何情况，我们就得到一个 6 个点以及每两点之间共有 15 条连线的图。因此，这个定理是说不管两点之间的连线你如何选，那么这个图中一定存在一个三角形，它的三边都是同一颜色，或都是红色，或都是蓝色。

道理很简单，如图，从 A 点出发有 5 条线。这 5 条线涂上两种颜色，无论怎样涂，至少有三条是一种颜色，其实这就是抽屉原理。假设 AB，AC，AD 三条边是红色，我们看 BC，CD，BD 这三条边，当然有这两种可能性：

一种可能是 BC，CD，BD 中有一条红边，例如 BD 是红色，那么 ABD 就是全红三角形。

另一种可能是 BC，CD，BD 中没有一条红边，那它们都是蓝边，这样一来 BCD 就是全蓝三角形。

因此，不管怎么说，总有一个同一颜色的三角形。由于在我们的证明中只用到抽屉原理，所以后来由拉姆赛定理也称为广义抽屉原理。不过我们还得补充两点：如果只有五个点或更少，拉姆赛定理不一定成立。这只要找一个图，没有同色三角形。如果图 2 中的五边形和它中间的五角星是两个颜色，那这图中就没有全色三角形。如果多于六个点，当然拉姆赛定理一样成立。6 就是存在同色三角形的最小数目，称为拉姆赛数，记作 r（3， 3）＝6。

拉姆赛定理只不过是拉姆赛理论的出发点，它已经有了许多推广，但求拉姆赛数是一个极为困难的问题。现在只知道 r（4，4）＝18，也就是只有 18 人或 18 人以上的集会中才一定有四个人互相认识或互相不认识。更大的拉姆赛数尚不知道。

代数

61.符号的重要

——从算术到代数

小学的算术问题在学会代数之后,一般可以迎刃而解,其中的奥妙何在？这里的关键是符号化。通过符号的使用，算术的运算规则凸显出来，它们实际上只是一些机械的操作，除了四则运算之外，还包括移项和消去的法则，而四则运算则遵守加法交换律、结合律，乘法交换律，结合律以及加法、乘法的分配律。有了符号之后，运算完全程序化、机械化，代数成为寻求算法的技术了。

其实近代数学是以符号化为其显著特征，而这特别以近代代数学为其典型代表。早期的近代代数学的产生是两种潮流的结合产物：一是数值计算过

程的符号化，二是数系的拓广，这两个过程都经历了漫长的过程，前者到 18 世纪中期才稳定下来，而后者直到 19 世纪中期仍有一些学者有异议。不过这时以方程论为中心的代数学已发展到比较成熟的地步。

数学史家纳赛尔曼认为代数符号化过程经历了言辞代数、缩写代数、符号代数三个时期。其实任何数学分支的符号化大抵都经历了类似的阶段，而且各阶段往往互相重叠，并不能截然分开。

数学分支的符号化大致可分五个部分：

（1）对象的符号化，首先是数字的符号化。由于印度—阿拉伯数码的引进，并于 15 世纪在欧洲普遍推广，到 16 世纪，现在普遍使用的位值制及十进小数计法也得到广泛应用，这大大推动了计算技术的发展。

（2）运算的符号化，加、减的符号（“＋”“－”）早在 15 世纪中期已经有了，16 世纪已普遍使用。乘、除的符号则稍晚，根号在 17 世纪也已产生。乘幂的符号到 19 世纪初才固定下来。

（3）关系的符号化，等于、大于、小于是 16 世纪中期开始引进的。

（4）区别已知量、未知量、未知量系数的符号，这是韦达的最大功绩。

（5）更一般对象的符号化，如集合、曲线、方程、函数乃至语句的符号比则是比较晚的事，一般到 18 乃至 19 世纪才开始萌芽。

如果只把符号化看成一个以字母代表数字的过程，那就太简单了。实际上符号化包含了计算对象（例如数）可以扩大到什么程度，也就是数系的扩大化的问题，另外还包岔运算的形式化或程序化与规则的公理化。这些不仅可以用于对数、量的计算上，同时也可以应用于更一般的对象上。经过这样的处理，我们可以说一门学科得到代数化。而代数化则直接导向数学的机械化，这开辟了构造数学的新方向。

62.怎样解题

——代数方程和代数方程组

在几千年历史发展的长河中，人们碰到大量的问题，这些问题经过许多人努力，往往能够化为数学问题、从天文问题、航海问题一直到下棋、赌博、经济、金融、运筹、决策、管理等问题，现在都能采取数学的形式，并产生相应数学的学科，这样问题就转化为求解数学问题。许多数学问题，经过已知数及未知数的适当选择就化成代数方程及代数方程组的问题，因此，求解方程就成为解数学问题的最常见的手段，而这构成经典代数学的主要课题。

代数方程的求解问题可以分成多少相互独立的三大领域：

（1）线性方程及线性方程组理论。大量数学问题都可以简化为线性方程组，许多问题已经化为数以千计的变元的大型线性方程组了。现在已经有了完整的求解线性方程组的方法，并形成了包括行列式、矩阵、线性空间及线性变换在内的线性代数理论。

（2）单个代数方程理论。2 次或 2 次以上的代数方程的求解，经历了近 2000 年的发展到 19 世纪才形成较为完整的理论。古代已有解特殊二次方程的各种方法，到 16 世纪才有了解 3 次和 4 次方程的方法，这时自然产生虚数及复数的概念。从 16 世纪到 19 世纪初，数学家寻求 5 次及 5 次以上方程的解法，始终没有成功。到 19 世纪 20 年代，挪威年轻数学家阿贝尔首先证明

一般 5 次及 5 次以上代数方程没有根式解。1830 年左右，法国数学家伽罗华通过置换群理论解决代数方程可解性问题。其后代数方程论沿着四条途径发展：

①求代数方程的近似解。

②求代数方程的解析解。5 次及 5 次以上代数方程一般虽然没有根式解，但是按照代数学基本定理，n 次代数方程总有 n 个根，代数方程的解还是存在的，它不能用根式表示却可以用超越函数的值来表示。例如 5 次方程的根可以用椭圆模函数表出，这些解是精确的，不是近似的。

③判别高次方程的可解性及不可解性。阿贝尔、伽罗华证明，一般的 5 次及 5 次以上方程不可解，并不等于说其中没有可根式解的方程，因此，给出一个方程需要判定其是否根式可解。例如 5 次方程。

$$x^5 - ax - b = 0$$

当 a，b 可被素数 p 整除，b 不能被 p^2 整除，且 $4^4a^5 > 5^5b^4$ 时，这方程不能根式解。同样可证明

$$x^7 - ax - b = 0$$

当 $6^6a^7 > 7^7b^6$ 时不能根式解。

④置换群理论。代数方程的可解性理论引导到置换群理论的研究，其后又引向更一般、更抽象的群论的研究，群论连同域论、环论在 19 世纪末及 20 世纪初导向抽象代数学的产生。

（3）代数方程组理论。这个方向在 19 世纪先以消元法，后以代数几何学的形式发展，到 20 世纪成为当代数学的前沿，正在蓬勃发展之中。从抽象代数的观点看，它构成交换环理论的主体部分。

63.分圆与开方

——虚数与复数

在求解一二次方程时，我们势必要碰到虚数。虚者，不实也。因此，长期以来，许多人不接纳虚数，其中甚至有许多大数学家。但是，虚数的引进不仅在数学上很有必要，而且还有很大的实用价值，例如 19 世纪后半期发展起来的电工学，复数就是开发交流电技术不可少的工具。

数学理论上需要虚数，首先是必须明确 n 次代数方程到底一共有多少根。无论从历史上还是从现实上来看，容许哪些根不容许哪些根一直是有争议的。在 16，17 世纪，甚至到 19 世纪，许多数学家不仅仅对虚根，连负根也不容许。这样一来，同样是二次方程就可能出现二个根、一个根或没有根多种情况。这种偏狭的见解最终还是被宽容的意见所取代。把实根、复根都包括在内，n 次代数方程总是不多不少有 n 个根，这就是代数学基本定理。在这种情况下，数学形成一种完美的理论，不会产生歧见。

最简单的代数方程求解实际上就是开方。开方是乘方的逆运算，但开方与加法和乘法的逆运算——减法和除法不同，它不止有一个解，即开几次方就有几个解。如果没有列出全部解，那就是受到这样那样的限制，因为开方无非是解二项代数方程。

$$x^n - a = 0$$

我们从最简单的开始，求$\sqrt{4}$显然求$\sqrt{4}$相应于解方程$X^2 = 4$，它应该有

两个解＋2 和－2。因此不管一个正数开方出来是有理数还是无理数，它都有两个解，大小相等，一个正，一个负。按照这个规则，就产生一个负数的开方，我们规定。

$$X^2+1=0$$

的解为＋$\sqrt{-1}$，－$\sqrt{-1}$，通常用＋i和－i表示，那么一个负数的开方$\sqrt{-b}$（$b>0$）就等于$i\sqrt{b}$和$-i\sqrt{b}$。

对于开高次方根如$\sqrt[n]{a}$（为简单起见，设$a>0$），我们同样可如法炮制，也分成两步走：一步是求$\sqrt[n]{a}$的正根，一步是求$\sqrt[n]{1}$也就是单位根。求1的单位根有一个简单直观的方法，这就是分圆术。取一个单位圆即以原点为圆心半径为 1 的圆，然后把它分成 n 等分，其中一个分点是＋1，那么另外 n－1 个单位根的值可由它们的坐标得出，如坐标是（X，y），则它的值就是 X＋iy。如图是 1 的 6 次根如下：

$$+1,\frac{1}{2}+\frac{\sqrt{3}}{2}i,-\frac{1}{2}+\frac{\sqrt{3}}{2}i$$

$$-1,-\frac{1}{2}-\frac{\sqrt{3}}{2}i,\frac{1}{2}-\frac{\sqrt{3}}{2}i$$

把单位根乘上$\sqrt[n]{a}$自然就得出$\sqrt[n]{a}$的n个值了。

64.多个变元多个方程

——线性方程组

从算术及代数的简单练习到科学、技术、经济、社会上许多复杂问题，在许多情况下，最后都归结为一个或多个变元代数方程或代数方程的求解。其中最简单的也是最实用的是线性方程组的求解。初等数学中常见的是两个或三个变元的线性方程组，也称为联立线性（或一次）方程组，如

$$a_{11}x_1+a_{12}x_2=b_1$$
$$a_{21}x_1+a_{22}x_2=b_2$$

和

$$a_{11}x_1+a_{12}x_2+a_{13}x_3=b_1$$
$$a_{21}x_1+a_{22}x_2+a_{23}x_3=b_2$$
$$a_{31}x_1+a_{32}x_2+a_{33}x_3=b_3$$

对于这种方程，通常用消去法来求解，这种方法在中国的《九章算术》当中，就已经广泛使用。19 世纪初，高斯把这种方法应用到更多变元的一般线性方程组。

对于 n 个变元 n 个线性方程组

$$a_{11}x_1+a_{12}x_2+\Lambda+a_{1n}x_n=b_1$$
$$a_{21}x_1+a_{22}x_2+\Lambda+a_{2n}x_n=b_2$$
$$\Lambda\ \Lambda$$
$$a_{n1}x_1+a_{n2}x_2+\Lambda+a_{mn}x_n=b_n$$

对这个线性方程组，我们已有系统的解法，即解

$$Xi = \frac{\det Ai}{\det A} \qquad i = 1,2,\Lambda\ n$$

其中 det A 表示系数行列式

$$\det A = \begin{vmatrix} a_{11} & a_{12}\Lambda\ a_{1n} \\ a_{21} & a_{22}\Lambda\ a_{2n} \\ a_{n1} & a_{n2}\Lambda\ a_{nn} \end{vmatrix}$$

detAi表示在系数行列式中，第i列换成$\begin{pmatrix} b_1 \\ b_n \end{pmatrix}$后所得的行列式。行列式有固定的计算机方法，因此，只要 detA≠0，我们就可以求出方程组的解来。

在现代科学技术问题中，如结构分析、网络分析、大地测量、经济分析、运筹及规划问题以及大量数据处理以及非线等方程及微分方程的数值计算中，变元的数目达到几十、几百乃至成千上万，上述解法不太现实，这时要用电子计算机进行数值求解，并且发明了许多方法及技术，主要是直接法和迭代法，并根据具体问题创造出稀疏技术、加速收敛技术等等。

65.数的阵列

——行列式

行列式是线性代数学最主要的计算工具，它是由 n 行 n 列 n^2 个数构成的数的阵列。

$$\begin{vmatrix} a_{11} & a_{12}\Lambda\ a_{1n} \\ a_{21} & a_{22}\Lambda\ a_{2n} \\ a_{n1} & a_{n2}\Lambda\ a_{mn} \end{vmatrix}$$

它称为 n 阶行列式，这 n！个数通过一种固定的方法计算得出一个确定的数，例如：

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & k \end{vmatrix} = aek - afh + bfg - bdk + cdh - ceg$$

n 阶行列式共有 n！个项，每项为 n 个元素的乘积，各项前面的系数一半为正，一半为负。

行列式有许多简单的性质，使得计算方便。

（1）对于行列式中行成立的定理，对于列也成立。

（2）行列式中有两行或两列相同，则行列式等于零。

（3）行列式中某一行加上另一行后，行列式的值不变。

（4）如两个 n 阶行列式｜aij｜，｜bij｜的乘积是 n 阶行列式｜cij｜则

$$cij=\sum_{k=1}^{n} aikbkj$$

一个特殊的行列式是范德蒙行列式

$$\begin{vmatrix} 1 & 1\Lambda 1 \\ x_1 & x_2\Lambda\ x_n \\ x_1^2 & x_2^2\Lambda\ x_n^2 \\ \Lambda\ \Lambda & \\ x_1^{n-1} & x_2^{n-1}\Lambda\ x_n^2 \end{vmatrix} = \pi(x^{j-xi}) \quad 1\le i\le j\le n$$

行列式的用处很多，首先是解 n 元 n 次的线性方程组。其次在解析几何中可以方便地求解几何问题，例如过（X_1，y_1），（X_2，y_2）的直线方程为

$$\begin{vmatrix} x & y & l \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0$$

三个不同直线 $a_iX+biy+ci=0$ i＝1，2，3 交于一点的充分必要条件是

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = 0$$

以（x_1，y_1），（x_2，y_2），（x_3，y_3）三点为顶点的三角形的面积等于

$$\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & l \end{vmatrix}$$

66.投入与产出

——矩阵

矩阵是是线性代数的核心，但在历史上出现很晚。矩阵这个词首先由西尔维斯特在 1850 年使用的，为了与行列式的数列区别开。现在意义下的“矩阵”是 1857 年由凯雷引进的，目的是为了表达方程系数。用方阵表示线性交换的思想，高斯在他的二次型理论中就已提出，他还提出把两个线性交换结合成第三个线性变换（也就是矩阵的乘法）的思想。

如

$$A=\begin{bmatrix} a_{11} & a_{12}\Lambda\ a_{1n} \\ a_{21} & a_{22}\Lambda\ a_{2n} \\ \Lambda\ \Lambda & \\ a_{n1} & a_{n2}\Lambda\ a_{nn} \end{bmatrix} \quad B=\begin{bmatrix} b_{11} & b_{12}\Lambda\ b_{1n} \\ \Lambda\ \Lambda & \\ \Lambda\ \Lambda & \\ b_{n1} & b_{n2}\Lambda\ b_{nn} \end{bmatrix} \quad AB=C=\begin{bmatrix} c_1 & \Lambda\ c_{1n} \\ \Lambda\ \Lambda & \\ \Lambda\ \Lambda & \\ \Lambda\ \Lambda & \\ c_{n1} & \Lambda\ c_{nn} \end{bmatrix}$$

则$c_{ij}=\sum_{k=1}^{n}a_{ik}b_{kj}$

高斯这种符号代数后来为爱森斯坦所发展，他在自己的研究中，认识到线性变换可以相加、相乘，可以求逆及幂，尤其是他注意到乘法不交换，这与行列式有原则不同。

矩阵作为一个运算对象，与数最的运算有很大的不同，量主要的差异是乘法一般不可交换，也就是不满足乘法交换律，即

AB＝BA

不一定成立。例如：

$$\begin{pmatrix}0&1\\0&0\end{pmatrix}\begin{pmatrix}0&0\\1&0\end{pmatrix}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$$

$$\begin{pmatrix}1&1\\0&0\end{pmatrix}\begin{pmatrix}0&0\\1&0\end{pmatrix}=\begin{pmatrix}0&0\\0&1\end{pmatrix}$$

其次乘法消去律不成立，即

AD＝AE

不一定推出 D＝E，例如

$$A=\begin{pmatrix}0&0\\1&0\end{pmatrix}\quad D=\begin{pmatrix}1&0\\0&0\end{pmatrix}\quad E=\begin{pmatrix}-2&0\\2&0\end{pmatrix}$$

$$AD=\begin{pmatrix}0&0\\1&0\end{pmatrix}=AE\text{但}D\neq E$$

还有非零因子存在，即 A≠0，B≠0，但 AB＝0，例如

$$\begin{pmatrix}0&1\\0&0\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix}=\begin{pmatrix}0&0\\0&0\end{pmatrix}$$

矩阵在现代科学技术中是不可少的工具。在量子力学、电网络分析、大型结构分析、工业工程、统计分析、运筹学以及投入产出分析等领域是必不可少的。

67.根的置换

——置换与置换群

在代数方程论中最重要的概念是置换及置换群。当时处理的对象是根的置换，由于 n 次方程有 n 个根，因此不妨把 n 个根排列成一定的顺序，可以用（1，2，3，……，n）来表示。所谓一个置换，就是把这个排列换成另外一种排列，例如（2，3，1，5，4…n，n－1）。一般来讲，不管是 n 个根也好，n 个事物也好，n 个字母也好，我们关心的不是它们某种特殊的排列，而是把一种排列换成另一种排列的动作或操作，这个动作或操作就是置换。数学中我们通常研究的对象是数和形，置换的引进可以说是一个特殊的研究对象。

正如研究其他对象一样，首先研究一下置换的表示。由于置换不是一个排列，而是把一个排列变成另一个排列，因此，置换无非是把两个排列放在

一起，即

$$\begin{pmatrix} 1,2,3\Lambda & \Lambda, n-1, n \\ 2,3,1\Lambda & \Lambda, n, n-1 \end{pmatrix}$$

现在置换成为我们的研究对象，那我们就可以定义两个 n 元置换的乘法，所谓乘法，无非是先做一次置换后接着再做一次置换，两次置换以后也相当一个置换，这就是这个置换的乘积。具体讲，我们讨论三元置换

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

它们的乘积怎么算。很简单，第一个置换把 1 变成 2，第二个置换又把 2 变成 1，两个置换连续作用（即相乘）的结果是把 1 还变成 1。这样一来，两个置换相乘的积是一个置换。

$$\begin{pmatrix} 1,2,3 \\ 1,2,3 \end{pmatrix}$$

这个置换使排列仍变成自身，称为单位置换，它的作用如同数乘法中的 1 一样。

有了置换乘法的概念之后，我们把所有 n 元置换放在一起，构成一个集合，对于 n 元置换，显然共有 n！个元素。这个集合称为对称群，因为它满足群的四个公理。

（1）群里存在一个二元运算——乘法，两个置换的乘积仍是群里的一个置换。

（2）这个乘法满足结合律，但是不一定满足交换律，例如

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}=\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}=\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix}$$

结果是完全不一样的。

（3）群中存在单位元，即单位置换，它与任何置换之积还是那个置换，所有它的作用就同乘法中的 1 一样。

（4）群中每个元素都存在逆元素，也就是一个置换和它的逆置换相乘等于单位置换，例如。

$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$的逆置换就是$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$

对称群中的一部分元素如果也满足这四条公理，也构成一个群，这些群就称为置换群。

68.伽罗华的悲剧

——高次方程的根式解

伽罗华是数学史上的传奇人物，他是一位天才，20 岁就同人决斗而丧生。生前他用置换群理论证明一般五次及五次以上方程不能用根式解，但是

这个天才的思想却长期得不到当时权威们的理解。一直到几十年后，才逐渐得到人们的认识，他的理论也逐渐发展成为抽象的群论，成为现代数学的基石之一，而且在物理、化学等领域有着重要的应用。

代数方程是最古老的方程，有着重要的应用，古代就已知一次、二次代数方程的解法，16 世纪也发现三次、四次代数方程的根，它们都可以表示为方程系数（通常为有理数）的加、减、乘、除以及开方来表示，这样表示的根称为方程的根式解。很自然，人们就寻求五次及五次以上的方程求根式解的算法。但是，经过 300 多年的努力，未获成功。这时就已经有人意识到，可能五次及五次以上方程不能根式解，但是这样一个大定理，需要严格的证明，这个证明最终被挪威年轻的数学家阿贝尔给出来。但是，这个结果是消极的，它只是说，不必再徒劳无功地去寻求五次及五次以上方程的用根式表根的一般公式了，但是没有给数学增加太多新内容。伽罗华的理论则不一样，它产生许多积极的成果，开辟了完全崭新的代数学领域。

首先，伽罗华把每个有理系数代数方程同一个置换群连系在一起，这群称为该方程的伽罗华群，并且证明伽罗华理论基本定理，即有理系数代数方程可以根式解当且仅当该方程的伽罗华群是可解群。这样，他不仅证明一般五次及五次以上方程不能根式解，而且也指出不是所有的五次及五次以上方程都不能根式解，有的也可以根式解，而且给出这些方程可根式解的条件。

为什么一般五次及五次以上代数方程不能根式解呢？这是因为一般五次方程的根的置换群是 5 元对称群，它共有 120 个元素，它包含一个 5 元交错群，它共有 60 个元素，这个群是单群不能再分解成简单的群了，因此不是可解群，所以一般五次方程不可根式解。但如果某个五次方程的置换群只是 5 元对称群中一个小的子群，那它就是可解群，从而就可以根式解了。例如

$x^5+x-a=0$

当 a 为 2 或 6 时可根式解，而 a＝3，4，5，7，8，9，10，11 时不可根式解。

其次，伽罗华指出研究群及群的结构的重要性，这样开辟了现代抽象群论及抽象代数学的广阔天地。

69.1＋2＋…＋100

——等差级数及其他

小学生都知道大数学家高斯的故事：高斯小时候，老师出了一道题，求 1＋2＋3＋…＋100＝？他的同学一步一步傻算时，高斯已把答案 5050 写在小石板上。高斯的方法是，首尾相加，1＋100＝101，第二项和倒数第二项相加，2＋99＝101，第三项和倒数第三项相加 3＋98＝101，如此下去，最后 50＋51＝101，这样就得到 50 个 101，这就是答案。实际上，高斯的方法并不限于这个特殊的级数。实际上，对于任何等差级数都适用。所谓等差级数，就是不管那一项，它与前一项之差都是相等的，这个差称为公差，记作 d，我们通常这样表示等差级数 $a_0+a_1+a_2+\cdots+a_{n-1}$，n 称项数，它满足：

$$a_k-a_{k-1}=d \quad k=1,\cdots,n-1$$

由此可以得出等差级数的各项之间的关系：

$$a_k = a_0 + (k-1)d$$

$$a_k = \frac{a_{k-1} + a_{k+1}}{2}$$

为了用高斯的方法求和，我们可以验证一下：

$$a_0 + a_{n-1} = a_1 + a_{n-2} = a_2 + a_{n-3} = \Lambda\ \Lambda$$

实际上它们都等于 $a_0 + a_0 + (n-1)d$。当 n 为偶数时，等差级数的和 Sn 为

$$Sn = \frac{n}{2}(2a_0 + (n-1)d) = na_0 + \frac{n(n-1)}{2}d$$

当 n 为奇数时，这个公式也成立。高斯的级数中，首项 $a_0 = 1$，公差 $d = 1$，项数 $n = 100$，结果 Sn 也是 5050。而这个公式则对任意首顶、任意公差、任意多项式都成立，这就是代数的一般公式的威力，用这个公式可以证明：

$$1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$$

$$2 + 4 + 6 + \cdots + 2n = n(n+1)$$

$$1 + 3 + 5 + \cdots + (2n-1) = n^2。$$

另一个常用的级数求和公式是等比级数

$a_0 + a_1 + a_2 + \cdots + a_{n-1} = a + aq + aq^2 \cdots + aq^{n-1}$ q 称为公比，因为

$$\frac{a_k}{a_{k-1}} \quad k = 1, \Lambda\ , n-1$$

对等比级数，我们有

$$a_k = aq^{k-1}$$

$$a_k = \pm\sqrt{a_{k-1}a_{k+1}}$$

$$S_n = \frac{a(1-q^n)}{1-q} = \frac{a - aq^n}{1-q}(q \neq 1)$$

而当 n→∞时，等比级数成为无穷级数，而这只有当 q＜1 时，才有和。

$$S = \frac{a}{1-q} \quad (q < 1)$$

70.$1^2 + 2^2 + \cdots + 100^2$

——等幂和问题

除了等差级数和等比级数之外，一般的有限的级数求和问题非常困难，一般没有好方法。但是，当数学家得到一个一般的求和公式之后，证明它却不难，这时的代数公式的证明工具就是数学归纳法。

历史上最重要的求和问题是等幂和问题，也就是求

$$1^2 + 2^2 + \cdots + 100^2$$

$$1^3 + 2^3 + \cdots + 100^3$$

更一般地求

$$S_k = 1^k + 2^k + \Lambda + n^k$$

经过经年累月的探索，最终得出下列一般公式：

$S_2 = \frac{1}{2}n（n+1）（2n+1）$

$S_3 = （\frac{1}{2}n（n+1）^2$

$S_4 = \frac{1}{30}n（n+1）（2n+1）（3n^2 3n-1）$

$S_5 = \frac{1}{12}n^2（n+1）^2（2n^2+2n-1）$。

到 17 世纪，费尔马等人发现得出这些公式的一般方法，他主要用

$$(m+1)^{k+1} - m^{k+1} = (k+1)m^k + (\frac{k+1}{2})m^{k-1} + \cdots\cdots + 1$$

从 S_{k-1} 求 S_k。例如要求 $1^2+2^2+\cdots+n^2$，用

$(m+1)^2-m^3=3m^2+32m+1$

然后两边求和

$$\sum_{m=1}^{n}(m+1)^3 - \sum_{m=1}^{n}m^3 = 3\sum_{m=1}^{n}m^2 + 3\sum_{m=1}^{n}m + n$$

从而得出 $(n+1)3-1 = 3\sum_{m=1}^{n}m^2 + \frac{3}{2}n(n+1) + n$

这样就得出 S_2 的公式。

17 世纪中，帕斯卡更得出一般的证明方法——数学归纳法。简单的数学归纳法可以如下操作：

（1）验证公式对 1＝1 时成立。

（2）证明公式对 1＝n－1 时成立，则 1＝n 时也成立。例如要证 S

$S_2 = \frac{1}{6}n（n+1）（2n+1）$

（1）当 1＝1 时显然成立。

（2）设 1＝n－1 时公式成立，即

$$1^2 + 2^2 + \cdots\cdots + (n-1)^2 = \frac{1}{6}(n-1)(2n-1)$$

则

$$1^2 + 2^2 + \cdots\cdots + (n-1)^2 + n^2$$

$$= \frac{1}{6}(n-1)n(2n-1) + n^2 = \frac{n}{6}[(n-1)(2n-1) + 6n]$$

$$= \frac{n}{6}[2n^2 - 3n + 1 + 6n] = \frac{1}{6}n(n+1)(2n+1)$$ 证毕

同样，我们可以得出奇数的等幂和公式

$$1^2 + 3^2 + 5^2 + \cdots\cdots + (2n-1)^2 = \frac{1}{3}n(4n^2 - 1)$$

$$1^3 + 3^3 + 5^3 + \cdots\cdots + (2n-1)^3 = n^2(2n^2 - 1)$$

71.项数无穷的困惑

——无穷级数

对于有穷多个数，进行加加减减，我们最终总可以得到一个唯一正确的答案，而且加减法服从加法交换律和结合律，也就是最终的答案与运算的顺序无关。你可以从头计算，也可以颠倒顺序，也可以在中间的几项上加上括号先行计算，无论怎样，答案都是相等的。

可是当项数普成无穷时，无穷多个数加加减减的问题就是无穷级数和求和问题，这时麻烦就来了，例如说，求无穷级数

1－1＋1－1＋1－1＋……

的和，如果把它写成

（1－1）＋（1－1）＋（1－1）＋……

它就变成 0＋0＋0＋……，当然得 0 了，要是把它写成

1－（1－1）－（1－1）……

它就变成 1－0－0－0－……，这就是得 1。引进 x 之后，算术就变成代数，无穷算术当然也变成无穷代数，由中学代数就知道：

$(1+x)(1-x+x^2-x^3+x^4-x^5+\cdots\cdots)=1$

所以

$$\frac{1}{1+x}=1-x+x^2-x^3+x^4-\cdots\cdots$$

当 x＝1 时，右边就变成我们要求的无穷级数，而左边变成 1/2，这就是我们的第三个答案：

$$\frac{1}{2}=1-1+1-1+1-1+\cdots\cdots$$

18 世纪最伟大的数学家欧拉换一种表示，也得到同样的结果，他用的公式是

$$\frac{1}{1-x}=1+x+x^2+x^3+\cdots\cdots$$

当 x＝－1 时，就得

$$\frac{1}{2}=1-1+1-1+1-1+\cdots\cdots$$

因此，无穷即使在算术阶段也产生麻烦。

麻烦还没完，在欧拉公式中，代入 x＝2 就得出

－1＝1＋2＋4＋8＋……

另一方面，概括公式或常识

∞＝1＋2＋3＋4＋……

当然 1＋2＋4＋8＋…＞1＋2＋3＋4＋…，所以－1＞∞。根据这点，欧拉说，无穷大（∞）就好像 0 一样，把正数和负数隔开，过去是正数大于 0，0 大于负数，现在是负数大于无穷大，无穷大又大于正数，所以只要你认为无穷大是个数，就会得出这种似是而非的结论。

因此，有穷级数的运算规律不能任意推广到无穷级数上，要想处理无穷级数还必须谨慎从事，首先要考虑无穷级数的收敛与发散。

72.比较就能判别

——收敛与发散

无穷多项级数求和与有限多项求和有很多差别，不仅许多问题不能求和，而且加法交换律及结合律也不成立，但是是否无穷级数不能求和呢？

无论是中国还是希腊，似乎都有原始的极限观念。中国的“一尺之捶”与芝诺悖论都涉及无穷级数

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots\cdots$$

之和等于 1。亚里斯多德也认为这种公比小于 1 的几何级数有和。阿基米德在求抛物线的弦截面积实际上是求无穷级数的和

$$1+\frac{1}{4}+\frac{1}{4^2}+\frac{1}{4^3}+\cdots\cdots=\frac{4}{3}$$

但是，希腊人谨慎地避开无穷，他们的证明都是用有限和严密地证明。

中世纪的神学家开始议论无穷，其中最杰出的思想家是奥雷姆，他明确几何级数有两种可能性，当公比≥1 时，无穷几何级数有无穷和，而当公比≤1 时，有有限和。

无穷研究的这个神学阶段实际上对无穷研究进了一大步。也就是证明无穷多有限量的和不一定是无穷，也就是有限量也可分解为无穷个有限量之和。更进一步，$a_n\to 0$ 只是级数收敛的必要条件而非充分条件。

不过一直到 19 世纪初,数学家才明确意识到不能够任意的去求无穷级数的和，要想求和，首要问题是判定级数收敛还是发散。只有当级数收敛时，才能谈到求和。这样一来，对于无穷级数找出收敛的判据非常重要。为了简单起见，先考虑单调递减的正项级数，也就是每项都是正的，而且后项总比前项小，且 $n\to\infty$时，$a_n\to o$。根据经验，我们能通过比较来判别。假如有两个无穷正项级数

$a_0+a_1+a_2+\cdots+a_n+\cdots$

$b_0+b_1+b_2+\cdots+b_n+\cdots$

如果 $a_0\le b_0$。$a_1\le b_1$，…，$a_n\le bn\cdots$

而且 Σbi 收敛，那么显然 Σai 也收敛。同样，还有比值判别法：

如：

$$\frac{a_n}{a_{n-1}}\le\frac{b_n}{b_{n-1}}$$

若 Σb_n 收敛，则 Σa_n 收敛。

我们最熟悉的无穷级数是等比级数

$a+aq+aq^2+\cdots\cdots+aq^n+\cdots\cdots$

当$q<1$时收敛，而且和为 $\frac{a}{1-q}$，而当$q\ge 1$时发散。以等比级数为判断标准，我们首先得到 Σai 的收敛判据：

（1）如果$\sqrt[n]{a_n}\to 1<1$，则 Σai收敛

（2）如果 $\frac{a_n}{a_{n-1}}\to 1<1$，则 Σai收敛。

73.对称无处不在

——群与群论

群是数学中第一个抽象概念，虽然其观念可追溯到很久以前，但直到19世纪末，数学家才对它有比较清楚的认识。从那时起，几乎所有大数学家无不强调其重要性，把它看成为数学的中心概念以及统一数学的基础。挪威著名的数学家李说：“群和不变式的概念，每日每时在数学中都占有优越地位，而且越来越趋于支配整个这门科学领域。”外尔说“没有群就不可能理解现代数学”。

群论是当代数学最重要的分支之一，也是发展得比较完整而成为现代抽象代数学的模式的一门学科，在伽罗华后的100年间，群论的研究大致有彼此多少重叠的五个方面：

（1）抽象群的定义及公理化

（2）具体的群的构造、刻划及分类

（3）抽象群的具体表示

（4）有限群的列举

（5）一般的结构定理

抽象群，它与具体群不同之处有三：一是摆脱掉具体的作为变换对象的集合，群的元素不一定是具体的变换、置换、运动、作用、运算或数，而是抽象的符号，元素的数目不加限制。二是用公理来刻划。三是各种同构的具体的群只不过是同一抽象群的不同表示，对此我们不加区分。

群的定义一个具有二元运算（如乘法×或加法＋）的集合称为群，如果它满足下面四条公理：

公理1（封闭性），集合中任何两个元素a，b相乘，其乘积也属于这个集合。

公理2（结合性），乘法结合律成立

（a×b）×c＝a×（b×c）

公理3（存在单位元素），集合中存在一个单位元素1，它满足，对于集合中任何元素a，有

a×1＝1×a＝a

公理4（存在逆元素），对于集合中每一个元素a，都存在集合中一个元素a^{-1}，使得

$a\times a^{-1}=a^{-1}\times a=1$

1981年，群的结构定理取得了重大突破，单群的分类最终完成，但整个证明需要上万页，尚未全部发表。

74.更上一层楼

——四元数

到了1830年左右，人们对于复数有比较清楚的认识，它们也满足人们熟悉的有理数的加减乘除等性质。由于人们生活在三维空间中，所以哈密尔顿

希望能找到一个三元数，也具有普通实数和复数的性质。如果能找到这样的数，那表示物理学中的速度和力以及空间的坐标就非常方便了。

哈密尔顿的出发点是复数，他虽然知道并且使用复数的几何表示，但他还是愿意把复数 a＋bi 表示成数对（a，b）。他向自己提出的任务就是找三重数（a，b，c），使它的乘法类似于复数的乘法，也就是要求

（a＋bi＋cj）（x＋yi＋zj）仍然是一个向量，并且满足

（1）可以逐项相乘。

（2）积向量的长度等于因子向量长度之积。他称这规律为“模律”。

为此，他曾经尝试过各种情形，至少复数情形应满足模律，因此必须假定

$ii=-1$

因此必需令

$jj=-1$

那么 ij 与 ji 有什么关系呢？他试过 $ij=1$ 或 $ij=-1$，结果乘积系数的平方总不等于

$(a^2+b^2+c^2)(x^2+y^2+z^2)$

他又试 $ij=0$，也不行，经过长期反复思考，始终未能成功。

1843 年 10 月 16 日，当他和他的夫人在都柏林散步来到布勒安桥时，他突然产生了灵感，于是四元数就降临了人世。实际上，他发觉满足所有普通运算规律的三元数是不存在的，所以新的数至少要包含四个分量。他在桥上想到的就是：除了复数单位的 i 之外，还要加上两个新的虚数单位 j，k，而且 i，j，k 之间要满足一些基本的方程，即

$ij=k=-ji$

$jk=i=-kj$

$ki=j=-ik$

$i^2=j^2=k^2=ijk=-1$

四元数可以像复数一样，表示为 a＋bi＋cj＋dk 的形式，四元数之间可以进行加减乘除，而且满足加法交换律、加法结合律、乘法给合律和分配律，唯一美中不足的是乘法不能满足交换律，所以它和复数相差并不太多，而且它还满足模律。

哈密尔顿对他的四元数热情极高，他相信四元数和微积分同等重要，会成为数学、物理中的重要工具。他本人也把四元数应用到几何学、光学和力学上。但是四元数的使用仍然有阻力，就是因为乘法不满足交换律，在运算上也很不方便。虽然有一些人支持四元数，但是随着向量的出现，四元数的地位大大地削减了。但是四元数却给代数学开辟了一个新的方向，也就是对于超复数的研究，即包含几个分量的“数”或称 n 无数。这种超复数最后导致结合代数乃至环的研究。

75.数学中的新家族

——抽象代数

现代代数学即抽象代数学是在 19 世纪末到 20 世纪初发展起来的。1930 年到 1931 年，范·德·瓦尔登的《现代代数学》一书问世后，抽象代数学成

为现代代数学的主流。二次大战后，自然而然、堂而皇之地去掉“抽象”及“现代”的字眼了。

与古典代数学不同，抽象代数学研究代数结构，即一个集合上的元素及子集合之间的关系。它的目标是对特定的代数结构进行刻划及分类。它所研究的对象已经完全不同于古典代数学那种符号的演算以及求方程的解之类的问题了。但是，它的对象仍然从古典代数学及几何学可以看到其来源。

现代代数学的对象一般有两个来源，一是由具体的数学对象中产生的，如群、环、域、可除代数、交换代数、结合代数、李代数、布尔代数等等；二是由抽象的数学对象衍生出来的，如整环、半群、拟群、圈、广群、具单元半群、半环、近环、交错代数、幂结合代数等，它们是由已有的结构经公理的增减及改动得出来的结构。

现代代数学的具体对象来自多种学科，主要是方程论、代数数论、代数几何学、不变式论、四元数论、超复数论、几何学、逻辑等。

抽象代数学最重要的领域除了群论之外，还有域论及环论。

伽罗华不仅是群论而且是域论的创始人，只不过他没有建立抽象域的观念。他的域是指由 n 个数 a_1，a_2，…a_n，经过加，减，乘，除（零不做除数）后所得到所有数的集合。这种域并不新鲜，有理数全体、实数全体、复数全体就是域的好标本。不过伽罗华域是一种由有限多个元素构成的域，它的元素就不是一般的数了。在 19 世纪，除了知道上述几种域之外，还对于代数数域和代数函数域进行许多深入的研究，形成了代数数论和代数函数论的两大分支。

1910 年德国数学家施泰尼茨对于域论进行统一的抽象处理，形成域论的基础。代数数域理论在 19 世纪末已由希尔伯特总结，他提出的一些猜想陆续由日本数学家高木贞治和奥地利数学家阿丁所解决，形成一个完整的体系。

抽象代数学中最深刻的一部分是环论。其中“代数”（一种特殊的环）理论还在 19 世纪就已经发展起来。19 世纪，复数在数学中所起的举足轻重的作用，给人留下深刻的印象。复数可以看成一对实数，它们可以加，减，乘，除，也能开方，自然就使大家去思考把它推广的问题。爱尔兰数学家哈密尔顿在 1843 年发现了四元数，也能加，减，乘，除，只是乘法不服从交换律，也就是和一般的数不大一样。英国数学家凯雷在 1845 年引进了八元数，可是乘法连结合律都不满足，也没有一定的除法了。这些是后来结合代数和非结合代数的前身。长期以来对于结合代数（也称超复数）进行深入的研究。摩林在 1900 年证明：复数域上维数≥2 的单（结合）代数都和复数域上适当阶数的矩阵代数同构。美国数学家魏德本在 1907 年得出结构定理：结合代数分解为幂零代数及半单代数。而半单代数可以表为单代数的和。而单代数可表为域上某个可除代数的矩阵代数。关于可除代数的研究有许多算术上的困难。为此内特引进交积的概念。由此内特等人证明“主定理”：代数数域上任何有限价中心单代数都是循环代数。所有交积都是中心代数。反过来，中心单代数是否都是交积，这到 1972 年才由以色列数学家阿米祖尔举出反例。

改变代数学面貌最显著的分支是交换环论中的理想理论。理想是戴德金发明的，他为了再建代数数域中不再成立的“唯一分解成素因子”定理，引进了适当的理想。代数几何学中也出现这种理想。希尔伯特研究不变式的主要工具之一是希尔伯特基定理，它断言每个多项式理想具有有限基。多项式的理想与代数数域的理想怎么样一起来处理呢？内特就是使用公理化方法发

展了一般的理想论，她提出了链条件，从一个理想出发，一个理想包含在另一个理想中间，这种理想的升链一定到有限步终上。理想都满足这样的条件的环称为内特环。内特环已有丰富的理论，特别是其中任何理想都可表为准素理想的交。交换代数的发展给代数几何学奠定了稳定的基础。法国数学家韦依在 1946 年的著作《代数几何学基础》中一扫过去不严格的地方，把几何建立在可靠的代数基础上。

1930—1931 年间，荷兰数学家范德瓦尔登的《近世代数学》出版，从此，以群、环、域为中心的抽象代数学成为代数学的中心课题，同时它也构成现代数学的基础，对现代数学的发展有着举足轻重的影响。

几何

76.数与形的联姻

——坐标

解析几何学的诞生是数学思想的一次飞跃，它代表形与数的统一，几何学与代数学的统一。解析几何学的基本内容是：

（1）引进坐标，使点（乃至更一般的几何对象）与数对应。

（2）使方程与曲线（或曲面）等相互对应。

（3）通过代数或算术方法解决几何问题，反过来对于代数方程等给出几何直观的解释。

由于几何学的代数化或算术化大大扩展了几何学的研究领域并弥补了综合方法的不足，为后来的数学发展指出一条康庄大道。

朴素的坐标观念在古希腊甚至古埃及就已经有了，经纬度观念也早就有了。阿波隆尼乌斯在研究圆锥曲线时也使用过坐标。希拔楚斯已开始对天球上的点引进坐标。而在中世纪欧洲，奥雷姆在 1350 年左右引进直角坐标系的原始形式。但坐标轴一直到 17 世纪中叶才引入。到 17 世纪末斜角坐标系才普遍使用，这时其余坐标系也在考虑。牛顿在 1671 年写成的《解析几何》的手稿，除了以直线为参照系的斜角坐标系及直角坐标系之外，还提出另外八种坐标系，其中包括极坐标系及双极坐标系，并利用它们研究一系列几何及微积分问题，特别用极坐标研究螺线。该书出版过迟，雅各、伯努利已于 1791 年正式首次发表极坐标系，并用来研究费尔马的抛物螺线。瑞士数学家赫尔曼则于 1729 年正式宣告用极坐标研究轨迹同笛卡儿坐标一样好。他用 ρ、$\cos\theta$、$\sin\theta$ 为变元，分别用 z、n、m 表示。他还给出由直角坐标转换成极坐标的一般公式。

最常用的平面解析几何的坐标变换公式为直角坐标（x，y）与极坐标（ρ，θ），的相互变换，其公式如下：

$$\begin{cases} x = \rho\cos\theta \\ y = \rho\sin\theta \end{cases}$$

及

$$\begin{cases} \rho = \sqrt{x^2 + y^2} \\ \tan\theta = \dfrac{y}{x} \end{cases}$$

77.曲线与方程

——解析几何

虽然坐标观念很早就有，但有意识地把曲线与方程对应起来和用代数方法解几何问题则是解析几何学诞生以后的事。费尔马首先提出解析几何学的一般原理：“只要在最后的方程里出现了两个未知量，我们就得到一条轨迹，其中一个未知量的端点就描绘出一条直线或曲线”，他接着说“直线是唯一的，而曲线则是无穷多，包括圆、抛物线、椭圆等等”。

解析几何学的出现标志着方法上的重大进步，用以前的综合方法，每一个问题都要用特定的方法去求解。在许多情况下要求有高度的智慧和技巧。而解析几何学的方法，用统一的语言来表述对象，把几何问题化为同一种形式——可以处理的代数形式，这样可以通过标准化的代数程序使得几何问题顺利得到解决。这样为解几何问题提供了一种普遍的方法，而且是一种机械化的方法。这样一来，能解的问题大大增加。例如：

（1）通过计算来解决各种作图问题。

（2）求出具有某些性质的曲线（或曲面），特别是能够解决各种轨迹问题，而且通过曲线、曲面等几何对象的方程可以确认它到底可归入哪一类。

（3）可以发现或证明新定理，这在笛卡儿时代已有所表现。近年来，我国著名数学家吴文俊教授已把这方法推向一个新的高度，并推广到极为广阔的领域。

（4）反过来，通过代数和几何的语言互译，也有利于用几何方法解某些代数问题，如方程求根的问题。不仅如此，解析几何学使微积分大大简化，而且扩大了数学研究范围：如提出作图可能问题、曲线和曲面的分类问题、奇点问题等，大大推动代数的发展，并衍生出代数几何学及微分几何学这些直接后裔。

费尔马利用由距离表示的两个未知量，求出一些曲线的轨迹的方程，用现代的写法就是：

（1）过原点的直线方程 $\dfrac{x}{y} = \dfrac{b}{a}$

（2）任意直线的方程 $\dfrac{b}{d} = \dfrac{a - x}{y}$

（3）圆的方程：$b^2-x^2=y^2$

（4）椭圆方程：$a^2-x^2=ky^2$

（5）双曲线方程：$a^2+x^2=ky^2$

（6）双曲线方程：$xy=a$

（7）抛物线方程：$x^2=ay$

其后他和笛卡儿等人还引进更高次曲线并定出其方程。

78.形形色色的曲线

——代数曲线与超越曲线

自古以来，人们发现各种曲线，从数学上讲，解几何三大问题带来许多圆锥曲线以外的曲线。解析几何出现了以后，发现它们的方程各式各样，有的方程是高次代数方程，有的方程却表示不成多项式。因此。后来把这两类曲线分别称为代数曲线及超越曲线。不过，通常这是按照直角坐标来区分的，而代数曲线的极坐标往往也是三角函数。例如，蔓叶线在直角坐标系的方程为

$$y^2=\frac{x^3}{2a-x}$$

而在极坐标系的方程为

$r=2a\sin\theta\tan\theta$

蚌线的方程为

$(y-a)^2(x^2+y^2)=k^2y^2$

而割圆曲线及螺线则是超越曲线，割圆曲线的方程为

$$y=x\tan(\frac{xy}{2})$$

阿基米德螺线在极坐标方程是

$r=a\theta$

解析几何学出现以后，一下子曲线增加了不少，这是由于我们只需在方程上动一点手脚，而不必像过去那样一定要找到作图的方法。例如，费尔马稍稍改动等边双曲线及抛物线的方程，就得到一大批高次曲线如：

$x^my^n=a$

$y^n=ax^m$

解析几何的另一个优越性是把许多来源各异的曲线用统一的方法来处理，最典型的是卡西尼卵形线族

$(x^2+y^2)-2a^2(x^2-y^2)=k^4-a^4$

当 $a^2=k^2$ 时，就是著名的伯努利双纽线。

伯努利双纽线还可以纳入另一系列曲线族，它们的极坐标方程是

$$r^2=a\cos n\theta$$

其中 n 是有理数。

n	曲线
−2	等边双曲线
−1	直线
$-\frac{1}{2}$	抛物线

$-\frac{1}{3}$ 车恩豪斯三次曲线

$\frac{1}{2}$ 心脏线

1　圆

2　伯努利双纽线

超越曲线中最显赫的家族是螺线，它们的极坐标方程是

$r=f(\theta)$

其中 f 是单调增加函数，其中包括

阿基米德螺线　$r=a\theta$

对数螺线　$r=ke^{a\theta}$

双曲螺线 $r=\frac{a}{\theta}$

等。其他最常见的超越曲线有正弦曲线、指数曲线及对数曲线。

79.怎样分类

——曲线的分类

解析几何学的优越性之一在于对于形形色色的不同来源、不同性质的曲线用统一的代数方法加以处理，这样我们对曲线有一个统一的认识及分类。

在区分平面曲线方面，第一步是分为代数曲线及超越曲线，代数曲线最主要的分类依据，显然是定义曲线方程的次数。因此，对于每个次数，我们可以进行更细致的分类。一次曲线最为简单，它只有一类，即直线。

二次曲线情况就复杂得多，也正是：二次曲线显示解析几何学的威力，一二次曲线一般的定义方程为

$ax^2+2bxy^2+cy^2+2dx+2cy+f=0$

通过坐标系的平移及旋转，可以化简方程，使方程变成比较标准的形式。但是在坐标变换的过程中，存在着不变式，它反映曲线的本质，而不受坐标的线性变换的影响，因此根据不变式就可以对曲线进行分类。

二元二次方程有三个不变式：

$$I_1=a+c,\ I_2=\begin{vmatrix}a & b\\ b & c\end{vmatrix},\qquad i_3=\begin{vmatrix}a & b & d\\ b & c & e\\ d & e & f\end{vmatrix}$$

那么可依据 I_1，I_2，I_3 得到如下分类：

当 $I_2\neq0$，$I_3\neq0$ 时，如果曲线不是空集（虚曲线），则为有心圆锥曲线：

$I_2>0$，$I_3<0$ 椭圆

I_1 $I_2>0$ 虚椭圆

$I_2<0$ 是双曲线

当 $I_2=0$，$I_3\neq0$ 时，曲线为抛物线

当 $I_3=0$，$I_2\neq0$ 时

$I_2>0$，曲线退化为一点

$I_2<0$，，曲线退化为两条相交直线

当 $I_2=0$，$I_3=0$ 时，曲线是空集或一条直线或两条平行直线。

由此可见，原来通过不同方法截出或作出的圆锥曲线都可以分别纳入二次曲线的范畴当中。反过来，二次曲线无非就是圆锥曲线及其退化情形，但是由于二次代数方程不完全是实曲线，因此，二次曲线中也包含虚曲线，这通常超出解析几何学的范围。

80.把图形画在纸上

——射影几何

文艺复兴时期，大画家如列奥纳多、达·芬奇和丢勒等人对于如何把自然景观和人物正确地表示在画布上进行了细致的研究，其结果形成透视方法。但是，其中一些理论问题有待于数学家解决，而这促使射影几何学的产生。

射影几何学，也称投影几何学，研究图形在射影变换之下的不变性质。更具体讲，一个物体与截景有什么关系，两个不同的截景有什么共同的性质。促使射影几何学研究的是艺术家及建筑师，他们早在古希腊到中世纪欧洲对透视法有许多研究。到文艺复兴时代，已经形成透视法的数学体系，但真正开创射影几何学的是法国建筑师德萨格,他在 1639 年出版了一本书中研究圆锥曲线的投影性质，他在书中引进了无穷远点和无穷远线，其实，这些概念在画家看来很直观，在他们的画布上，两排平行的树木越来越靠近消失在画布上一点。从这点来看，这是与欧几里得相抵触的。另外，在射影之下，长度和角度也有所改变，但是交比却是不变的。在他的影响下，帕斯卡真正用射影不变性把许多关于圆的定理推广到圆锥曲线上。

射影几何的复兴开始于 18 世纪末，主要来自蒙日、卡诺等。蒙日从实用的观点出发，建立了画法几何学。虽然有人用画法几何学来称射影几何学，但画法几何考虑的是平行投影。他的着眼点是工程画，他的画法几何成为建筑师及工程师必学的课程。其后，还发展出其他的投影方法，包括中心投影法，而这些在 20 世纪在航空测量、卫星摄影等领域继续起着重大的作用。

81.共点与共线

——德萨格定理

德萨格是法国工程师，他是射影几何学的先驱，他在 1636 年发现了著名的德萨格定理：

如果两个三角形的对应顶点三条连线相交于一点（共点），则这两个三角形的对应两边的三个交点位于一条直线上（共线），反过来如果两个三角形的对应两边的三个交点共线则这两个三角形的对应顶点的三条连线共点。

如图所示，两三角形△ABC 和△A'B'C'当 AA'，BB'，CC'交于 O 时，这两个三角形成透视对应。显然三角形△ABC 上每一点一方面通过 O 对应另一三角形△A'B'C'上一点，另一方面对应由 O 出发的一条射线，这样一方面我们有一系列点列，另一方面有一系列线束。假如点列每一元素（点）落在线束

的相应元素（线）上时，这点列和线束也称为透视对应。同样如果两个线束的每对对应射线相交于一直线——透视轴时，这两个线束也称为透视的。透视对应是两个图形射影对应的最重要的典型情况。

什么叫两图形之间存在着射影对应或者它们之间射影关连呢？定义是这样的：如果一个图形的每四个元素的交比与另一个图形四个对应的元素的交比相等时，就称两者射影对应。对于一条直线上四点 A，B，C，D，它的交比（ABCD）是

$$\frac{AC}{BC}:\frac{AD}{BD}$$

而对于一个线束里四条射线 a、b、c、d、它的交比（abcd）是

$$\frac{\sin ac}{\sin bc}:\frac{\sin ad}{snbd}$$

其中 ac 表示两条射线 a 与 c 之间交角。

两个透视图形相互射影对应是根据帕普斯定理，而帕普斯是公元四世纪希腊数学家，他收集许多古希腊数学典籍，使它们得以流传至今。

帕普斯定理很简单：一个线束里四条射线的交比与任意直线与这四条射影的四个交点的交比相等。

这样一来，两个相互射影对应的点列或线束总能转换到一个透视位置，由此不难证明德萨格定理。

交比的概念是射影几何学的中心概念，我们定义好圆上四个点的交比，圆锥曲线上四个点的交比，圆锥曲线上四条切线的交比等等之后，我们也能够研究它们的射影几何了。

82.点与线的对应

——对偶定理

射影几何学对几何学乃至数学的一大贡献是对偶性的发现。通过对偶性，一个定理对应与它对应的对偶定理。两个定理虽然很不相同，却是同一定理的两个方面。对偶性的概念不仅可以使数学家发现定理、证明定理，而且在许多当代数学前沿领域如抽象代数和拓扑学中有着重要应用。

对偶定理的最主要例子是帕斯卡定理和它的对偶定理——布里安香定理。帕斯卡定理是 1640 年在一个 6 页的小册子中首先发表的。当时他只有 17 岁，他证明这个定理是受了德萨格在 1639 年的著作的启发而进入射影几何领域的。帕斯卡的定理的对偶定理却是在 170 年后于 1810 年由布里安香首先发表的，这两者之间的对偶关系一直到 19 世纪 20 年代才由几位数学家通过不同途径独立发现的。

帕斯卡定理是指，内接于圆锥曲线的六边形中，三对对边的三个交点共线。

帕斯卡定理是个非常一般的定理，它不仅包括任何圆锥曲线，不管是椭圆、抛物线还是双曲线，而且还包括退化的情形，如图以及两条直线。它的内接六边形也不必按照我们想象的是一个凸的六边形，这六个顶点可以任意排序，1，2，3，4，5，6，一旦排定以后，六条连线 12，23，34，45，56，61 就是六边形的边，这时对边也完全确定，也就是边 12 与 45，边 23 与 56，

边 34 与 61 互为对边，这三对对边的三个交点所在的直线称为帕斯卡线，六边线称为帕斯卡六边形。当然，帕斯卡六边形的六个顶点也可以有的重合在一起，退化成五边形、四边形甚至三角形，也有相应的帕斯卡定理。

帕斯卡定理的逆定理也成立，如果一个六边形（其中没有三个顶点在一直线上）的三个对边相交于一直线上，则它们的六个顶点位于一个圆锥曲线上。

注意！这个逆定理可不是对偶定理。帕斯卡定理的对偶定理是外切于圆锥曲线的六边形中，三条对顶线共点。它同样具有帕斯卡定理的一般性，外切于圆锥曲线的六线形当然也是个六边形，不过我们强调它的六条切线Ⅰ，Ⅱ，Ⅲ，Ⅳ，Ⅴ，Ⅵ为它的边而依次两切线的交点 Ⅰ Ⅱ，ⅡⅢ，ⅢⅣ，ⅣⅤ，ⅤⅥ，ⅥⅠ称为其顶点，而顶点ⅠⅡ与ⅣⅤ，顶点ⅡⅢ与ⅤⅥ，顶点ⅢⅣ与ⅥⅠ称为对顶点，连接一对对顶点的线称为对顶线，实际上就是通常所说的主对角线。这个布里安香定理的逆定理也成立，如果一个六边形（其中没有三边过一点）的对顶线过一点，则六边形的边是一个圆锥曲线的切线。

现在我们可以看出了它们之间的对偶关系

点←——→线

点的连线←——→线的交点

对边的交点←——→对顶点的连线

交点共线←——→连线共点

因此，对偶关系的实质是点与线的对应，共线与共点的对应。因此，线与点一样具有作为几何元素的资格。

83.唯我独尊

——欧几里得几何

虽然自 19 世纪中期起，已经发生大大突破欧几里得几何学的框架的革命，但是，我们还不能说我们已经完全摆脱欧几里得几何学观念的束缚。因此，有必要检查一下欧几里得的观念，以及对它的批判产生的结果。

（1）欧氏空间是现实的空间形式，几何学同物理学一样是以现实物理世界为对象的自然科学。时至今日，许多人把数学，特别是几何列入自然科学，其主要根据也在这里。其实，这种观点是极为错误的，问题在于

①数学并不研究物理空间。实际上，空间的概念最早是开普勒在 17 世纪初引进数学之中的，用它代替主体，这在后来也是沿用为了区别开二维与三维的几何图形。数学实际上只是研究几何图形，虽然常常假定它是安装在一个大空间中，而这也只是在解析几何中取坐标的数目时才默认的，许多具体问题是与空间不相干的。

②认为我们所在的现实空间是欧氏空间是大哲学家康德的观念，实际上康德硬性规定空间只此一种，因此几何自然也就只有欧氏几何了。

③欧几里得几何学实际上是经验的概括，只是这个公理系统得不出来相互矛盾的结果。因此，这里有两种真理观：一种是欧氏几何的结论符合事实，而这只能说是欧氏几何可能是我们这个世界的模型，但它并不能排除在逻辑上不自相矛盾的几何学体系。而另一种可能性则是逻辑上不自相矛盾的体系。例如非欧几何学体系。因此在确定我们现实世界是什么样时，首先要解

决两个问题：

A、有多少种可能的几何学体系或者空间模型。

B、哪一种符合现实。

前一问题是数学问题，后一问题是实测问题，从而康德独断地认为欧几里得空间是没有任何根据的，而且从这时起，数学也不是自然科学的一个分支了。

（2）欧几里得的一些具体做法也不断受到挑战，每一次斗争都导致数学的新发展。

①平行公理的否定导致非欧几何的建立，打破欧氏几何的一统天下。

②去除三维的限制，打破数学必须局限于“现实空间”。

③欧几里得公理体系不完善，必须严密建立新公理学。

④局限于整体大于部分的公理的突破，使得无穷数学得到发展。

总之，近 200 年数学发展的历史正是欧几里得唯我独尊一统天下的逐步破灭过程。

84.对欧氏几何的挑战

——非欧几何

欧几里得的《几何原本》第一卷中提出 23 个定义、5 个公理、5 个公设。麻烦都出在第五公设，也就是平行公设上。《原本》的平行公设是这么叙述的：平面上两直线被一直线所截，如果截线一侧的两内角之和小于两直角，则此二直线必相交于截线的这一侧。欧几里得的第 23 个定义是平行线，它们必须在同一平面上，并且两边无限延长后均不相交。这个公设不像其它公理公设在直观上明显，而且直到命题 29 才需用第五公设来证明。因此，从古到今激起许多人企图利用其他公理和公设来证明这个平行公设。

长期以来，这种“证明”并未取得成功，但是，却大大简化了第五公设，使得它变成直观上明显的等价命题。其中最常用的是苏格兰物理学家、数学家普雷范尔的命题：过不在给定直线上的一点，可做一条且仅可做一条直线与该给定直线平行。他在 1795 年证明这命题与平行公设等价之后，成为平行公设的通用提法。实际上，公元五世纪普洛克洛斯已经提到过。沿着这条路走下去得出一些与平行公设等价的命题：

（1）任一三角形的内角之和等于两直角，后来这命题可简化为（1），至少存在一个三角形，其三个角之和等于两直角。

（2）存在一对相似三角形但不全等。

（3）过任意不共线的三点可以做一圆。

（4）在小于 60° 的角内一点，总可以做一直线与该角两边相交。

（5）两条相交直线不能同时平行于第三条直线。

这些命题也都同平行公设一样不能单独从其他几条公理中推出，如果推出，必定都隐含运用一条等价的公理，这样一种办法行不通。于是许多人采取第二种办法：即用平行公设相矛盾的命题代替平行公设，即①罗氏公设，过一直线外一点可作两条直线与该直线平行；②黎氏公设，过一直线外一点不能作直线与该直线平行，并试图推出矛盾但没有成功，而导致非欧几何学的成立。

85.平行线不止一条

——罗巴切夫斯基几何

现在公认非欧几何学的奠基人是高斯、罗巴切夫斯基和波耶，他们是彼此独立地得到所谓罗巴切夫斯基几何（克莱因称为双曲几何）的。高斯早在1792 年就开始深入研究这个问题，但他并没有公开发表过任何著作。1799 年他声称他已掌握一种违背平行公理的新几何学。1813 年他发展他的新几何学，他称之为反欧几何，后称星空几何及非欧几何。他在后来的书信中一再明确指出，平行公理不能由其他公理证明，而且认为这种新的几何确有实际应用。他曾实际测量三个山峰构成三角形的内角和是否 180°，由于实际误差没能得到确切结论。他认以到只有更大的三角形才会显示明显差别。罗巴切夫斯基 1826 年宣读非欧几何第一篇论文而且 1837 年用法文和 1840 年用德文发表在西欧的主要杂志上，但直到去世也没有得到理解和重视。波耶早在 1823 年就告诉他父亲法卡斯·波耶，说他已作出一个新的平行理论，并在他父亲的书《为好学青年写的数学原理》（共二卷，1832—1833）作为附录发表。他的绝对几何学同高斯及罗巴切夫斯基几何也是等价的。双曲几何在否定平行公设的假设前提下引进一些新内容：

（1）平面上两条直线上存在三种关系，除了两条直线相交及平行之外（这时它们都不能有公垂线），还可能有第三种关系称为超平行，它们没有交点，但有唯一一条公垂线。

（2）平行角π（a），设 AB 为一直线，C 为直线外一点，C 到 AB 的垂直距离为 a，过 C 的直线可分为两类，一类与 AB 相交，一类不与 AB 相交，则不与 AB 相交的直线与 CD 成角度最小者定义为 a 的平行角，记作π（a）。罗巴切夫斯基和高斯独立得出

$$\operatorname{tg}\frac{\pi(x)}{2}=e^{-\frac{x}{k}}$$

k 称为空间带数。

（3）三角公式，罗巴切夫斯基认识到，他的几何的三角公式就是球面三角公式中，边长及角度均用虚数即可得到。

（4）面积公式，三角形 ABC 的面积与π－（A＋B＋C）成正比，这个数值称为角亏。

（5）从境界线（极限圆）的导入，罗巴切夫斯基在《平行线论》中导入境界线和境界球面的概念，境界线是平面上曲线，其所有弦的垂直等分线（称为轴）都互相平行。而境界球面是以境界线绕其中任意一轴所旋转出的球面。境界球面的几何学是欧氏几何学，从而可以看出欧氏几何学是罗氏几何学的极限情形。

86.群与几何学

——埃尔兰根计划

1872 年克莱因发表埃尔兰根计划，目的是用群的观点来统一各种几何

学。他的基本观点是：每种几何学都由变换群所刻划，每种几何学所要研究的就是几何图形在其变换群下的不变量。而一门几何学的子几何学就是研究原来变换群的子群下的不变量。射影变换群的一个子群是仿射变换群，仿射变换群保持一条直线 l_∞ 不变，因此仿射几何学是射影几何学的子几何学，射影变换下的不变量除了射影几何学的不变量之外，还有把直线变成直线、平行直线变成平行直线等性质。仿射几何学虽然早已出现在欧拉及莫比乌斯的著作中，但克莱因在他的计划中并没有提到。克莱因进一步考虑了主变换群，实际上是相似变换群，相应的几何是相似几何学，相似变换群的群是运动群，相应几何学是欧几里得几何学。

类似地，他进一步刻划双曲度量几何，也就是研究射影平面上使一个任意的、实的、非退化的二次曲线保持不变的所有变换所构成的子群下的不变量，这个子群叫作双曲度量群，相应的几何学叫乍双曲几何学，其中的不变量是合同有关的那些量。同样，单纯椭圆几何学所研究的变换是使射影平面上一个虚椭圆不变。而二重椭圆几何学则使三维空间中的球面保持不变的三维射影变换群的子群。这样他统一了射影几何学及各种度量几何学。

克莱因的计划不仅在于统一点几何学，而且从变换群观点出发，把其他类型的几何学与任意空间元素的几何都统一在一起，并且指出与它们相对应的代数理论：例如二次型理论同具有一条基本圆锥曲线的平面射影几何学等价，空间射影几何学与四元二次型理论等价，而线几何学与五维射影空间中二次流形的射影几何等价，平面反演几何学与四维射影空间中二次流形的射影几何等价，而且复二次型理论可以通过实球面射影几何学来表示。因此，二次型理论、反演几何学、直线几何学是互相对应，它们的区别只是变无数目不同，而这自然把几何引导到高维中去。这样几何的对象再也不是空间中的曲线、曲面和立体了。

克莱因进一步推广了这种观点，他提出更一般的问题，给了一个流形和这个流形的一个变换群，以在这个变换群的变换之下其性质保待不变的观点研究这个流形的实体。在这广义的意义下，克莱因考虑的不仅仅是通常以点为基础的几何学，而且考虑以任何一种点集为基础的几何学。例如圆几何及球几何也可以看成研究某些射影变换群的某些子群的不变性质，而且还更进一步扩大他的计划的应用范围：代数几何学研究双有理变换下的不变性，拓扑学研究连续变换下的不变性等。虽然并非所有几何学都可以纳入克莱因的分类框架，但是这种观点至今对几何学仍有影响，特别是强调变换下的不变性，对于力学及物理学思想的推动，大大超出了数学的范围之外。

87.笔直与弯曲

——曲率

弯曲性是最直观的几何性质之一。当我们走到街上时，我们不难判断道路的弯曲程度。北京的许多街道一眼望不到头，它们可以横平竖直，而山城重庆、青岛的许多道路弯弯曲曲，还要上上下下，开起车来总要左拐右拐，改变方向。数学家并不满足于直观的感觉，他希望对于曲线和曲面的弯曲程度有一个严密的定义和精确的量度，这个题材构成了微分几何学的主要内容。

数学是一门抽象的科学，研究任何对象必须抓住弯曲的本质。为此，我们先考虑最简单的平面曲线。平面曲线中直线是直的，圆是典型的曲线，它们的不同之处在于沿着直线向一个方向运动，方向永远不改变。因此，要想度量弯曲的程度，只要看一下方向随着曲线改变的程度。对于圆来说，当绕圆走上一段圆弧长度为 S 时，方向改变可由两点切线之间夹角φ来确定，我们可以定义圆的曲率为：

$$\frac{\varphi}{S}=\frac{1}{R}$$

这里φ用弧度来表示，R 正好是圆的半径，显然它是弯曲程度的度量，因为它符合我们的直观。一个圆越小弯曲程度越大，也就是曲率越大，一个圆越大曲率也就越小也就越发平直。当一个圆半径无穷大，曲率也就变成 0，它也变成一条直线。直线是曲率为 0 的曲线，它当然最直。

一般的平面曲线与圆不一样，它每一点的弯曲程度都不一样，因此，对于一般曲线，必须时每点区别对待，这也同道路一样，有的地方弯曲，而有的地方比较直。这说明曲率并不是曲线的整体性质，而是局部的性质，也就是在一点局部的性质。求曲线上每点的曲率也可以用微积分，例如我们对于曲线

y＝f（x）

它的曲率为

$$k=\frac{|y''|}{(1+y'^2)^{3/2}}$$

y'及 y"为 y 对 x 的一阶及二阶导数，不过不懂微积分也不要紧。我们在曲线每一点画一个密切圆，也就是这个圆不仅同曲线在这点相切，而且也和曲线在这点的切线相切，换句话说，这点的密切圆和曲线在这点有公共切线。这时显然曲线在这点的曲率就是密切圆的曲率也就是密切圆的半径的倒数。

曲率还有力学的意义。根据牛顿的惯性定律，质点在不受外力的情况下，应沿着这点切线方向前进。因此，如果质点沿曲线匀速（速度 v 大小不变）运动的话，曲率则表示曲线偏离切线的速度。可以证明，在该点的（法向）加速度 w 为

$w=kv^2$

对于空间曲线来说不仅有曲率还有挠率。挠率反映曲线偏离平面曲线的程度。而对于曲面，就有更为复杂和深刻的曲率理论。

88.地球上两点间最短航线

——测地线

当你走在旷野中，如果没有特殊的道路，自然会选择连结两点的最短路线，因为在一个平面上两点之间直线最短。可是，如果在一望无际的茫茫大海中航行，或者在 10000 米的云端飞行，什么是最短的路线呢？从数学上看，也就是求一个曲面上两点之间的短程线或测地线，它相当于曲面上的“直”线。

平面上的直线有三个突出的性质：

（1）方向不变；

（2）不弯曲；

（3）是两点间的最短距离。

但是，曲面上的测地线显然不能完全满足这个要求，但是我们尽量接近这些要求：

（1）曲面上每一点有一个切平面，它在这点的垂线称为这点的法方向。曲面上的曲线每一点有切线，与切线垂直的有一个法平面，法平面上有两个互相垂直的方向，称为主法线和从法线。对于测地线，它每点的主法线与曲面的法方向平行，也就是测地线相对于曲面保持方向不变的性质。

（2）弯曲程度是用曲率来衡量的，直线的曲率等于 0。曲面上的曲线一般很复杂，曲率甚至挠率都一般不等于 0，因此，我们期望曲率尽可能小。如果是球面，连接其上两点的平面曲线有大圆（平面通过球心）和小圆（如两点同一纬度的纬线圈），显然，大圆的曲率小，小圆的曲率大。因此，球面上的测地线是大圆弧。因此，从美国到东北亚的飞机航线不是通过夏威夷，而是要通过北极圈，因为大圆才是最近的航线。对于一般的曲面上的曲线，我们定义一个测地曲率，与直线的情形相类似，测地线的测地曲率处处为 0。

（3）在一个球面或更复杂的曲面上，两点之间可以用任意长的曲线连结，因此不存在最长线，但是测地线也不一定就是最短线。以球面为例，连结两点的大圆可以分成短弧和长弧，只有短弧是最短线，而长弧部分就不是。对于一般曲面，通过微分几何学只能证明，两个无限接近的两点之间存在唯一一条测地线。但在一般情况下，这问题就比较复杂了，这时就涉及大范围的问题。历史上一个著名问题是一个凸闭曲面上至少存在三条没有自交点封闭测地线，这个定理在 30 年代已被证明，现在知道一般的紧闭曲面可以存在无穷多条闭测地线，这正如球面上存在无穷多个大圆一样。

89.张在铁丝架上的肥皂膜

——极小曲面

1849 年起，比利时物理学家普拉托进行一系列实验，用金属丝弯成各种形状的框架，浸入肥皂水中，拉起来后看框架是否张上肥皂膜。这样的肥皂膜由于表面张力之故是所有张在同样框架上的曲面中面积最小的，用数学的话讲，以某个封闭曲线为边界的所有曲面中求面积最小的曲面。这个面积最小曲面称为极小曲面。这个问题也称为普拉托问题。事实上，早在普拉托实验之前一百多年，数学家已经研究极小曲面问题。

1760 年还在知道具体的极小曲面之前，已经由拉格朗日作为变分问题提出来，他证明，如

$z=Z(X, y)$

是极小曲面方程，则 z 满足一个非线性偏微分方程。其后蒙日发现，这方程解曲面的面积极小的条件为平均曲率 $H=0$，因此以后称 $H=0$ 的曲面为极小曲面。但 $H=0$ 只是一个必要条件。

利用变分方法得到非线性偏微分方程，然后在给定边界条件下求解是非常困难的数学分析问题，经过近 200 年的努力，我们对于解的形式以及解的存在性是否唯一等都有了相当丰富的结果。由于运用复变函数的理论，已经知道极小曲面的一般形式，并可以用参数表示，由此，得出一系列极小曲面。

18 世纪只知道三种极小曲面，除了平面之外，还有欧拉在 1744 年发现的用悬链线的一段绕 X 轴旋转产生的悬链面是极小曲面。后来又有人发现正螺旋面是极小曲面。1830 年起许多数学家又发现许多新型的极小曲面，特别是德国数学家舍尔克发现的具有单周期和双周期性的极小曲面，后者定义在一个国际象棋棋盘上，每个白方块上的曲面和每个黑方块上的曲面都是一样的。

19 世纪中叶起，又发现许多特殊的极小曲面，一种是恩涅配尔极小曲面，它具有自相交的曲线。另一种是边界不是一条或几条曲线而是一个立方体的框架为边，也存在极小曲面。但是，是否有亏格≥1 的曲面也是极小呢？如果是封闭曲面这是不存在的，但是在曲面有扎了的空洞还是可能的。这些一直到十几年前才能得到，这一次很大程度上借助于计算机的帮助，1982 年首先得到亏格为 1 的完备极小曲面，它有三个末端（空洞）。它的图像要是没有计算机是根本做不出来的，后来陆续发现，对每个亏格≥1，都存在完备的极小曲面，末端数≥3。因此，近 20 年来，极小曲面成为数学的一大热门，无论是数学分析，还是高维几何学乃至物理学都从这一发展得到好处。

90.走向新空间

——各种几何学

19 世纪中叶以前，欧几里得空间是头一个教学空间，也是唯一的数学空间。欧几里得空间是通过欧几里得几何学的公理来刻划的，它反映人们对空间的平直性（线性），匀齐性，各向同性，包容性（作为万物的容器），位置关系（距离），无穷延伸性，三维性，连续性，无限可分性等等性质的初步认识。

欧几里得三维空间可以从许多方向上进行推广：

（1）维数上的推广：由 3 维推广到 4 维、5 维……甚至到无穷维。

（2）变换群的推广：我们抛弃度量性质只研究射影变换下的不变性质就得到射影几何学，同样可以得到仿射几何学及保形几何学。相应的空间自然也就是射影空间、仿射空间等等了。

（3）空间元素的推广：以前空间元素我们默认是点。19 世纪中普吕克尔就创立以直线为元素的空间的几何为线几何学，直线或曲线还可以构成线汇及线丛或者网、罗等等。同样也有圆几何学和球几何学。19 世纪后半叶这方面的几何学研究得十分热闹，所有这些空间也可以从变换群角度来研究。

（4）度量上的推广：非欧几何的出现是欧几里得空间推广的第一个方向，非欧空间——双曲空间与椭圆空间打破了欧氏空间独一无二的绝对地位。由于非欧平面可以在欧氏空间的曲面上实现，就把空间的平直性去掉了。我们可以概括成为“常曲率空间”，把曲率处处相等这个条件再推广，可以成为黎曼空间，研究它的是黎曼几何学。

（5）基域的推广：作为欧几里得空间的基域——实数域可以推广到一般域上，首先是复数域，其次是有限域。有限域上射影空间、仿射空间的几何学是 20 世纪重要的研究课题。

（6）由局部到整体的推广：局部欧氏空间可以拼接成各种流形。

（7）在数学中常用的空间，随着泛函分析的发展，许多以函数或算子为

背景的抽象空间应运而生，成为分析研究的新对象（几何化的分析），它们都构成二级大学科。

（8）最一般的拓扑空间，当空间摆脱了度量及范数之后，就变得比较抽象。空间中点与点的关系可通过邻域或开集、闭包来定义。最常见的是豪斯道夫空间。常见的空间都是豪斯道夫空间，但代数几何学中还有一种非豪斯道夫的拓扑，即查瑞斯基拓扑。对于拓扑空间还有种种推广，如一致性空间、接近空间以及上百种推广，从中已经看不到一点通常空间的直观了，它们是一般拓扑学的研究对象。

如果说数的推广我们还能理解，空间这种无穷无尽的推广已经使我们一点具体形象也感觉不到了。几何学由具体直观的学问已变成十分抽象难以捉摸的东西了。

拓扑学

91.哥尼斯堡的七桥

——笔画问题

第一个拓扑问题是欧拉在 1736 年解决的哥尼斯堡的七桥问题。

哥尼斯堡是东普鲁士的首府，这座历史名城产生过大哲学家康德和大数学家希尔伯特。普雷格尔河横贯哥尼斯堡城中，其中有七座桥。

问题是，我们能否散步经过每一座桥，而且只经过一次，欧拉证明这不可能，他的数学才能就表现在他把这问题化成图的形式，然后证明不存在一笔画法把这图画出。

如果经过这七座桥所有可能路线都试一下的话共有 5040 种路线,如果每一条路线都试一下证明行不通的话，显然既费时日，也太盲目了。欧拉作为一个数学家，在这里显示出他的威力。欧拉没有到过哥尼斯堡，更不去盲目地乱试，他只是把问题简化，去掉不必要的因素，例如桥的长度，然后把用河隔开的四块区域缩成四个点，这样七桥就变成 4 个点间、7 条连线组成的图，而七桥问题就变成这个图能否一笔画成。那么一个图能否一笔画成，依赖于点和线的数目。连到一点的数目如是奇数条，就称为奇点，如果是偶数条就称为偶点，要想一笔画成，必须中间点均是偶点，也就是有来路必有另一条去路，奇点只可能在两端，因此任何图一笔画成，奇点的数目不是 0 个就是 2 个，否则不能一笔画成。现在看欧拉的七桥简图，4 个点均为奇点，因此，七桥问题找不到一个肯定的路线。

92.多面体的秘密

——欧拉公式

研究图形拓扑性质的方法，可以说是受到欧拉公式的启发，欧拉公式是多面体的顶点数，棱数和面数之间的一个关系，尽管多面体可以千变万化，这个关系却一定满足，也就是这个关系反映了多面体的本质的拓扑性质。

无论在中国还是外国，正多面体都是最早的研究对象，对于任何正多面

体，若用 V 表示顶点数，E 表示棱数，F 表示面数，那么我们数一下，就可以得到下面的表：

	V	E	F
正四面体	4	6	4
正六面体	8	12	6
正八面体	6	12	8
正十二面体	20	30	12
正二十面体	12	30	20

尽管这些数目各不相同，却可以发现，V－E＋F 都等于 2，这是不是巧合呢？进一步研究更复杂的多面体，发现这公式也对，笛卡儿是知道这个公式的。1750 年，欧拉首先宣布他发现这个公式，随后又给出了一个证明，尽管他的证明不对，这个公式后来还是被称为欧拉公式。对于某些特殊情形，有些数学家证明了这个公式，但在 1813 年，瑞士数学家路易耶发现欧拉公式并非对任何多面体都成立。比如，一个正方体中挖去一个小的正方体，数一下就知道 V－E＋F＝4，要是把小立方体上下都挖通，则 V－E＋F＝0。

那么欧拉公式在什么条件下才成立呢？经过这许多数学家研究，发现只要多面体是实心的（里面没有空洞），只有一个外表面，并且这个表面可以变形成为球面，那么欧拉公式就成立，一个多面体，只有一个外表面，而且外表面可以变形为环面（和内胎相似），那不管多面体如何变形，都有 V－E＋F＝0。

总而言之，拓扑学研究图形的某种性质，这种性质在连续变形下不变，这种性质我们可以称为拓扑性质，而在连续变形下不变的量称为拓扑不变量。比如上面讲的 V－E＋F 就是一种拓扑不变量。

93.把平面分为两部分

——若尔当定理

几何学最简单的概念是曲线。但是在拓扑学里，曲线并不太好定义。法国著名数学家若尔当在 1882 年用参数方程加以定义：曲线是平面上的点集，其坐标是参数 t 的连续函数

$x=\phi(t)$

$y=\phi(t)$

t 在区间[a，b]取值。所以若尔当定义下的曲线无非是区间的连续。曲线两端连在一起就是闭曲线。如果闭曲线没有自交点，称为简单闭曲线。简单闭曲线实际上是圆的连续象。如果把这条闭曲线画在橡皮膜上，当橡皮膜连续变形时，曲线的长度及包围的面积可以改变，但是根据拓扑的定义，它的拓扑性质不会改变，其中一个拓扑性质从直观上极为显然：平面上一条简单闭曲线 C 恰好把平面分成两个区域，一个是内部，一个是外部，它们均以 C 为边界。这样，平面上的点分成两类：C 内部的点和 C 外部的点。同一类中两任意点都能用一条不与 C 相交的曲线相边接，而连接分属两类的点的曲线必定与 C 相交。而且，这两个区域一个是有界的，一个是无界的。对于有界区域中任何一点 x，以及 C 上任何一点 x_0，总可以连接一条简单弧以 x 及 x_0 为端点。且除了 x_0 之外，整个弧都在有界区域之内。这称为舍恩夫列斯定理。

由于这个定理直观上很显然，一直到若尔当才认识到有必要正式表述这个定理并加以证明。若尔当的证明非常长而且复杂，后来还发现其中有问题，最主要的许多直观的概念，特别是曲线和内部外部这些概念需要更严密的分析，拓扑学就是考虑这些问题的。

由于问题的复杂性,若尔当曲线定理一直到 1905 年才由美国数学家维布仑首先严格证明。这个定理很快推广到高维：n 维欧几里得空间 R^n 中（n－1）维子流形如果同胚于（n－1）维球面 S^{n-1}，也把 R^n 分为两部分，且该子流形是两部分的公共边界。法国数学家勒贝格证明 n－3 的情形，荷兰数学家布劳威尔证明一般的情形。

库拉托夫斯基证明了强化的舍恩夫利斯定理：存在一个由闭单位圆 D 的同胚，它把 D 的边界 S'映到 C 上，把 S'内部映到 C 的内部上。但这定理在高维情形是错的。作为反例，亚历山大举出他那著名的“ 角球 ”。

94.填满空间的曲线

——维数

众所周知，我们所在的世界是所谓“ 三维空间 ”，意思是你要在空间中确定某一点位置，只需三个数就够了。反过来，任意三个实数都对应空间中唯一的一点。因此，空间每个点和三个实数之间存在“ 一一对应 ”的关系。同样，平面是一个“ 二维空间 ”，直线就是一维空间，孤立的点自然就是零维的。

因为在三维空间中，一条曲线上的点和一条直线的点可以一一对应起来，所以我们可以说，曲线也是一维的。同样，曲面是二维的。三维空间中一块实心的东西，有长、有宽、有厚，我们也可以说它是三维的。

但是，用参数的数目来决定维数只对于欧几里得空间才对，对于一般流形只是局部才对。而一般维数的概念很长时期中并不清楚。比如说，一段曲线是一维流形，可是皮亚诺在 1890 年发现一段曲线可以盖满一个正方形。他的想法是把一维线段 AB 和正方形 PQRS 对应起来，怎样对应呢？把 AB 分成 16 段，把 PQRS 分成 16 块，这样每一段就对应每一块中的黑线段，而且保持首尾相接。再把这小线段分成 16 段，把小正方形再分成 16 块，同样可以对应起来，如此分下去，线段越分越短，正方形越分越小，它们之间仍能保持一一对应关系。当线段分得很细很细，正方形也很小很小时，线段也自然通过正方形每一个点，从而覆盖整个正方形了。这样一来，一维的线段和二维的正方形没有什么区别了。古老的维数概念发生动摇，以至 19 世纪末大数学家克莱因曾感叹，那时最头疼的问题是“ 什么是曲线？ ”

20 世纪初，法国大数学家庞加莱发现维数是空间的“ 拓扑 ”性质，他通过三维空间的一些现象总结出来维数的本质：一个点可以把一个线段分成不相连结的两段，也就是零维空间可以把一维空间分成不相连结的两部分，但是一个点不能把一块曲面，比如说皮亚诺那个正方形，分成不相连结的两部分，由这样的性质可以判断一个线段和一块曲面的维数不一样。同样，一个一维图形——圆圈可以把二维的曲面，比如说球面和环面，分成不相连结的里外两部分，但是一维图形却不能把一块三维的实心球分成不相连结的两部分。二维图形却可以把三维空间分成两部分。这样我们就可以一步一步归纳

地定义维数了。这是维数的严格的拓扑定义，接着布劳威尔证明了维数的拓扑不良性。20 年代，门格和乌雷松独立地给出维数的严格理论，正式建立起维数论这样一门拓扑学的分支，这时空间的观念也大大超出我们经验的空间和物理的空间之外，而出现应用范围极广的抽象的“拓扑空间”了。

95.单侧的带子

——莫比乌斯带

莫比乌斯带是最简单的单侧的有边缘曲面。双侧面面与单侧曲面有明显的不同。设想曲面是由薄膜做成的。如果从薄膜的一侧出发存在一条道路到另一侧，中途不允许越过边缘，也不许穿过薄膜，那么这种曲面称为单侧曲面，否则称为双侧曲面。另一个特征是双侧曲面，我们总可以分成内外两面，各涂上一种颜色，例如红色和蓝色，它们不会相混，而单侧曲面只能涂上一种颜色。

莫比乌斯带在日常生活中也可见到。当把皮带扭个弯再插到孔中，就得到一个莫比乌斯带，这时你会发现皮带的反面翻到上面来。更具体的做法是找一条细长的长方形纸条 ABCD 如果把两边 AB 和 CD 粘起来，使 A 与 C 重合，B 与 D 重合，则得一圆筒形带，正确的皮带也是这样的，它有一个正面、一个反面，可以涂上两种不同的颜色。可是要在两端粘在一起之前，扭转 180°，使得 A 与 D 重合，B 与 C 重合，就得到莫比乌斯带，显然这种曲面是单侧的。小虫沿着 PP'直线爬行时，它不越过边缘就从一侧爬到另外一侧。

莫比乌斯带和圆筒形带还有两个明显的不同之处：

（1）圆筒带沿 PP'剪开后成为两个圆筒带，而莫比乌斯带沿 PP'剪开后还是连在一起成为更大的莫比乌斯带。

（2）圆筒形带的边缘是两条闭曲线（圆周），而莫比乌斯带的边缘只是一条闭曲线。

单侧曲线为我们引进一种最重要的拓扑性质——可定向性。一个曲面的可定向性，是指整个曲面都存在一种相互谐合一致的方向规定，这种规定不能由点在曲面上的移动而破坏。在莫比乌斯带上我们在 PP'可以规定一个 x 轴，方向沿 PP'的走向，而 y 输方向，是垂直于 x 轴且从 x 轴到 y 轴是逆时针方向。这样在莫比乌斯带上，当沿着 PP'线运动时，每点的方向都是谐调一致的，也就是两点周围一个小团曲线如是逆时针的，那么在它们相交之处的点周围的小闭曲线都是逆时针的。但是，再次走到 P'＝P 点时，这个定向反了过来，小闭曲线由逆时针变成顺时针，或者 y 轴由向上变成向下。这种现象在圆筒带或球面上是不存在的。这些曲面称为可定向的。因此，不可定向性与单侧性是同一拓扑性质。

96.直线构成的空间

——射影平面

射影几何教导我们的一个最重要的事情是点与线的对偶性。过去我们只把点当成构成几何图形的元素，空间或曲线、曲面都是由点构成的。现在直

线甚至圆、球面等等都可以和点一样构成空间的元素。它们所构成的空间是怎样的呢？为了简单起见，我们只考虑直线构成的空间。

平面上通过原点的所有直线构成了一个一维空间，称为射影直线。它到底是什么样呢？如果我们用一个点来代表一条直线的话，它就是一个圆周，因为每一条直线都与一个圆周交于一条直径的两个端点，因此这两个点就代表这条直线，把这两点粘合在一起看成一个点，也就得到一个半圆把两端粘起来，它仍然是一个圆周，因此圆周就是一维射影直线的图像，只不过这时以点为元素了。

射影平面是通过三维空间中所有通过原点的直线构成的，我们如法炮制，它相当于把单位球面的直径对顶点粘在一起构成的曲面，这个曲面可不再是球面了，而是射影平面。

射影平面可以通过下半球面加上赤道直径对顶点粘合而成。如图所示，我们把赤道用 ABCD 四边形表示，目的要把 AB 与 CD 粘合，DA 和 BC 粘合。为此，我们把 A 和 C 拉高，B 和 D 拉下，最后把 A 和 C，B 和 D 粘在一起，而且 AB 和 CD，DA 和 BC 相重，这样就得到自交于一条直线的闭曲面，它就是射影平面。看起来它像球面，但它不是球面。如果我们切掉下半的一部分，我们得到的是一个交叉帽，它是又一个莫比乌斯带的模型。因此射影平面也是单侧曲面，即不可定向曲面。

交叉帽有一条自交线和两个奇点。希尔伯特的学生鲍伊造成一个射影平面没有奇点，但自交线还保留，这实际上是不能去掉的。要想造出又没有奇点，又没有自交线的射影曲面只有到四维空间才行。而且我们可以给出四维空间（x，y，y，t）中射影平面的代数方程，即

$$y^2z^2+y^2t^2+z^2t^2=yzt$$

$$y(z^2-t^2)=xzt$$

97.装不进水的瓶子

——克莱因瓶

1874 年德国大数学家克莱因设计了一个瓶子，它装不进水，当然也就卖不出去，不过这个克莱因瓶却使人们，特别是数学家大开眼界。

要使瓶子能够装水，必须有两个条件：第一，它盖上盖子后必须是封闭的，这是不言而喻的。一般人觉得这个条件就够了，在现实生活中也的确如此。可是克莱因说，单这个条件还不够，这个封闭曲面还必须是双侧的，也就是这个曲面必须有里外之分，换句话说，水在里面不能跑到外面来。可是哪里有里外不分的瓶子呢？在莫比乌斯带的启发下，克莱因认为有这种可能性。

克莱因的办法是从一个两头开口的管子出发，如果把管子两头弯曲，一直弯曲到两头相遇，然后把它们粘接在一起，这样就得到了一个环面。现在改变一个办法，先将管子一端弄得一端稍粗一些，一端稍细一些，这从拓扑上讲是容许的，它还是圆柱形。然后把两头弯曲，把细的一头插到粗的一头的管壁之内，然后两头从里面相接近，在接口处可以看到两个同心圆，大的一头套着小的一头，最后把小口放大，大口缩小，两口重合粘接在一起，这

样得到的曲面如同环面一样也是封闭的。但是，它是单侧的，“外面”的一个小虫沿曲面爬，最后也可以爬到里面。

美中不足的是，克莱因瓶在三维空间中一定存在着自交点，只有到了 4 维空间，它才有一个正则的嵌入。

正如射影平面可以通过环面对顶点相粘合而成。克莱因瓶也可以通过球面对顶点相粘合而成。

我们可以把克莱因瓶沿着一条闭曲线剪开，这样我们就得到两条莫比乌斯带。因此，我们也可以把克莱因瓶看成二条莫比乌斯带沿公共边界相粘接而成，它的单侧性就更容易看出来了。

98.球不分大小

——球面拓扑学

在空间图形中，球面是最简单也是最有趣的曲面，它的几何性质非常突出，也非常之多。但是随着球面的连续变形，许多性质消失不见，而留下来的是拓扑性质。

（1）度量性质，这是特殊球面的具体性质，例如直径、表面积及容积、曲率也是具有刚性的度量性质。

（2）微分几何性质，球面具有常数高斯曲率，也具有常数平均曲率。

（3）极值性质，一定体积的立体，以球的表面积为最小，一定表面积的立体，以球的体积为最大。

（4）对称性质，球面的刚体运动群是 3 参数群，属于这类的曲面只有球面和平面。

（5）凸曲面性质，球面是凸的，也就是球面是凸体的表面。球面有许多性质可以推广到凸闭曲面，特别是球面的测地线均为闭曲线，而球面上有无穷多条闭测地线。不仅球面，而且所有闭凸曲面是刚性的。

（6）常宽及常廓，常宽指球面在任意两平行切平面间的距离是常数，这样一个球就可以在两个平行切面间自由滚动。一般人也许认为，球面由这个性质唯一决定，但实际上存在许多凸闭曲面，其中有的没有奇点，也是常宽的，而且常宽立体所有轮廓都是同一种常宽的曲线。即具有常廓性。对于凸面性，常宽性和常廓性等价。球面还具有等照性，即立体的所有投影并不一定周长相等，而是面积相等。除球之外，还有其他等照曲面。

（7）交截性质，球面的特征是所有平面的交口是圆，对于二阶曲面其交口是二次曲线。

（8）映射性质，保形映射的类与曲面的拓扑性质有关。

（9）组合性质，不论如何进行三角剖分，顶点数一棱数十面数＝2。另外，画地图四色足够。

（10）拓扑性质，如双侧性，可定向性、圆的可缩性，任何圆周把圆分为两部分等等。

以上 8、9、10 三类性质对于球面的连续变形均成立，它们是最一般球面的拓扑。

99．轮胎上的几何

——环面拓扑学

从直观上一眼就能看出，球面与环面有很大的差别，最明显的就是环面有一个空洞而球面没有。因此，如果不准撕坏、粘合，由一个球胎是不能变形成轮胎的，反过来，轮胎也不能变成球胎。这反映它们拓扑性质有本质的差别，曲面上的这个空洞数是曲面的最重要的拓扑不变量，称为亏格，用 g 表示。球面的亏格为 0，轮胎面的亏格为 1，显然还可以造出 g＝2，g＝3 等封闭曲面，它们具有 g 个空洞，曲面的拓扑性质都与 g 有关。

我们来看一下球面与环面的差别。在球面上画一个圆圈，你会发现，它可以在球面连续变形，最后缩成一点，这个性质称为单连通。但是在环面上，不是所有的圆圈都能这样连续的变形，例如，当这个圆是绕环面的圆时就不能缩为一点，因此，环面不是单连通的。同样，有 2 个、3 个乃至多个空洞的环面也不是单连通的。为了描述单连通性，庞加莱引进一个拓扑不变量——基本群，用π_1表示。当$\pi_1=0$时，曲面就是单连通的，而$\pi_1\neq 0$则不是单连通，π_1越复杂，表示它与单连通差距越大。

另外一个拓扑性质是分割性。一个圆总可以把球面分成两块，它们互不连通，但都以这个圆为边界，但是，一个圆不一定能把环面分成两块，例如把自行车轮胎垂直剪断，它没有变成两段，还是一段，只不过变成一个弯管，两个圆有时也不行，但是，要是在环面上再剪一个圆圈，那么任意这样三个圆圈一定可以把环面分成两块（或两块以上），这个发现很有一般性，对于 g 个空洞的环面存在 2g 条闭曲线，不能把这个环面分成两块，但任意 2g＋1 条闭曲线，总可以把亏格为 g 的环面分成两块。因此我们把 2g＋1 称为曲面的连通度。这就是说，g 越大，把它们剖分开越困难。

同样，对于有 g 个空洞的环面来说，虽然我们在环面上能画 2g 条闭曲线，但最后环面还连成一片，但是 2g 条闭曲线必定是相交的，要想闭曲线互不相交，又能使划分结果连成一片，最多只有 g 条闭曲线，多了就一定相交，不信你试试看。

100．各种曲面
——闭曲面的分类

我们能摸得着、看得见的曲面多种多样，把它们加以分类是非常困难的事，但是如果我们只给出粗糙分类，也就是拓扑分类，有时还是可以完成的。从拓扑上看，曲面可以分为三类：开曲面、有边缘曲面、闭曲面。其中开曲面非常复杂还不能完全分类。有边缘曲面则依赖于闭曲面的分类。而经过长时期的努力，数学家对于无边缘的封闭曲面可以给出一个完满的拓扑分类，这是数学中很少几个漂亮的结果之一。

在分类过程中通常是找区别不同曲面的拓扑性质及拓扑不变量，当把这些性质和不变量找完全时，拓扑分类也就完成了。幸好，闭曲面的拓扑不变量只有一个，这就欧拉一庞加莱示性数 x，它可以从剖分得出，即

x＝顶点数－棱数＋面数它同直观的拓扑不变量——亏格 g 也有关系，

$x=2-2g$

区别闭曲面的拓扑性质是可定向性，也就是它是双侧曲面还是单侧曲面。因此，闭曲面分成两个系列，它们可以通过标准的方法造出来。

对于定向的闭曲面，我们可以从球面出发，加上若干环柄构成。环柄可以看成两头开着的圆管，在球面上开两个圆孔，再把圆管两端分别与这两个圆孔粘在一起，这样就得到了一个孔洞的环面，它的亏格为 1，x＝0，同样在球面上，其他地方再开两个圆孔，再接上一个环柄，就可以得到两孔的环面。总之，所有定向曲面都可以采用这个方法在球面安上若干环柄构成，它们的环柄数＝g，而 x＝2－2g。

不可定向曲面也可以如法炮制，不过这一次在球面开的圆孔上安装的不是环柄，而是莫比乌斯带。由于莫比乌斯带是一头封死的交叉帽，所以它的边是一个圆周，因此球面上每个圆孔上都可以安装一个交叉帽。球面上安装一个交叉帽就是射影平面 x＝1，安装两个交叉帽就是克莱因瓶（x＝0），因此安装 n 个交叉帽的不可定向曲面 x＝2－n

这样一来所有闭曲面不是同安装若干个柄的环面同胚。闭曲面的分类大功告成。对于有边缘的曲面，它们同胚除了所在的闭曲面同胚之外，还要求边缘的圆周数目相等。

101．有边无边的争论
——无边缘的流形

直到现在，我们还常常听到这样的说法：时间无限，空间无限。但是从拓扑学的观点看，这种提法是不严格的。从数学上讲，有限和无限至少包含两方面意思：

（1）有没有边缘，说时间无限，是指时间没有起始点，也没有终点，而对于空间来讲，有没有一个边界。

（2）度量上是有界还是无穷大，通常我们指空间的距离一直延续到无穷远，而且往往还推出宇宙中物质总质量无穷大。

另外在研究时空问题时，还多少有一个默认的假定，空间正如曲线、曲面一样嵌入或安装在一个大的空间之中。

我们姑且不谈这些观点与当代宇宙论及天体物理学的事实的背离，而只考虑数学的概念，上面的两个观点实际上并不是统一的，实际上存在着有界而没有边缘的曲线及曲面，例如圆及球面就是。在有边无边的问题上，人们在历史上已经犯过一次错误，这就是关于地球形状的争论。各民族最初都有所谓“天圆地方”的说法，只不过他们坚持这个说法多久有所差异。既然要坚持地方的说法，就要说明在可达到的范围之外是什么东西。有的讲是一望无际的茫茫大海，永无尽头，这就是无边论；有的讲有四四方方的边界，这就是有边论。现在知道，无论是无边论及有边论都是错误的，因为地球表面是有界但无边缘的球面。

现在的一些人再一次挑起这个争论，只不过是人们对于三维流形的几何学不像二维曲面那样可以摸得着、看得见罢了。但是拓扑学可以讲清楚，三维流形也正如二维曲面一样，存在着有界无边缘的流形，当然也存在着无界无边缘的流形和有边缘的流形。究竟我们的宇宙空间属于哪一种，就要由观察的事实来决定，数学家只能告诉你，可能性是多种多样的。

还有一个误区是说，如果我们的宇宙空间是个三维球，那它“外”面是什么？事实是不存在什么“外”面，那只是我们地球人的空间观念的延伸。我们总设想一个空间，把一切东西都装在里面。实际上从数学上看这没有必要，这就是内蕴几何学的观点，而且即使存在着外围空间，那里发生什么我们无从知晓，因为至今我们无法接收任何四维空间的信号。

102．庞加莱的妙法
——拓扑不变量

在多面体欧拉公式的启发下，我们如果碰到了一个图形 X，就不妨将它割成一些简单的多面体，然后数一数这样分割的结果，算一下顶点 V，棱数 E，面数 F，体数 K……最后计算出 V－E＋F－K＋……我们把这个数称为 X 的欧拉—庞加莱示性数。它是一个最重要的拓扑不变量。假如有两个图形，由一个可以连续变形（当然不允许撕破和粘接）成为另外一个，也就是两个图形从拓扑上看没有什么区别，我们就说它们同胚，也就是拓扑等价。显然两个图形如果同胚的话，它们的欧拉—庞加莱示性数 x 就应该相等。庞加莱正是用这种分割的办法来研究图形，从而创立了组合拓扑学（1895 年）。

但是，只有一个拓扑不变量 x 是否能够完全解决问题呢？设有两个图形 A 和 B，我们把它们分割，算出 x（A）和 x（B）。假如 x（A）和 x（B）不相等，我们就能够判断出 A 和 B 不同胚，也就是从拓扑上看起来它们不一样。但是，如果 x（A）＝x（B），是否就能够说明 A 和 B 同胚呢？上面我们看到（空心的）环面的 x＝V－E＋F＝0，算出克莱因瓶的 x 也等于 0，可是它和环面是多么不同啊！一个双侧曲面，一个单侧曲面，差别如此之大（当然不能相互变形），可是欧拉—庞加莱示性数都相等，都是 0。

这个例子说明只用欧拉—庞加莱示性数来研究拓扑性质，还嫌太粗糙一些，看来拓扑学家还必须寻找更加精细、更加有力的武器。首先庞加莱仿照顶点数、棱数、面数……定义了贝蒂数 b_0，b_1，b_2……挠系数 t_1，t_2…，…以及基本群 π_1，这些武器果然比欧拉—庞加莱示性数 x 要厉害，x 所不能区别开

的环面与克莱因瓶，用贝蒂数 b_1 就可以区别开，环面的 b_1 等于 2，克莱因瓶的 b_1 等于 1。而且亚历山大证明了它们也是拓扑不变量。庞加莱推测，贝蒂数、挠系数、基本群加在一起说不定可以完全解决流形的同胚分类问题，也就是说，如果两个流形的两组不变量对应相等，那么这两个流形同胚。遗憾的是，这个梦想很快就破灭了。1919 年亚历山大创造出透镜空间，并举例说明庞加莱所创造的全部武器还不足以区别开互不同胚的透镜空间，拓扑学家还得加紧努力。

本世纪 20 年代，伟大的德国女数学家诺特开创了抽象代数这个崭新的须域，抽象代数不局限于古老的研究对象，而去研究象群、环、域这些抽象代数结构，她启发拓扑学家 为什么一定要把拓扑不变量都弄成一个一个数呢？为什么不能把贝蒂数和挠系数看成一个群的元素呢？在她的影响下，拓扑学家开阔了眼界，把贝蒂数和挠系数又扩展成为同调群，这样一来，拓扑学的组合方法被推广成为代数方法，从而形成近 40 多年来的代数拓扑学。

103．走向更高的维数
——流形

微分流形是现代数学中最主要的概念之一。流形可以看成是曲线和曲面的推广。黎曼在 1854 年已经引进这个概念，他的观念是从物理上来的，曲线可以看成点运动的轨迹，曲面可以看成曲线运动地轨迹，这样下去，n－l 维流形运动的结果就成为 n 维流形。1895 年庞加莱在引进拓扑学的同调概念的同时，也引进了流形的概念。他的办法并不新鲜，是在 n 维空间中用 p 个方程来定义的（如果有边界的话，再加上一些不等式），满足这些方程的点的集合就是流形。从拓扑学的角度来看，这种定义来细致了，比如说，一个足球，你踢上一脚或踩它一下，它的方程马上就改变，但只要足球不破，它的拓扑性质并没有变化，另外，这种定义还要依赖于它所在的 n 维空间，每次研究都要拿个“ 容器 ”把它装在里面，实在多此一举。因此，我们希望用“ 内在 ”的性质来定义微分流形。外尔首先在他 1913 年出版的名著《黎曼面的观念》中给出二维复流形的内在定义，这种定义后来成为微分流形定义的模范。为了理解这个定义，我们可以举一个简单的例子：自行车内胎可被看成一个标准的环面，它可以通过一个圆周环绕一个点走一个圆周而生成，这是黎曼的想法，也可以通过三维空间的方程来定义，这是庞加莱的想法。现在“ 内在 ”的定义大致是这样的：新车胎是完完整整的一个好环面，在骑的过程中难免出现破洞，有了破洞就要补，补的时候大都用一块圆皮，这块圆皮不仅把洞补上，也把洞口旁边都给覆盖上。假如旁边再破一个洞，补这块洞的圆皮和原来那块圆皮就会有互相重叠的地方。如果这内胎用久了，就有可能每个地方上面都覆盖着圆皮，这种内胎骑起来和没补过的内胎也还是差不太多，它仍旧是一条内胎，但是已经成为（有限多块）互相重叠的补钉（圆皮）的集合了。我们描述内胎也就只需要了解：（1）补钉是什么，（2）补钉是怎样相互重叠的。这样就不必管它外围空间是什么，也不必考虑它的定义方程了。对于微分流形来讲，补钉就是欧氏空间（或其中一块开集）。补钉相互重叠的方法：有的是光滑的（好像是光光溜溜的新轮胎），这就称为微分结构；有的是见棱见角的，就称为分段线性结构；有的是不怎么光滑的，但还保持连续性（比如坑坑蹩蹩的），就称为拓扑流形。这样一来，流形从局部看来（就和补钉一样）都差不多，差别只是它们相互重叠的方式。比如你用几块补钉补满一个球，利用同样这几块补钉补满一条内胎，补钉之间从整体上看，它们的连接方式是不会一样的。

104．网络能布在平面上吗

——嵌入问题

在我们现在信息时代，最常见的就是复杂的电路，电路的结点可以成千上万，抽象来看，就是由顶点及其间连线构成的图。图可分成三类：一类图画出来，其中每条连线均彼此不相交，这类图称为平面图；一类图画出来时，连线有的相交，但是经过移动后，可变成没有相交连线的平面图，这种图称为可平面化的图，或者称为可以嵌入到平面上的图。还有的图无论怎么样改造，总也不能变成平面图，这种图称为不可平面化的图，最典型的不可平面化的图是 k_5 及 k（3，3）

一个有 p 个顶点，q 条连线的图可平面化的条件是什么呢？它必须满足

$$q \leq 3P-6 \quad (P \geq 3)$$

而且可平面的图一定包含一个顶点，它的阶（即过顶点的连线数目）≤5。由上面不等式我们可以证明 K_5 不可平面化。K_5 有 5 个顶点和 10 条连线，因此

$$10 > 3\times 5-6=9$$

与上面不等式矛盾。但 K（3，3）不可平面化需要其他工具。

一个线路能否布在平面上而不自相交叉，这类问题称为嵌入问题，如图 K_5 和 $K_{3,3}$ 两个线路图就无法布在平面上而不自相交叉。现代的电路十分复杂，我们当然希望知道它能不能布在平面上不自相交叉，要是不能嵌入在平面当中，交叉就要用分层、穿孔等办法弄到空间去，线路画起来、看起来都麻烦，这就要靠拓扑学了。1930 年，波兰数学家库拉托夫斯基证明：K_5、$K_{(3,3)}$ 两图是一个图不能嵌入到平面中去的仅有障碍，也就是说，一个图能够嵌入到平面中去的充分必要条件是它不含有像 K_5 和 $K_{(3,3)}$ 那样的部分图。

从拓扑上来讲，我们也可把图嵌入到球面和其他走向曲面之上。

105．现代数学的女王
——拓扑学

拓扑学研究几何图形在连续变形下的不变性质。1895 年庞加莱引进剖分方法以及一系列拓扑不变量，开创了组合拓扑学的新领域。经过半个世纪的努力．拓扑学与抽象代数相结合形成了代数拓扑学的新领域。1945 年以后的 20 年间，代数拓扑学引进一系列新工具、新技术，不仅使拓扑学本身得到发展而且成为研究所有数学领域，从数论、代数到几何分析、微分方程的必不可少的工具。特别是流形上一层一层的结构，如微分结构、复结构等的存在及分类问题，更是由拓扑学的不变量来鉴别，这些问题可以统一用纤维丛概念来概括，而纤维丛的示性类则是解决上述问题的有力工具。正是由于代数拓扑的工具以及纤维丛和示性类的概念与计算，把数学提高到一个崭新的阶段。因此著名数学家狄厄多内说“代数拓扑学和微分拓扑学是现代数学的女王”。

拓扑学取得这个重要地位，主要因为从 50 年代到 60 年代，拓扑学取得一系列辉煌成就：首先是纤维丛理论同伦论有着巨大突破以及托姆创立了配边理论，这实际上是流形的粗糙分类。1956 斩，米尔诺发现了今人惊异的事实：在七维球面上存在着互不相同的微分结构。1960 年斯梅尔证明五维以上流形广义庞加莱猜想成立，也就是流形如果同伦等价于球面，则它与标准球面同胚。其后又证明对于复合形和流形主猜想不成立。从此这导致在数学中，拓扑学无处不在，从代数数论到偏微分方程，从抽象代数到概率论，拓扑学中的主要概念和方法都在发挥着巨大的威力。同调代数学可以看成抽象代数和拓扑学的杂交产物，在现代的群论、环论、域论中，你很难不碰到拓扑学的术语，什么同调、上同调，这在 30 年前是连影子都没有的。偏微分方程的看家工具（不动点理论、拓扑度等）也都是从拓扑学中借去的，就连在拓扑学中比较生僻的“上同伦”也在最近的文献中出现了。现代数学不仅仅可从拓扑学中得到几个玄乎的概念，更重要的是能产生别的方法所根本办不到的深刻结果。1973 年德林证明了著名的魏伊猜想，靠的是引进新的上同调，这样得出的多项式丢番图方程的解数的精确估计，是解析方法所根本不能望其项背的。

最近，在物理学、化学、生物学甚至经济学中，拓扑学也取得重要应用。基本粒子理论中有所谓拓扑解、非拓扑解，拓扑学中的纤维丛和示性类理论仿佛是规范场理论定做的工具，分子生物学在研究主宰传宗接代的 DNA 的超螺旋结构时，碰到了像“环绕数”这种很占老的拓扑慨念，而且生物化学家也的的确确鉴定出来几个拓扑异构酶。70 年代以来拓扑学更是跨出数学领域。

分析

106．联系四个量的美妙公式

——欧拉公式

数学中最重要的四个量是1，虚数单位$i=\sqrt{-1}$，圆周率π和自然对数底e，这四个量都是绝对的常数。前三个的来源众所周知，而e却是个来头很不简单的量。

对于n个正数a，b，c，…，u，v，一个基本的不等式是它们的算术平均值总是大于或等于其几何平均值，算术平均值是所谓

$$\frac{a+b+c+\cdots+u+v}{n}$$

而几何平均值是

$$\sqrt[n]{abc\cdots uv}$$

那么总有 $\frac{a+b+c+\Lambda+u+v}{n}\geq\sqrt[n]{abc\Lambda\ uv}$

等号只有在a＝b＝c＝…＝u＝v时成立。从这个不等式出发，我们可以证明当a＞b＞0时，

$$(1+\frac{1}{b})^{b}<(1+\frac{1}{a})^{a}$$

同时

$$(1+\frac{1}{b})^{b+1}>(1+\frac{1}{a})^{a+1}$$

这样一来，当正的自变量x增大时，函数

$$\varphi(x)=(1+\frac{1}{x})^{x}$$

递增，而函数

$$\Phi(x)=(1+\frac{1}{x})^{x+1}$$

递减。而且对每x

$$\varphi(x)<\Phi(x)$$

随着x的值增大，ψ（x）值越增大，而φ（x）的值越减，从而ψ（X）和φ（X）的距离越来越接近。当X趋于无穷大时，中（X）、ψ（x）的距离趋于0，这样ψ（x）和φ（x）的极限值重合，这就是欧拉的e，e的值为

e＝2.718281828459045236……

同圆周率一样，也是一个无穷不循环小数，而且也是一个超越数。

对于每一个值z（虚值或实值），我们都可以把函数e^z展开

$$e^{z}=1+z+\frac{z^{2}}{2!}+\frac{z^{3}}{3!}+\Lambda$$

当z＝x＋iy时（x，y为实数），就得到欧拉公式

$$e^{x}+iy=e^{x}(\cos y+i\sin y)$$

当x＝0时

e^{iy}＝cos y＋inin y

当 $y=x$ 时，我们就得到著名的联系四个量的欧拉关系式

$$e^{i\pi}+1=0$$

107．不断反复的波形曲线

——三角函数

古希腊人早在公元 1 世纪研究天文时，就考虑一角所对的弧和弦的长度，由此逐渐形成角度的正弦、余弦等概念。到 15 世纪，可以通过三角形两边之比定义正弦和余弦并计算它们的数值。由于天文及航海的需要，人们花了很多力气进行正弦、余弦数值的汁算，由此造出精密的数值表，并引导到对数的产生。

1665 年牛顿首先得出计算正弦和余弦的一般公式，这时他的角度 x 的值是以弧度来表示的，即180度为π（弧度），90度为$\frac{\pi}{2}$，30度为$\frac{\pi}{6}$，如此等等，这样牛顿推出

$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots$$

$$\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots\cdots$$

由于这级数是无穷级数，在计算时可以只取有限多项，而且误差不会超过略去诸项的首项。这种方法实际上非常快也非常精确，例如计算 1° 角（$x=\frac{\pi}{180}$）的正弦，取前项

$$\sin x=x-\frac{x^3}{6}$$

得到的数值精确到小数点后 10 位。

直到这时还没有函数也没有三角函数的概念，因为这时首先要把角度扩大到 90° 以外，然后再根据公式

$\sin x=\sin[(2n+1)\pi-x]$

$\sin x=\sin(2n\pi+x)$　　$n=0,\pm1,\pm2\cdots\cdots$把角度扩展到所有角度，这样就得到正弦函数的概念。由于正弦函数以 2π为周期，因此，图像经 2π后完全重复，得出的是一个波形曲线，它在反映周期运动时非常有用，同样余弦函数的图像也是波形曲线。

同样，可以定义正切、余切、正割、余割函数，它们也都是周期函数，但是不是波形曲线，而且求它们的展开式也较为复杂，如

$$\sec x=1+\frac{1}{2}x^2+\frac{5}{24}x^4+\frac{61}{720}x^6+\frac{277}{8064}x^8+\cdots\cdots$$

$$\tan x=x+\frac{1}{3}x^3+\frac{2}{15}x^5+\frac{17}{315}x^7+\Lambda\Lambda$$

但是在 17 世纪时，却已得到反正切函数

$$\arctan x=x-\frac{x^3}{3}+\frac{x^5}{5}-\frac{x^7}{7}+\Lambda\Lambda\quad |x|\le 1$$

由此不难得出莱布尼茨级数

$$\frac{\pi}{4}=1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\Lambda\Lambda$$

虽然用这公式计算π是不合适的。

108．越来越快的增长

——指数函数与对数函数

在 50 年前电子计算机问世之前，繁复的计算对于人力是一个极大的挑战，因此寻求方便、快速、准确的计算方法是数学家责无旁贷的任务。而计算方法一次重大的突破是 17 世纪初对数的发明,对数把乘除简化为加减的运算，大大减轻了人们的劳动。而据此原理制造的计算尺长期以来一直是科学家和工程师的必备工具，其影响之大不可低估。

对数的原理很简单，基本上是把它指数化，例如要计算 xy，就把 x，y 变成常用对数，即

$x=10^{\log x}$　　$y=10^{\log y}$

这样

$xy=10^{\log x}10^{\log y}=10^{\log x+\lg oy}$

我们只须把 x 的常用对数求出即可。

在数学上常用的是自然对数，即以 e 为底的对数，我们可用 lnx 表示，它和自然对数的关系是

$\log x\times\ln 10=\ln x$

$$\log x=\frac{\ln x}{\ln 10}\text{Mln}x$$

其中 M＝0.4342944819……

自然对数乘以 M 后才得到常用对数的值。17 世纪中叶，出现了计算对数值的公式，即

$$\ln(1+x)=x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4+\Lambda\ \Lambda$$

由于这个级数收敛太慢，不适合计算，我们改进成

$$\ln\frac{1+x}{1-x}=2[x+\frac{1}{3}x^3+\frac{1}{5}x^5+\Lambda\ \Lambda$$

它收敛很快，只需要很少几项就可以相对精确地求出对数值来。

对数函数的反函数就是指数函数，我们常用增长得越来越快的情形称为指数增长。指数函数是最晚得到的函数，但是它最简单，而且与三角函数和对数函数密不可分。指数函数的展开式是

$$e^z=1+z+\frac{z^2}{2!}+\frac{z^3}{3!}+\Lambda\ \Lambda$$

当 z＝iy 时

$$e^z=e^{iy}=1+iy-\frac{y^2}{2!}-i\frac{y^3}{3!}+\frac{y^4}{4!}+i\frac{y^5}{5!}-\Lambda\ \Lambda$$

$$=\ (1-\frac{y^2}{2!}+\frac{y^4}{4!}-\frac{y^6}{6!}+\Lambda\ )\ +i\ (y-\frac{y^3}{3!}+\frac{y^5}{5!}-\frac{y^7}{7!})$$

＋……)

由 cos y 及 sin y 的展开式，可知

e^{iy}＝cos y＋i sin y

这就是欧拉公式。

109．分析的中心

——函数

函数是数学分析的中心概念。从某种意义上来讲，数学分析就是研究函数的演算及其性质的。但是数学分析开创时期，函数的概念并不清楚，当时的分析是用几何的语言或算术一代数的语言来叙述的。1692 年莱布尼茨才首先用函数这个词，其后经过约翰·伯努利、欧拉、拉格朗日、柯西、狄利克雷、黎曼一直到康托尔及戴德金，经历了 200 年才得出现代一般的抽象的函数观念。

虽然一般的函数概念还不太清楚，数学家从 18 世纪起已开始对具体的函数进行研究。当时的主要倾向是把函数用计算公式表示出来。由于多项式有局限性，把函数展开成幂级数是最常见的，这在当时形成一个数学分支——代数分析。第一本代数分析的著作是欧拉的《无穷分析引论》第一卷（1748 年）。

到 19 世纪初，人们熟悉的函数并不多，除了多项式之外，还有指数函数、对数函数、三角函数、反三角函数、双曲函数等。对它们的幂级数展开以及无穷乘积表达式有一定了解。随着解微分方程方法的发展，通过微分方程定义的函数也多起来，其中特别是圆柱函数（贝塞尔函数）和球函数（勒让德函数）以及它们的推广——超几何函数。另外，通过积分表示产生出 B 函数及 Γ 函数，它们一般有非常好的性质，如可展开成收敛的幂级数。而在整个 19 世纪，研究最充分的就是椭圆函数。真正确切的一般函数定义是狄利克雷给出的。他还第一个引进著名的狄利克雷函数，即

$$f(x)=\begin{cases}0，当X取有理值\\1，当x取无理值\end{cases}$$

这是一个在无穷多不孤立的点的不连续函数，这种函数对于 19 世纪的数学分析已经足够一般的了。狄利克雷的函数定义在两个重要方面改进欧拉的定义，一是他明确指出函数不一定非得具有解析表达式。二是他把不连续函数纳入函数的范围。这两方面对后来分析影响是巨大的。但他的定义仍局限于实变量的实函数，后推广为复变量的复函数，对于定义域及值域的推广来自康托尔及戴德金。康托尔把函数看成对（a，b）的集合，其中 a∈A，b∈B，A 和 B 均是集合。戴德金在 1888 年出版的《数是什么，数应该是什么》中，把函数作为集合间的映射来定义，而映射是一种对应规则，在这个规则 Φ 之下，集合（他称之为系统）S 中元素 S 对应于确定的对象 Φ（S），但他没有对值域有任何刻划。

110．速度与切线

——微分法

微积分出现前，有四种类型问题与微分法的产生有关：

（1）已知物体在一定时间之内走了多少距离，求物体的速度和加速度。这类问题直接与运动有关系。如果物体以匀速直线运动，问题当然很简单。但是如果物体运动时快时慢，就不能用总的时间去除总的距离来计算速度了，因为这样得到的只是平均速度。因此就要把时间间隔缩小，因而在这段时间间隔物体所走的距离也很小，当都小到 0 时，就出现 0/0 的困难，如何求瞬间速度成为主要问题。

（2）求曲线的切线，这不仅是几何学问题，而且在应用上极为重要，比如光线通过透镜反射或折射，都要确定透镜曲面上每一点法线的位置，法线与切线垂直，所以要求知道怎样求曲线的切线。在曲线运动中也有求切线的问题。

（3）求极大值和极小值的问题，比如求炮身的仰角多大，打出炮弹的射程最远。求行星的椭圆轨道的近日点和远日点，都是求极大值和极小值的问题。

（4）求曲线乃至曲面的弯曲程度即曲率等。

微分法的发现经历了漫长而曲折的历程，至今已是很简单了。所有上述问题归结为求导数问题。设 y＝f（x）是光滑的函数，则其导数的求法可分两步走，一是把 f（x）表示成多项式或幂级数，二是求 f（x）＝ax^n 的导数，经过许多人的努力

$$f'(x)=\frac{d(ax^n)}{dx}=anx^{x-1}$$

而幂级数的导数等于每一项导数之和。在 17 世纪，导数就是这么简单地求出来，由此可以得出，速度 v

$$v=\frac{ds}{dt}$$

加速度

$$a=\frac{dv}{dt}=\frac{d^2s}{dt^2}$$

平面曲线 y＝f（x）的切线与 x 轴夹解 A，则切线的斜率为 tanA，

$$\tan A=\frac{dy}{dx}$$

由此不难求出任意点的切线。

111．面积与体积

——积分法

涉及积分的问题要比涉及微分的问题产生早得多，它最直接的来源就是求面积。直线形的面积不难求出，但曲线围成的面积就比较困难，首当其冲的是圆的面积。在古希腊公元前 5 世纪最有名的问题是求月形的面积，而这个问题到 20 世纪才完全解决。阿基米德和他的前人曾用“穷竭法”求面积和体积，取得了重要的成就，但是到 17 世纪微积分问世之前，一般曲线围成的面积还没有系统的方法。另外，推动积分法产生的问题还有求曲线的弧长，一般物体的体积和表面积以及与这类问题有关的求物体重心等。

虽然积分法源远流长，但微积分发明之后，一般函数的微分法并不太成问题，而一般函数甚至很简单函数的积分都不是容易的事，积分法仍是数学家研究的主要对象之一，而且积分的概念对整个数学的发展也至关重要。

积分法为不定积分和定积分，一个简单函数如多项式及 x^n 的积分的不定积分

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C \qquad n \neq -1$$

$$\int x dx = \log x + C$$

更一般讲

$$\int f(x)dx = F(x) + C$$

F（x）称为 f（x）的原函数，而 C 是任意常数，在求弧长面积时通常用定积分

$$\int_a^b f(x)dx = F(b) - F(a)$$

从历史上讲，这些公式也不是一下子得出来的。例如 17 世纪初开普勒就对体积问题很感兴趣，他注意到酒商求酒筒体积的方法不精确，他用切成小圆锥的方法来计算球的体积。卡瓦利埃利在这方面迈进了一步，他摆脱掉穷竭法，把曲线下的图形看成平行线段组成的，面积就看成许许多多无限狭小的长方形的面积之和。其后许多人的积分概念实际上就是无穷小量的求和。因此微积分在当时被称为无穷小演算，用这个方法得到了重要的积分公式。

112．开门与关门——微积分基本定理

牛顿与莱布尼茨的微积分比起前人来显著的特点是其普遍性，也就是建立起一个普遍算法，它能处理各种特殊问题。概括说来，求速度、作切线、求极大极小、求曲线在某点的曲率等问题可统一用微分法来解，而求面积、求曲线弧长、求体积、求重心、求距离等问题可以统一用积分法来解。以前大都特殊问题采用特殊解法，它们之间的联系是模糊的。

其次，牛顿及莱布尼茨建立了微积分的一般的算法体系，这个算法体系在莱布尼茨那里十分清楚，他明确给出“函数的”和、差、积、商、幂、根的微分公式：

（1）如 y=v±w≠a，则 dy=dv+dw±0

（2）如 y=avw，则 dy=avdw＋awdv

（3）如$y=\frac{v}{aw}$，则$dy=\frac{wdv-vdw}{aw^2}$

（4）如$y=w^z$,则$dy=zw^{z-1}dw$

（5）如$y=\sqrt[z]{W}$，则$dy=\frac{\frac{1}{z}dw}{w\frac{z-1}{z}}$

积分也有相应的公式，这样微积分的整个算法体系建立起来了，为它成为解决实际问题的有效工具奠定了基础。

另外，他们都为进一步发展微积分本身指出了方向，他们都提出过换元方法，莱布尼茨特别指出 ydx 的换元下不变性，这是极为深刻的观点。他们都用过无穷级数，牛顿十分强调其普遍意义，这样为一般问题的解决提供了前景。

最重要的是，他们建立了“微积分基本定理”，在当时，微分法和积分法是如此不同，以致很难看出它们之间的联系，一个是无穷小量之商，一个是无穷小量求和。在函数出现之后，明显看出这两个运算的互逆性，就像加法与减法、乘法与除法一样。具体说，如果

$$\frac{dF(x)}{dx}=f(x)$$

则 F（x）就是 F（x）的原函数

$$\int_z^b f(x)dx=F(b)-F(a)$$

而微积分基本定理就是

$$\frac{d}{dx}\int_z^x f(dx)=f(x)$$

113．多项式走向无穷——泰勒展开

当一个函数是一个代数函数，也就是

$$y=f(x)$$

可以用多项式表示出来时，求它的数值、它的极大极小，求函数的导数和积分都是不难的，可以代入现代公式。但是当函数是超越函数时，通常是不能用多项式来表示，这时运算就带来不少困难。实际上连计算函数值也很困难，这在造对数表、三角函数表时已经碰到。因此当我们把多项式扩展到无穷，就成为一个幂级数。如果幂级数在某一个区间内收敛，那它在区间内就代表这个函数。因此，个函数展成幂级数的问题就变成确定它的系数的问题。例如 $f(x)=a_0+a_1x+a_2x^2+\cdots+a^nx^n+\cdots$

在非常理想的情况下，函数非常光滑，可以逐项微分，那我们就可以得到

$$f'(x)=a_1+2a_2x+3a_3x^2+\Lambda+na_nx^{n-1}+\Lambda$$

$$f''(x)=2a_2+6a_3x+\Lambda+n(n-1)a_nx^{n-2}+\Lambda$$

$$f'''(x)=6a_3+\Lambda+n(n-1)(n-2)a_nx^{n-3}+\Lambda$$

因此，令 x=0，我们就得出每项系数

$f(0)=a_0$

$$f'(0)=a_1$$

$$f''(0)=2!a_2$$

$$f'''(0)=3!a_3$$

……

$$f^{(n)}(0)=n!a_n$$

这样我们就得到 f（x）的幂级数展开式

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\Lambda+\frac{f^{(n)}(0)}{n!}x^n+\Lambda$$

同样，f（x）在收敛区间内，可展成（x-a）的幂级数

这个幂级数称为函数在 x=a 时的泰勒展开。

利用泰勒展开，很容易计算函数的近似值，例如

$$ex=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\Lambda+\frac{x^n}{n!}+\Lambda$$

它对所有 x 都收敛，令 x=1，取子项就可以把 e 计算得精确到四位小数，同样用

$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\Lambda$$

它对一切 x 均收敛，取三项就可算准到第五位数。

由于它们可以逐项积分，对于复杂的积分可以很容易算出。

114．半个世纪的争论——傅立叶级数

把函数表示成幂级数当然是最自然不过的事。但是，能用幂级数表示的函数并不多，因为要想用幂级数表示，起码这函数得无穷次连续可微，而且即使无穷次可微也不一定就行。这样，要想用无穷级数表示一般的连续函数

就得另觅他途。我们熟悉的比多项式稍复杂的函数就是三角函数，三角函数图像清楚，有表可查，当然是最有力的候补者。但是，一个任意的（连续）函数能否被三角级数表示，在历史上还是有一番争论的。这个争论的一方是丹尼尔·伯努利早在 1732 年他就从物理上说明振动弦有较高的振动模式，在后来的论文中说明振动弦的许多模式能够同时存在，他进而认为所有可能振动的初始曲线都可表为

$$f(x)=\sum_{n=1}^{\infty}a_n\sin\frac{n\pi x}{I}$$

因为有足够的常数 a，使级数适合任一曲线。他并没有能给出数学上的证明，克莱洛在 1757 年论文中支持了伯努利的观点。与他们相对立，欧拉及达兰贝尔从不同的角度反对他，欧拉在 1753 年一篇论文中说，正弦函数总是一个奇周期函数，但是任意的函数肯定不可能被如此表示。达兰贝尔在 1757 年出版的《百科全书》第 7 卷中甚至不相信所有奇周期函数也能表为如上的三角级数，因为这级数是二次可微的，而任意奇周期函数未必如此。这争论持续了二、三十年，1759 年拉格朗日也参加了争论，他用自己的计算支持欧拉，但他一直坚持任意函数不一定能展成三角级数。1779 年拉普拉斯也加入争论，他支持达兰贝尔，他们在函数的三角级数表示问题上都未能得出数学上正确的结论。这一直到傅立叶也彻底解决。但是，18 世纪的大数学家都得出过特殊的函数的三角级数展开式以及系数表达式。例如欧拉在 1777 年就从三角函数的正交性得到三角级数的系数。即从

$$f(x)=\frac{a_o}{2}+\sum_{n=1}^{\infty}a_k\cos\frac{k\pi x}{I}$$

得出

此即傅立叶级数系数表达式。傅立叶在论文中，坚持认为任意函数可以展成三角函数的级数。

后来，数学家为傅立叶级数理论打下严密的基础。傅立叶级数在函数概念、积分概念、三角级数的展开及收敛性都大大扩展了以前的分析领域，最后推动集合论、测度论和积分论乃至泛函分析的发展。

115．函数值的计算——函数逼近论

在实际生活中，我们总是需要知道具体的数值，测量的数量很多无法求得其精确值，这时我们总是用小数或分数来逼近无理数。例如$\sqrt{2}$或π

对于函数，也有类似的情况，实际上遇到的函数，常常没有比较简单的表示方法，在具体运算和求函数值时，我们往往同样需要进行逼近，用来进行逼近的函数当然要求是比较简单的，什么是比较简单的函数呢？最简单的是多项式，其次是有理函数，再就是三角函数的多项式。这些函数的好处，特别是多项式，可以很顺利地在计算机上运行，是逼近任意连续函数的最好工具。但是重要的是逼近总是产生偏差，因此对于偏差的估计至关重要。

原始的逼近方法，使用插值多项式，例如用 n 次多项式来逼近 f（x），我们选 n＋1 个点，使 f（x）与 P（x）在这些点取值相同，即，由此可以求出逼近多项式 P（x），即：

这就是拉格朗日公式，其后利用泰勒展开及余项公式可以求出逼近多项式及偏差。但是这些只对较好的，也就是相当光滑的函数才行得通。而确实有一些坏的函数，不能用这样的方法逼近。

函数逼近论的基本定理是外尔斯特拉斯在 1885 年证明的 在有界闭区间（a，b）上的任何连续函数 f（x）都可用一个多项式（或三角多项式）来一致逼近，这是存在性定理。进一步研究还需要研究被逼近函数 f（x）与逼近函数（比如说 n 次多项式 P_nX）之间的误差。最小的误差弥为最佳逼近值。1918 年哈尔给出线性无关连续函数组ψ_0，$\psi_1\cdots$，ψn 实现最佳逼近的函数唯一性的充分必要条件，由此推出实现最佳逼近的 n 次多项式及三角多项式是唯一的。最佳逼近理论在 20 世纪初由俄国—苏联数学家伯恩斯坦、美国数学家杰克逊、瓦莱·布桑等所奠定，特别是 1911 年杰克逊首先得出三角多项式 n 阶最佳逼近值的上界的阶的估计，开辟了定量理论的新方向。其后法瓦尔、柯尔莫哥洛夫等人继续得出更精确的结果。

116．处处没有切线的连续曲线

——不可微连续函数

19 世纪以前的数学，对于微积分的许多概念并没有严密的理解，许多重要的概念仍诉诸几何直观。函数的概念仍不能完全脱离曲线的直观，函数的求导和微分与曲线的切线看成是一回事，函数的连续性无非就是曲线的连续性，这样，对数学分析进一步发展无疑有很大的阻碍。

柯西第一个为数学分析奠定严密的基础，他的功绩之一是为函数的连续性和可微性给出严密的定义，也就是建立在极限观念的基础上。不过在连续性与可微性的关系上，仍然没能摆脱以前的看法。当时普遍认为连续函数只要没有间断点，一定可导。著名的物理学家安培甚至多次给出证明，处处连续的函数一定处处可微。换言之，连续曲线一定处处有切线存在，而且这些错误的证明长期保留在教科书当中，直到 19 世纪末。

最早认识连续性与可微性的区别的是哲学家波尔查诺,他在 1834 年首次给出一个连续函数，处处没有有限的导数。但是他举出的例子是没有解析表达式的曲线，这在当时也不为人接受，他的结果没有什么影响，其严格证明到 1922 年才发表。因此真正区别开连续性及可微性的主要数学家是黎曼和外尔斯特拉斯。

黎曼早已认识到由函数的连续性不必然推出可微性,早在 1854 年就在他为取得讲师资格论文中定义了一个连续函数，在任意小区间内，都有无穷多点没有导数。1860 年他引进一个连续且处处不可导的函数。长期以来认为没有什么问题。一直到 100 多年后，才发现在许多点处可导，因此真正处处不可导的连续函数首先是外尔斯特拉斯在 1872 年讲课中首先宣布的，此即

其中a是一个奇整数，$1<b<1$，$ah>1+\frac{3\pi}{2}$。后来许多数学家又发现许多新的例子。值得注意的是，著名波兰数学家巴拿赫证明在所有的处处连续函数当中，处处不可微的连续函数占有绝大多数，比较起来可微函数的比例小得可以忽略不计。

117．走向极端——极大极小问题

无论在科学技术领域还是在日常生活之中，我们常常要遇到极大、极小、最好、最坏、最优、最省的问题。在数学中，也有一些极值问题的定理，如平面上两点间直线最短，球面上两点间大圆弧最短，平面等长闭曲线中圆所包围的面积最大，空间等表面积闭曲面中，球面围成的体积最大。在物理中，我们有许多变分原理，例如光程最小原理，由此可推出几何光学的反射定律及折射定律。在实际生活中，这类问题更多。例如，造船、汽车、飞机都要使阻力极小，造容器使材料最省等等。这类极值问题，原来是具体问题用具体方法求解，但到 17 世纪之后，发明两种不一般的方法为极值问题提供解答：

一种是微分法，它处理函数的极值问题。

一种是变分法，它处理函数族的极值问题。

微分法考虑的是求函数值的极大或极小，如求函数 y=f（x）的极大极小值，首先求满足

$\frac{dy}{dx}=0$的点P。

如果在P点处 $\frac{d^2y}{dx^2}$，则P点为极小点。

如果在P点处 $\frac{d^2y}{dx^2}$，则P点为极大点。

如果在P点处 $\frac{d^2y}{dx^2}$，则P点为拐点。

当函数是多元函数 z=f（x，y）时，我们求达到极大或极小的点（x_0，y_0）的方法是用偏导数

偏导数 $\frac{\delta}{\delta x}$ 的意思是对x；求导数，而把y看成常数不予考虑。例如造一个无盖的长方形盒，体积为 V、盒底长 x，y 是多少，材料才用得最省？显然盒子的高度

$$h=\frac{V}{xy}$$

盒子表面积 $S=xy+\frac{V}{xy}(2x+2y)$

要使它达到极小

这样我们就得到最省料盒子的长、宽、高为

变分法的问题，参见 118、119 两条。

118．一根绳子圈出的面积——等周问题

等周问题是最古老的变分法问题。它是在所有给定周长的平面封闭曲线中，求其中一条曲线使它所围成的面积最大。这个问题可追溯到古希腊以前的时代的一个传说。据说，古代腓基尼的公主狄朵离开家乡，定居北非的地中海沿岸，在那里，她可以用一张牛皮圈出一块地归她所有，当然，圈的土地越大越好。聪明的狄朵把牛皮切成很细很细的牛皮条，然后把牛皮条一根一根缝起来，这样就形成一条相当长的牛皮带，用它去圈土地。当时狄朵圈的土地靠海，因此有一半不用牛皮带，她正确地把牛皮带圈成半圆，这样她围成的土地面积是最大的，因此，这个问题常称为狄朵问题，但是，它还不是严格意义下的等周问题，因为狄朵的曲线不是封闭的。

真正的等周问题来源于公元前 2 世纪的芝诺道卢斯，他写了一本等周形的书，但已失传，其中证明等周问题的一些基本定理。

（1）周长相等的 n 边形中，正 n 边形的面积最大。

（2）周长相等的正多边形中，边数越多，其面积越大。

（3）圆的面积比同样周长的正多边形面积大。

（4）表面积相等的所有立体中，以球的体积为最大。

这些结果一直到 18 世纪才得到严格的证明。1744 年，欧拉用统一的方法处理变分问题后，又得出一系列新的等周问题的结果，例如欧拉得出旋转极小曲面（面积极小）是悬链面。

等周问题更精确的结果是等周不等式，它反映各种几何量之间的关系；例如，若尔当曲线 J 周长为 L 所围成的面积为 F，则有不等式

$$L^2-4\pi F\geq 0$$

等号则只当曲线为圆的时候成立，这个不等式还可以更进一步精密化：如果内接于 J 的最大圆的半径为 r，外切于 J 的最小圆半径为 R，则

$$L^2-4\pi F\geq (L-2\pi r)^2$$

$$L^2-4\pi F\geq (2\pi R-L)^2$$

$$L^2-4\pi F\geq \pi^2 (R-r)^2$$

成立。三维情形问题较难，对于卵形面。表面积为 S、内部体积为 V，则

$$S^3-36\pi V^2\geq 0$$

等号仅限于球面的情形成立。

等周不等式现已发展成为一门专门学科。一方面，它已推广到高维情形，另一方面，它已经包括图形的各种几何量及物理量，成为一个至今还在蓬勃发展的领域。

119.最快地逃离火海

——最速降线问题

在高层建筑失火时，首要问题是把高层的居民尽快地运送到安全地区。当有一条长软带总可以使人踩在上面滑到地面。出现灾情时，时间就是一切，什么样的曲线使人能最快地逃离火海呢？这就是所谓最速降线问题，也称捷线问题。

1696 年瑞士数学家约翰·伯努利提出这个著名的问题，向当时的著名的数学家挑战。他这样表述这个问题：“ 设想一个质点沿着一条没有摩擦的曲线只在重力的作用下由 A 点向较低的一点 B 下滑，那么沿着什么样的连接 A 和 B 的曲线，使该质点下滑所需时间最短？ ”

显然，直线不是最速降线，伽利略曾认为是圆弧，这也是错误的。1697 年牛顿、莱布尼茨等五位数学家独立地得出正确的解答：最速降线是上凹的旋轮线，由此产生数学的一个新分支——变分法。这门学科也是研究极大极小问题的，但是它与微分法中所讨论的极大极小问题有所不同。在微分法中的极大极小问题中要求极大极小的量只依赖于一个或几个数值变量，问题是求函数的极大极小值。而在变分法中，要求的极大极小的量，例如最速降线中的时间，却依赖整条的曲线（或函数）。因此这个量是曲线或函数的函数，后来我们称为泛函。所以变分法研究泛函的极值问题，它要求的是函数或曲线。从这个观点看，最古老的变分法问题不是最速降线，而是等周问题。但是最速降线的解是不久前发现的旋轮线，特别引起大家注意，因为旋轮线是 17 世纪下半世纪数学家关注的中心，也是引起他们争吵的金苹果。

旋轮线是圆沿着一条直线滚动时，圆上一点的轨迹。在此之前，已经知道它的许多有趣性质。大科学家惠更斯发现，一个质点（例如一个光滑的小球），在没有摩擦的情况下，无论把它放在旋轮线的哪一点上，它受重力作用来回振动，那么它的振动周期与振幅无关。因此旋轮线被称为等时性曲线或摆线。通常钟表的摆走的是一个圆弧，它的等时性或者说与振幅无关性只是近似的正确，300 多年来用摆制造的机械钟并不能说是绝对的准确，从而后来为石英钟、原子钟所取代。

120.从有穷代数到无穷代数

——数学分析今昔

从本质上来看，数学分析与代数并没有什么不同，实际上都是一种符号演算的技术。如果说有什么不同，那就是代数学一般涉及有限的数量，而分析涉及无穷的步骤以及无穷大、无穷小之类的量。这是由于代数方法也促进形式化的推广，扩大了函数的范围（如分指数和负指数的出现），也推动了无穷级数、无穷乘积、无穷连分数等“无穷”手段的出现，而这正是无穷小演算所要求的。

17、18 世纪的数学分析研究的主要都是演算，如微分演算、积分演算以及函数演算和变分演算。后来还有概率演算、算符演算、向里演算、张量演算等等。这种演算如同代数中的四则运算一样，当它们有了基础之后，分析中的主要问题就成了解微分方程了。求解微分方程是 18、19 世纪数学分析的主要课题，它在各方面有着重要的应用。正如代数学一样，为了求解方程，需要对方程的解（数的理论）及方程本身（方程论）进行深入研究，数学分析相应也对函数的理论及微分方程的理论进行研究。一位伟大的意大利数学家沃尔泰拉曾讲过“19 世纪是函数论的世纪。”的确，函数是数学分析研究的中心概念。不过，函数的研究远比数要复杂，它需要建立严格的基础，这是 19 世纪数学分析的第三条研究主线。分析基础的奠定导致现代数学的基本领域的形成：集合论、一般拓扑学与测度论。

在 19—20 世纪之交，数学分析的主流包括下面四大领域：

（1）现代数学分析的基础。

（2）一般函数论。

（3）泛函分析，它可以看成变分法的后裔，它处理函数集合上某函数（称泛函），泛函分析要考虑函数集合的各种结构，它们构成各种函数空间。这些函数空间及其上的算子是泛函分析的主要内容。特别值得一提的是广义函数论有特殊重要意义。

（4）一般微分方程论。

第二次以界大战之后，数学分析有了飞跃的发展，并表现出下列三大趋势：一是从单变量到多变量，二是从局部到整体，这表现在 60 年代中大范围分析的建立，三是从线性到非线性。而在方法上，也从单一分析过渡到与几何拓扑方法的统一，它预示着 21 世纪的大发展。

概率统计及运筹学

121.政治算术

——统计的起源

统计数字已经是现代社会不可缺少的因素了。它之所以重要，是因为它对社会的状况有一个整体的、定量的了解，而且也为各国的比较与历史的演进提供了素材。近代统计学一词来源于 17 世纪的国势学，这门学问主要是描述国家的状况、借助于表格和数字使描述一目了然，这对于各国政府无疑是决策的重要依据。

其实，这种古老意义下的社会统计，可追溯到很久以前，中国的历代史籍都有钱粮人丁户数的记载，显然政府对于它管辖下的人口数目很感兴趣，因为它必须征兵，征丁服役以及征收钱粮。但是，这种意义下的统计实际是普查，而不是现代意义下的统计。普查显然也是极为重要的，因为任何统计均有误差。普查在严密进行下所得数字是精确的，而且对于寻求规律性至关重要。因此现在普查的范围也日益扩大。从 1790 年起、美国每隔十年进行一次人口普查，其他一些领域普查的时间间隔更短。不过组织好普查并不容易做到。

但是，近代意义下的统计与普查不同：

（1）近代统计应用概率的数学方法，而普查只是简单的计数；

（2）统计采取抽象的方法，然后由样本对总体进行推断；

（3）统计注意大量观察结果的规律性，强调它的统计意义；

（4）为了更精确地进行统计，必须发展越来越先进的统计技术和方法。

1662 年，格劳特使用调查员调查伦敦市的死亡人数，是历史上最早出现的统计推断,而且他据此做出第一个现在人寿保险公司必备的死亡率表。1690 年，英国著名经济学家威廉·配第沿袭格劳特的方法，通过统计进行社会之间的相互比较，它的结果发表在《政治算术》一书中，这可以说是现代统计的滥觞。

122.赌场产生的问题

——概率

我们生活在一个变动不定的世界里，其中充满了不确定性。如果我们要对未来的事情发展做出预测，我们必须衡量这种不确定性。不确定性最常用的度量就是概率（过去也称几率、或然率，表示可能发生或不发生的程度）。

由于事件千变万化，如何计算概率需要具体问题具体分析。而只有我们有了概率模型之后，我们可以根据一定的原则计算概率，因此概率论一开始称为概率演算。从历史上看，概率演算的问题始于 1654 年帕斯卡与费尔马的通信，他们讨论的问题是关于赌场中的一些问题。因为赌博的模型最简单，而且大家也公认一颗骰子掷出某一点的概率是$\frac{1}{6}$。但是再稍微复杂的问题，概率的计算在当时有很大争议，而这种争论推动概率演算的产生。

帕斯卡的问题来源于一个赌徒，这位赌徒凭着多年赌博的经验认识到，掷一个骰子 4 次至少得出一次 6 点的概率要比掷两个骰子 24 次至少得出一次双 6 点的概率大。他请教帕斯卡，帕斯卡从理论上解决了这个问题。

他说，掷出一次6点的概率$\frac{1}{6}$，

掷不出6点的概率是$\frac{5}{6}$，

掷4次都掷不出6点的概率 $=1-\left(\frac{5}{6}\right)^4>\frac{1}{2}$

同时掷出两个6点的概率 $=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36}$

同时掷不出两个6点的概率 $=\frac{35}{36}$

24次都掷不出两个6点的概率 $=\left(\frac{35}{36}\right)^{24}$

24次中至少一次掷出6点的概率 $=1-\left(\frac{35}{36}\right)^{24}<\frac{1}{2}$

因此帕斯卡的理论证实了赌徒的结论。

引起更大争论的问题是所谓分赌本问题。甲、乙两方以 2000 元为赔本，不管赌仆么，赌一次赢者得 1 分，输者不得分，一直到谁先得 20 分谁获胜，就得到全部赌本。问题是在没有赌完的情况下，如何分赌本才算公平。例如甲再得2分，乙再得3分就可获胜，这时如何分赌本呢？如果不动脑筋，甲得$\frac{3}{5}$，乙得$\frac{2}{5}$，即6：4。帕斯卡的解法是，再玩4盘，肯定就可以得出最后结果，这时一共有 16 种情况（用 A 表示甲赢，用 B 表示乙赢）：

AAAA AAAB AABB ABBB BBBB

AABA ABAB BABB

ABAA ABBA BBAB

BABA

BBBA

显然前面 11 种情况甲赢，而后 5 种情形乙赢，因此，赌本应该照 11：5 来分配。这个结果一出来，许多人就提出不同意见，他们的意见表面上看来也有道理，他们说，只要一方赢了，比赛就终止了，所以真正出现的情况只有 10 种：

AA，ABA，ABBA，BAA，BABA，

BBAA，ABBB，BABB，BBAB，BBB

其中 6 种甲赢，4 种乙赢，也是 6：4。到底哪种结果对，这就牵涉概率论的基本原理：一是独立性原理，两个事件只有独立，它们的概率才等于每个事件的概率的乘积。二是等可能性原理，前面 16%是等可能性，而且只要玩完 4 盘，只有这 16 种情形。但后面 10 种虽然现实，但不是等可能性，也就 是概率不相等。因为玩二盘每种情况出现的概率 $=\frac{1}{4}$，玩三盘的概率 $=\frac{1}{8}$，玩四盘的概率 $=\frac{1}{16}$。这样甲赢的概率 $=\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{8}+\frac{1}{16}+\frac{1}{16}=\frac{11}{16}$。乙赢的概率 $=\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{8}=\frac{5}{16}$，同帕斯卡分析的一样.

123.观察与实验的必经之路

——最小二乘法

在近代数理统计学还没有建立之前，科学家已经用统计方法来处理他们观测和实验所得到的大量数据了。当时没有计算机，数据的处理都靠手算，他们处理数据的目的也希望由这些数据得出各种物理量之间的某种确定关系。例如，一定量的气体，压力、体积与温度之间的关系。但是由于观测和实验的误差，这些数据不都恰好落在一条直线或一条光滑曲线（或曲面）上。这样我们需要找一条曲线，使它最接近理想的曲线，这个方法就是最小二乘法。最小二乘法是 200 年前由勒让德和高斯独立发现的。他们的思想也很简单，用波瓦亚的例子说，把一头牛放在湖的中心，牛从哪儿上岸呢？显然这种哑巴动物一定会选择离它最近的岸边做为它的目标。用坐标来表示，那就是使距离

$$D = \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

极小的地方，这里（X_0，y_0）是牛的所在位置的坐标，（X，y）是要上岸的目的地的坐标。

由这种思想出发。我们所要求的曲线希望同所有数据

（x_1，y_1），（x_2，y_2），……（x_n，y_n）的距离极小，也就是当所求曲线为

$$y=f(X)$$

时，这些点到这条曲线的距离最小，但是用坐标之差表示的距离的值有正，有负，加在一起时有时互相抵消，因此，我们的要求是距离平方之和最小，这也就是最小二乘法的名称的由来。例如当 x，y 是正比关系时，即

$$y=\alpha x$$

我们只需求出 α 来，为此，使

$$\sum_{i=1}^{n}(y_i - \alpha X_i)^2$$

极小，这用初等方法或微积分就可以解出来。这种方法在统计上也很有用处，因为样本的平均值就是总体平均值的最小二乘估计。

124.数据的描述

——总体与样本

在科学和社会的实践中，我们对于要描述的对象的集合中的每个个体总要联系上一个数，这样就得到一大堆数据。个体的集合称为总体，而总体的一部分称为样本。实际上，我们很难对总体进行全面的描述，而且往往也没有这个必要，例如全国小麦每亩产量。实际上，我们只是从样本的观察来对总体下结论的，这个过程称为统计推断。20 世纪之前，对于总体与样本的区别是认识不够的。－yIJ 20 世纪初才弄清楚总体与样本之间的联系和区别，同时把统计的问题明确规定下来：

（1）规格问题　即选择什么样的概率分布来描述总体。例如某一个量的测量数据，我们往往选择正态分布来描述。（2）估计问题　即通过样本中算出的一些量，称为统计量.来估计总体分布中的参数。

（3）分布问题　即统计量的分布。

因此在 20 年代,统计学的基本任务就是由样本做出关于总体结论的统计推断。为了得出正确推断，这个过程要有一个标准程式：

（1）总体是一个客观存在的总体（例如本书所有句子构成的总体）。

（2）样本是一个随机样本，而不是一个有代表性的样本，例如 10000 人的人群中不能只选男性或女性。样本的选取要机会均等。为了得到随机样本，统计学家已编制了随机数表。

（3）根据样本的统计值，对总体的参数进行估计，然后计算不确定性的区间，大致相当于误差范围。

例如，总体是本书中所有句子中的字数，要估计参数是所有句子的平均字数。随机抽出 10 句，平均得出每句平均字数 15 个字，由于样本数太少，很少可能总体的平均字数也是 15 个字,我们可以根据统计理论计算出不确定性区间是 15±5 个字，也就是 10 个字到 20 个字的可能性较大。

（4）统计检验，也就是判断样本信息支持总体参数的理想值到什么程度。

125.两头小中间大

——正态分布

在测量及实验过程中，我们积累了大量的数据。这些数据的分布，往往呈现两头小、中间大的正态分布的现象。例如某地区，某个年龄段的男人或女人，他们的身高或体重的分布就近似的呈正态分布。例如成年男子超过 2 米的极为罕见，超过 1.90 米的也不多；另一方面不到 1.50 米和 1.60 米的也不多，大多数人都在 1.65 米到 1.75 米之间。为什么身高服从正态分布呢？原因是影响身高的因素有许多偶然因素，而这些因素相互之间是独立的，再有每一种因素的实际影响都是比较微小的。在这些因素的共同的、随机的作用下形成正态分布。以身高为例，影响它的因素很多，例如父母的身高、个人的营养状况、体育运动的爱好与否、得过什么疾病等等。这些因素中，具备全部有利条件的很少，具备全部不利的条件的也不多，而且大多数人既有长高的条件，也有不利于长高的条件，在这些因素共同作用之下，不高不矮的自然占了多数。它的分布曲线画起来就是一个倒钟形的正态曲线，正态曲线方程

$$Y=\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(x-\mu)^2}{2\delta^2}}$$

如果随机变量 X 的概率密度 Y，可由这个方程表示，就称 X 服从正态分布，正态分布曲线虽然都是两头小、中间大，可是分布与分布不同，它门的不同就反映在方程中的两个参数 μ 及 δ 之上。

参数 μ 称为均值，它表示数据的平均值，参数 δ 称为标准差，δ 称为方差，它是决定曲线形状的关键，δ 小时，曲线陡峭，表示数据分布比较集中，而当 δ 大时，曲线平缓，表示数据分布比较分散。

在实用上，我们往往用 $\mu=0$，$\delta^2=1$ 的正态分布，称为标准正态分布。统计学家已将标准正态分布下各部分面积列成表，而面积的实际意义就是数据落在 $a<x<b$ 之间的概率。例如某年龄段儿童的智商服从 $\mu=100$，$\delta=13$ 的正态分布，从查表立刻可以算出，智商从 87 到 113 的儿童的比例约为 68%，由此可见，正态分布是统计应用中最基本的工具。

126.多次试验出规律

——大数定律与中心极限定理

概率论及数理统计学主要研究随机现象，但是研究目的要掌握其中的规律性。大数定律和中心极限定理就是反映这些规律的。

大数定律主要讨论随机变量序列的算术平均值，当 n 增加时趋于一个常数的规律。最简单的形式实际上就是频率稳定性。大数定律的例子司空见惯。例如投掷硬币，投掷一次、两次甚至十次八次都可能出现正面、反面相差甚多的局面。但是，当次数很多很多时，出现正面的频率一般就越来越接近$\frac{1}{2}$，有人觉得这只不过是想当然，实际上，历史上许多人的确做过这个实验。法国科学家布封让一个小孩向空中掷钱币 4040 次，掷出正面 2048 次，反面 1992 次，掷出正面频率为 0507。法摩根的学生试验，掷出正面 2048 次，反面 2044 次，更加接近 0．501。耶方斯同时掷十枚硬币 2048 次，一共 20480 次中，正面出现 10353 次。英国统计学家卡尔·皮尔逊掷 12000 次，掷出正面 6019 次，正面频率为 0.5016，又一次他掷 24000 次，掷出正面 12012 次，出现正面概率为 0.5005，所有这些频率与概率 0.5 的偏差都不大。大数定律对于求近似值的算法有重要意义。最简单的是布封发明的投针方法计算 7t。布封在纸上划上两条间隔为 1 的一系列平行线，然后在其上随意投长度为 1 的针，则可以证明针压上某条平行线的概率为又因此可以靠投针方法近似计算 π 值。乌尔夫掷下针 5000 次，得出 7t 值为 3.1596，后来有人做 3408 次.得出 7t 值为 3.1415929，准确到小数点后 6 位。

中心极限定理则更进一步讨论频率与概率的偏差。它指出当把这个偏差标准化之后，这个偏差的分布趋向于正态分布。实际上，在自然界和生产中，许多现象受到许多相互独立的随机因素的影响，如果每个因素产生的因素都不大对，总的影响可以看成服从正态分布。中心极限定理是给出这定理成立的充分必要条件，并进一步加以椎广，它在数理统计的大样本理论中有重要的应用。

127.由小知大

——大样本与小样本

数理统计研究如何通过有效地收集、整理和分析带有随机性的数据，对于所考察的问题做出推断、预测乃至决策的方法。它的理论基础是概率论。经典的统计立足于全面观测或普查，不存在随机性的问题，因而不属于数理统计的范围。数理统计则立足于由少量观测推断总体性质，大致分为三大部分：一是抽象方法及实验设计，目的是有效取样。二是统计推断，三是各种各样的应用。因此由样本的性质去推断总体的性质是数理统计最主要的任务，实际上这是由一部分信息推断全体的信息，一般来说是片面的，有偏差的。但是，如果统计量选得好，统计方法合适，就可以用较少的样本得出比较精确的推断。从样本构造统计量的过程，实际上是对样本中所包含的总体信息进行加工和提炼，以求尽可能多地抽出总体的信息。一般来讲，对于容量为 n 的样本 $§_1$，$§_2$……$§_n$，可以选定一个不含未知参数的函数

$$g=g(§_1,\cdots\cdots §_n)$$

它也是一个随机变量。如果对任一个自然数 n，都能导出抽象分布的明显表达式，这种分布称为“精确分布”，它对样本容量较小的统计有用。在数理统计中，具有精确分布的问题称为“小样本问题”，最早的小样本统计量是 1908 年戈塞特提出的，他提出 t—统计量，服从 t—分布。

在此之前，主要应用大样本统计量，如果样本容量 n 无限增大时，统计量 $g(§_1,\cdots\cdots,§_n)$ 的极限分布存在，就称它为近分布，以它作为在样本容量 n 较大时抽象分布的近似分布。这样的问题称为“大样本问题”。

128.身材是怎样遗传的

——回归与相关

英国科学家高尔顿是优生学的创立者，他主要研究体质和智力的遗传。从一个家族一个家族来看，的确有“龙生龙、凤生凤”的倾向，但是从宏观来看，总体的分布与平均值经过许多世代，改变不大，这样他得出统计学上至关重要的“回归”概念。他说，身材高的父母有高个子的孩子，身材矮的父母有矮个子的孩子，但是世世代代人口总体的身高分布并没有明显的变化。他这样解释，一定身高的父母所生子女的平均身高，有朝着总体平均身高移动或回归的倾向。高尔顿的普遍回归定律后来得到近代数理统计学奠基人皮尔逊所证实，他收集了上千个家庭成员身高的纪录，发现在父亲身材高的一组中，儿子的平均身高比他们父亲身材低，而父亲身材低的一组中，儿子的平均身高比他们的父亲平均身高高。用高尔顿的话来说，这就是都回归到中等水平。

这个思想用到统计上就是所谓“回归”分析。以父亲的身高为自变量，儿子的身高有一个分布，它的平均值为因变量，这个函数关系是一条直线，称为回归直线，回归到中等水平也就是说这条直线的斜率小于 1。运用回归直线，我们可以估计和预测因变量总体平均值。自变量的数目 k＞l 时，称为多元回归，最常用的回归方程为

$$Y=\beta_0+\beta_1X_1+\cdots\cdots+\beta_KX_K$$

这时称为线性回归。通过变量代换，X_l 也可以选为 logX ’或 X_kX_1 等，因此，其他回归函数也可以表为线性回归。另外，因变量的数目也可以大于 1，这时称为多重回归。

对于两个或多个随机变量之间的线性关系，如果还要进一步测量这种线性关系的强度，就要引进“相关”方法。相关方法也是高尔顿引进的，其后也引入各种相关系数来描述相关性，最简单的是对于两个变量 X，Y 的相关性，对于 N 次观察（X1，y_1），…，（X_N，y_N）则相关系数为

$$r=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{xi-\bar{x}}{S_x}\right)\left(\frac{yi-\bar{y}}{S_y}\right)$$

其中$\bar{x}$为x_1，……X_n的平均值，y为y_1……y_n的平均值，Sx，Sy分别为x，y 的方差

$$S_X^2=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2$$

$$S_y^2=\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\bar{y}\right)^2$$

根据不同应用，可采用不同的相关形式，而且可以推广到多个变元，对它的估计和检验是统计的基础。

129.药真的有效吗

——假设检验

在生活中充满大最虚假广告，有的宣传药多么多么有效，治愈了多少多少人，有效率达 90%甚至更高；有的宣传产品精良，次品率很小很小；但是，这种宣传并不可靠，可是一般人也没有办法澄清事实。然而，统计学却有一套检验方法来判断假设是否真的成立。如果一种新药有疗效的话，什么叫疗效显著。

费希尔的显著性检验的步骤如下：

（1）假设的概率分布中参数的假设值；

（2）如假设值是真的参数值，有一个检验统计量，而且分布完全知道；

（3）来自总体的一个随机样本；

（4）由样本计算出统计量；

（5）当假设为真时，计算检验统计量取像实际观察到的那么大极端值的概率。这个概率称为是显著水平，根据它我们可以进行判断：如果这个概率很小，或者假设不真，或者假设是真的。但是却发生一件小概率事件。

1933 年，小皮尔逊和波兰统计学家耐曼发展了统计假设检验理论，他们的理论更为严谨，而且对统计的哲学思想有所突破。他们实质上把假设检验当作一个决策过程，要对于两个可能的假设做出接受其中一个的决策。一个假设称为原假设，另一个假设称为备样假设，如果经过一定的程序接受或拒绝原假设，我们有四种情况可能发生：

	决策	
实际情况	接受原假设	拒绝原假设
原假设为真	无错误	第一类错误
原假设不真	第二类情况	无错误

在这四种情况下，我们在两种情况下可能决策错误，这种情况在日常生活中常常遇到，一种是原假设为真，我们拒绝它，这种弃真错误称为第一类错误。另一种是原假设不真找们却接受它，这种容假错误，称为第二类错误。很自然我们想要消除这两类错误，但这是不可能的。因此，我们只能控制这两类错误的概率，这样判断显著性的概率由一个变成两个

a＝第一类错误概率＝显著性水平

β＝第二类错误概率

1—β＝检验的功能

我们想要α和β都很小，但它们的方向不同，要使α小，β就变大，而要使β小，，α就趋于变大。常用办法是固定α，再 加样本大小使β变小到可以接受。

这样数理统计向成为在不确定状况下的决策科学迈出了一大步。

130.醉汉的行走

——布朗运动

1827 年，英国植物学家布朗在显微镜下观察到悬浮在液体中的粒子做奇特的无规则运动，好像醉汉行走那样。他在 1828 年发表这种现象后，曾引起争论，长时间没有满意的定量解释。一直到 1905 年爱因斯坦和波兰物理学家斯莫卢霍夫斯基独立给出物理上的解释，也就是布朗运动是分子无规则碰撞的结果。他们还得出粒子在某一轴上投影的位置 X 服从正态分布。X 的平均值为0，均方差$\overline{x}^2 = \delta^2 t$。法国物理学家佩兰在$1908$年发表了他关于布朗运运的实验结果，并给出测定阿伏加德罗常数一个精确测定方法。这就无可辩驳地证明了原子和分子和实在性。1916 年，斯莫卢霍夫斯基把布朗运动归结为一个数学问题，但不是从随机过程的观点出发的。第一个从数学上深刻研究布朗运动的是美国数学家维纳。1923 年他第一次给出随机函数（即后来随机过程）W（t，ω）的严格定义，证明 X＝σW（t，ω）可以是布朗运动的理论模型。这种函数早在 1900 年已被法国数学家巴夏里耶在他的博士论文中考虑过，不过他的论文是研究投机理论的。他在寻求股票市场的涨落过程中，发现 W（t，ω）的许多特征性质。他实际上发展了一种随机过程理论，而且也提到布朗运动。不过，由于当时对概率论的普遍忽视，没有受到注意。维纳从样本路程的观念出发，研究“路程”的集合，引进维纳测度，并证明几乎所有轨道都是连续的。 1933 年英国数学家佩莱、波兰数学家齐格孟德与维纳合作证明布朗运动的所有轨道几乎肯定在每一点都是不可微的。这种连续而不可微函数在过去很长时间被认为是病态，而现在成为概率论中最重要的一类函数。维纳还证明布朗运动是马尔科夫过程。

马尔科夫过程是 1906 年由俄国数学家马尔科夫提出来的。粗浅地讲，它的含义是：一个体系将来的发展只与体系的现在状况有关，而与体系过去的历史无关。

从马尔科夫过程观点研究布朗运动的是保尔·列维。他从 1938 年起，对布朗运动进行系统的研究。他研究随时间增长的样本路程，总假定未来与过去无关这种马尔科夫性质。1944 年日本数学家角谷静天首先对于二维情形把马尔科夫过程与位势理论联系起来，形成战后概率论的一个重要方向。

131.运筹帷幄之内

——运筹学

现代运筹学已经发展成一个庞大的领域，有着多种多样的分支。其总的目标是运用科学方法，特别是数学方法，对于社会、经济、军事等复杂系统的运行、组织及管理等方面的问题加以分析，建立模型，提出解法以求最终得出正确的决策。运筹学与一般的决策科学不同之处在于它的定量化、数学化，而数学问题往往是最优化问题，这特别表现在人力、财力、物力、时间、空间等资源配置方面。在日常生活中经常出现物资运输及仓库存储、设备更新、工程顺序统筹安排、搜索及打捞等等问题，靠经验盲目处理的情况比比皆是，实际上由此造成大量的人力、物力、财力的损失以及资源的浪费。而运筹学正是在大量可能决策中寻找最有效、最科学、最优的办法的重要途径。因为我们面对的已经是比二次大战前远远复杂的系统了。

运筹学的产生是第二次世界大战之中的事。1933 年希特勒在德国掌权，英国就开始进行适当的准备来防御可能发生的空袭。结果在 1937 年末设计出雷达和飓风式战斗机。但是，1938 年 7 月进行的空战演习中，雷达和战斗机临时凑合，不能形成一个有效的空防体系。因此，当时英国东海岸的雷达研究工作的领导人罗维建议进行关于雷达—战斗机系统的运用方面（与纯技术方面相对立）的研究工作。他还创造出“运用的研究”——即运筹学这个词来称呼这种研究工作，这可以说是运筹学正式诞生。这种研究工作直接导致英国防空体系根本上的改进，在 1940 年 8、9 月间经受了决定性的考验。美国参战之后，在 1942 年底，美国也进行类似的研究工作。由于从事战时工作的科学家在战后大大倡导，而使运筹学的理论和应用在战后得到了蓬勃的发展，有关运筹学的数学问题也得到许多新的解法。

132.多快好省搞四化

——线性规划

一般认为，线性规划随着 1947 年美国数学家丹齐格制定单形法而建立，但数学规划问题早在 1781 年法国数学家蒙日在研究筑城术时已涉及开控与充填问题，即如何运输挖出来的土，使重量 X 运程为极小。线性规划所涉及的数学问题是解线性不等式。一般有限线性不等式组的解在 1936 年由莫获金得到。最早应用线性规划解决实际问题是苏联数学家康托洛维奇，他在 1939 年出版《生产组织与计划中的数学方法》，其中提出得出线性规划问题最初解法，但未受到重视。荷兰经济学家库普曼斯在 40 年代再次独立得到类似方法。但一直到丹齐格的结果发表之后，他们的工作才受到重视，而且因此共同荣获 1975 年度诺贝尔经济学奖。线性规划不仅解决一大批实际问题，而且推动了对算法理论的研究。至今几千个变元及约束条件的计算已经可以在计算机上进行，而且线性规划问题占有计算机工作的相当分量。

标准的线性规划问题是

求 $f(x)=c_1x_1+c_2x_2+\cdots\cdots+c_nx_n$

的极小值，而且这些 X_1，……X_n 满足约束条件：$a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1$

$a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2$

……

$a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m$

$x_i\geq 0$，$j=1$，$2\cdots\cdots n$

式中 $m\leq n$，a_{ij}，c_j，b_j 为常数。由于目标函数及约束条件均为线性的，所以称为线性规划。运输问题是一类特殊的线性规划问题，1941 年由希奇科克最早提出特殊的线性规划问题——运输问题，这问题在物资调运工作中至关重要。某种产品如粮食的产地为 a_1，a_2，$\cdots a_n$，它们的运出量分别为 $B_1\cdots B_2$，运往的粮食量分别为 b_1，$\cdots b_N$，如果从 Ai 运到 Bj 运费，如何调运使总的运费最小，这显然是在约束条件

$$\sum a_i = \sum b_j$$

之下，求

$$\sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}X_{ij}$$

的极小值的线性规划问题，1950 年以后，又根据目标函数、约束条件以及变元取值等，发展出非线性规划、整数规划以及几何规划、参数规划等，其后又发展出组合规划、动态规划、多目标规划等新兴领域，它们都有着非常重要的应用。

133．0.618

——优选法

在日常生产、生活以及科学研究和工程设计中，时常面临如何以最少的步骤、花最少的时间，得出最终的结果。例如寻求两组分产物最好的比例，得出最好的配方，可以进行几十次，几百次，也不一定找出最好的结果。一个最简单的例子是，百分之多少的酒精消毒效果最好，通常试验方法往往是随便选几个样品，例如 5%，15%，25%，…95%的酒精，然后看看哪些效果好，再选择几个中间样品做实验。这样的盲目搜索，显然少、慢、差、费，而优选法正是寻求减少步骤而得出同样结果的方法，其中 0.618 法最为人称道。

0.618 法也称为黄金分割法，也就是在[0，1]区间上先选出两点 X_1＝0.618 和 X_2＝1—0.618＝0.382。我们的目的是求一个只有一个极大值（或极小值）的函数 f（x）在哪一点达到极大值（或极小值），如 f（X1）＜ f（X2），就去掉[x_1,1]，这时再找[0，X_1]上相当于 0.618 或 0.382 的点，由于 0.618 的点就是 X_2，只要找相当于 0.382 的点 X_3，如此做下去，直到达到所需要的精度为止；反之要是 f（x_1）＞f（x_i）就去掉[0,x_2]，找到[x_2,1]中相当于 0.618 的点 X_3，如此做下去，直到达到所需要的精度为止。这样，除了第一次取两个样品进行实验以外，以后每次都只增加一个样品，几次之后就把所求函数的极大值（或极小值）局限在一个很窄的范围之内了。

这个方法之所以有这样的优越性与 0.618 这个数值有关。0.618 实际上就是著名的黄金分割点ω，也就是

$$\frac{\omega}{1}=\frac{1-\omega}{\omega}$$

由此不难算出

$$\omega=\frac{-1+\sqrt{5}}{2}=0.6180339887\Lambda$$

由于ω是无理数，只取 0.618 就足够好了。

当函数 f（x）定义区间不是[0，1]而是[a，b]，也不难找到 X_1 和 X_2 点，它们是

$$x_1=a+\omega\ (b-a)$$

$$x_2=a+(1-\omega)+(b-a)$$

其中ω就是 0.618，而 1—ω是 0.382。

70 年代，由于我国著名数学家华罗庚在全国推广优选法，这种方法为广大群众所掌握，产生了巨大的经济效益。

134.囚徒的两难

——对策论

我们生活在一个竞争的社会当中，在每一种竞争当中都面临着不只一种选择，由于每个参与竞争中不同的选择，产生出各种不同的结果。当然，每个人都想取得最有利的地位，但是，这种情况是否存在是个问题。而在有的情况之下，的确是竞争者处于进退两难的处境。

著名的囚徒疑难就是面对这样一种局面。两个同谋犯共同犯下一桩罪行，两个人被抓起来，分开关着，彼此没有办法通消息。检查人员对于每个囚犯都提出下面的处理办法，还告诉他们，处理两人的办法相同，而且两人都知道这事。“你们两人我们都掌握许多具体证据，如果你们两人都声称自己无罪的话，我们总能定你们的罪，每人都处两年徒刑。如果你承认有罪并且提供证据给你的同伙定罪，那我就把你开释，而判你的同伙五年徒刑。”可是这还不是所有局势的全部，要是两人都认罪呢？他还交待的很清楚，那就得每人都判四年徒刑。

在这种情况下，囚徒总是处于两难境地，为了看清形势，对策论总是用所谓“支付矩阵”来表示，其中（—x，—y）表示甲被判刑 x 年，而乙被判 y 年。

		乙	
		不认罪	供认
甲	不认罪	（-2，-2）	（-5，0）
	供认	（0，-5）	（-4，-4）

由于甲，乙双方情况是对称的，所以我们只考虑一方，每一方都有两种选择，不认罪和供认。每一方也都有两个出发点；只考虑利己和考虑合作，即大家都有好处。显然，只考虑自己时，他最好的选择就是供认，而且期望对方不认罪，如果对方也这样考虑，两人就得每人坐 4 年牢。如果双方都考虑整体利益，都不认罪，即两人加起来坐牢数最少，那就各坐二年，这当然是整体最优解。但是个人来说，也是风险最大的，因为一旦对方背叛，就得蹲五年。

这类对策比经典的对策更复杂。经典的你死我活的（0，1）对策已有完整的理论。但是，由上面的囚徒疑难指出合作往往比斗争对双方更有利。近 20 多年，合作对策成为研究的热门，并且利用计算机进行模拟。例如上面的囚徒疑难如果还有更长期的竞争问题，那第一次是采取合作还是背叛的态度，而且当你受骗之后，采取宽容还是报复的态度，引导出许多战略，但这些战略的相对优越性仍有不同的意见。

135.什么时候大修

——可靠性理论

现在越来越多的复杂系统进入家庭，例如电视机和汽车。说它们复杂，是因为它们的元件、零件、部件成千上万，各有着不可替代的功能。实际上第二次世界大战以来，复杂的军事装备、飞机、大型计算机已经是复杂系统了。

复杂系统的主要问题就是失效，也就是坏了不能用，这种失效也可能是整体损坏，通常是一些元件报废，不管损坏能否修理，它都给使用者带来一系列的问题。

以一台电视机为例，电视机在无故障情况下工作的时间通常称为寿命。可是具体到某一台电视机，它的寿命只能到开始坏才知道。而且我们也无法知道它什么时候才坏，而且导致它损坏的诸多元件中，哪些最容易坏。所有这些都是随机变量。因此，当我们买一台电视机时，我们总是面临一些关键问题，其中最重要的就是它的可靠性问题，以及维修及更换问题。例如是不是我们能到一定时间就像机器一样拿去大修一次，什么时候更换什么零件可以延长它的寿命，还是像国外一样，看到一定时间一扔了事。这些问题就是可靠性理论研究对象。

可靠性理论的核心概念是可靠性，大部分系统可用寿命来表征，系统在失效前正常工作的时间称为寿命，它是随机变量 x 的分布函数

$$F(t) = p\{x \le t\}$$

其中 P 表示在 t 时刻之前损坏的概率，这样 $R(t) = 1 - F(t)$

称为系统可靠度，它可用元件无故障率表示出来。

对于每个元件的分布特征及系统的结构函数，我们可以对可修复系统的失效特征进行研究，其中关键是对于系统所有可能出现的故障进行系统分析，其中主要方法是故障树分析。

在可靠性理论中，还有一系列最优化问题，这主要是更换策略问题以及考虑到费用、时间、备件限制的条件下增大整机的可靠性问题。由于可靠性理论的实用性，它的数学理论正得到蓬勃的发展。

数学基础与数理逻辑

136.桌子、椅子、啤酒杯

——公理系统

第一个几何学基础是由欧几里得的《几何原本》奠定的，但欧几里得体系有着许多缺陷，遭到各方面的批评。例如有的公理（如所有直角都相等）是多余的，有许多概念诉诸直观（如重合）等等，但其主要的缺陷是：

（1）欧几里得的定义许多是无意义的循环定义，如定义点是没有部分，定义线“有长无宽”等。

（2）许多术语没有明确意义，实际上是承认直观的考虑，如一点在“两点之间”等。

（3）逻辑结构的缺陷，在证明的过程中有许多默认的假定，例如在证明命题 I_{16}，不自觉假定直线的无限性。

希尔伯特在一些人的影响下，他提出来“在一切几何命题中，我们可以用桌子、椅子、啤酒杯来代替点、线、面”，这种思想后来导致他完成了几何学基础的彻底革新。

1899 年希尔伯特发表的《几何学基础》不仅彻底清除欧几里得几何学体系中的缺陷，建立了新的几何学基础，而且树立了现代数学的公理化模式，发展了公理学，推动了整个数学基础的研究。其重点在于：

（1）提出一些原始的术语，这些原始术语并不作定义。

（2）原始术语的性质只由公理所反映出来的性质决定。

（3）公理系统中的每一公理是否符合我们的直观不予考虑，我们只考虑公理系统中的公理是否彼此之间没有矛盾，也就是相容性。

希尔伯特在第一版中提出把点、线、面在上，在……之间和全等作为原始概念，并举出 21 条公理，其后作了改动及调整。在他生前最后一版（1930）中，提出五组共 20 条公理的表：

第一组 结合公理（共 8 条公理）

第二组 顺序公理（共 4 条公理）

第三组 合同公理（共 5 条公理）

第四组 连续公理（共 2 条公理）

第五组 平行公理（1 条）

希尔伯特的公理系统非常适合开展一个几何理论的。他不仅为欧氏几何学奠定新的公理化基础，更重要的是建立一套模式来处理任何数学对象，并把它们建立在可靠的公理基础上。他明确提出对公理系统的要求：

（1）无矛盾性，也称为协调性、一致性、相容性，也就是由公理系统不能推出相互矛盾的结论。

（2）独立性，也就是公理系统中没有一个公理可以由其他公理推出。

另外人们还希望公理系统有完全性及范畴性两个比较好的性质，前者是指凡是有关原始术语的命题或它的否定均可由公理推出，后者是指凡是适合公理系统的模型均同构，也就是满足公理系统的对象基本上是唯一的。这个概念是维布仑在 1904 年首先提出的。由于一般的公理系统不一定有完全性，更难得有范畴性，一般对公理系统不能要求太高。但无论如何，无矛盾性是

公理系统的必要条件，因为不满足这条件的公理系统就根本毫无价值。

137.数学中永恒的问题

——无穷

从一开始，无穷只不过是一个简单的否定，也就是不是有穷多。因此，从古希腊到现代数学当中，并没有无穷的位置。但是，随着数学的发展，不可避免要涉及无穷，但多数数学家仍然反对把无穷作为数学的合法对象，一直到康托尔建立了无穷集合论，无穷才最终成为数学中的合法一员。

无穷至少是通过四条道路进到数学当中的。第一条路是“数”的引进。在人对数的概念认识过程中，应该说经历四次飞跃。

（1）区别 1 与多。

（2）区别少数与大数。

（3）区别有穷数与无穷数。

（4）区别无穷数的不同层次。

每一次飞跃都代表对数、对无穷的新认识。

第二条道路是“量”的引进。量是随着测量而进入数学的。数学家为那些不能由有限多个整数加减乘除来表示的量以三种无穷的表示：无穷级数（也包括无穷不循环小数）、无穷乘积和无穷连分数。这一次是无穷多次运算或无穷多次操作。

第三条路是 0 的引进，在数学家眼里，

$$\frac{1}{0}=\infty,\ \frac{1}{\infty}=0$$

从这意义上说，0 和∞是够相像的。在宗教神秘主义的图案中，一条蛇咬着自己的尾巴，构成一个圆圈。要是把这个圆圈看成数直线，小的正数和负数由 0 点隔开，大的正数和负数由∞隔开，∞和 0 一样成为数的合法成员了。

第四条通向无穷之路是几何。几何的对象是空间的几何图形，图形的一个特点是连续性，在求图形的面积和体积时，我们遇到很多麻烦。但是，有穷多次是算不出圆的精确面积来的，需要无穷多次，又是一个无穷。到了 17 世纪，数学家把无穷小量引进数学，构成所谓“无穷小演算”，这就是微积分的最早名称。所谓积分法无非是无穷多个无穷小量加在一起，而微分法则是两个无穷小量相除。由于无穷小运算的引进，无穷正式地进入数学，虽然它给数学带来前所未有的繁荣和进步，它的基础及其合法性仍然受到许多数学家的质疑。高斯说“我必须最最强烈地反对你把无穷作为一完成的东西来使用，因为这在数学中是从来不允许的。无穷只不过是一种谈话方式，它是指一种极限，某些比值可以任意地逼近它，而另一些则容许没有限制的增加。”这里极限的概念，只不过是一种潜在的无穷过程，这里高斯反对那些哪怕是偶尔用一些无穷的概念，甚至是无穷的记号的人，特别是当他们把它当成是普通数一样来考虑时。

柯西也不承认无穷集合的存在，他认为部分同整体构成一一对应是自相矛盾的事。

尽管外尔斯特拉斯、戴德金等数学家都感到需要一种无穷的数学，然而只有康托尔才是唯一真正创造无穷的数学理论大师。他所面对的是整个数学界的几千年几百年对无穷的疑惧，他只有靠个人的勇气，个人的意志才能征

服它，战胜它！

138.无穷与无穷也不一样

——康托尔集合论

康托尔在无穷问题上取得突破，其关键在于无穷与无穷不一样，正如不同的整数不等一样。在这种情况下，我们才能开展对无穷的研究。靠什么去区别两个无穷集合不同呢？康托尔用的是古老的区别两个集体数目不同的办法，这就是一一对应。杀死多少敌人的士兵，无需要把他们的尸首抬回来报功，而只要把他们的左耳朵（或别的什么）割下来，人与耳朵一一对应，人的数目当然也就是耳朵的数目了。康托尔就凭借这个唯一的武器，只身走入无穷的世界。

1873 年 11 月 29 日康托尔在给戴德金的一封信中，终于把导致集合论产生的问题明确地提了出来；正整数的集合（n）与实数的集合（x）之间能否把它们一一对应起来。实际上，这个问题他已经考虑了很久，特别是考虑连续性的本质时，他总是要碰到这个根本问题，他在信中写道：“ 取所有正整数 n 的集体，表示为（n），然后考虑所有实数 x 的集体，表示为（x）；简单说来，问题就是（n）和（x）是否能够对应起来，使得一个集体中的每一个个体只对应另一个集体中一个且唯一一个个体？乍一看，我们可以说答案是否定的，这种对应不可能，因为（n）由离散的部分构成，而（x）构成一个连续统；但是从这种说法我们什么结果也得不到。虽然我非常倾向于认为（n）和（x）不能有这样一个一一对应，但是我找不出理由，我对这事极为关注，也许这理由非常简单。 ”

1873 年 12 月 7 日，康托尔写信给戴德金，说他已能成功地证明，实数的“ 集体 ”是不可数的，也就是不能够同正整数的“ 集体 ”一一对应起来，这个日期应该看成是集合论的诞生日。康托尔的证明方法是反证法，比较繁复，特别是比起后来他发明的对角线方法大为逊色，但他后来发表的问题有了些改变。原来的问题是：

（实数的集体）=（有理数的集体）+（无理数的集体）

既然康托尔已经证明（实数的集体）不能同（整数的集体）一一对应，换句话说（实数的集体）是不可数的，他还证明（有理数的集体）可以同（整数的集体）一一对应，因此是可数的，由此立刻可以推出（无理数的集体）是不可数的。

现在康托尔把这一套思想方式再稍稍改动一下：

（实数的集体）=（实代数数的集体）+（实超越数的集体）

这一回并不太显然，因为实代数数的集体不仅包含全部有理数，还包含大量无理数。加进来这么多无理数之后，实代数数的集体还能是可数的吗？这就是康托尔所要证明的。

139.集合中的子集

——子集合有多少

一个集合 S 包含若干个个体{x_1，x_2，x_3，…}，它们之间的关系“属于”是集合的基本关系，用 xi∈S 表示 xi 属于 S。另一方面，S 中一部分元素构成 s 的一个子集合，例如{x_1，x_2，x_3}是 S 的一个子集合，集合和集合之间的包含关系用 $\subseteq$ 表示，如$\{x_1，x_2，x_3\}\subseteq S$。

这里我们必须注意区别开∈和 $\subseteq$。∈表示元素和集合的关系，如个人与集体、人与人类的关系，而 $\subseteq$ 表示两个集合之 间的关系，如班级与学校，亚洲人与人类的关系。可是元素不一定限于“个体”，也可以是集合。在数学上，可以把直线看成是点的集合。我们在研究平面上某些直线的集合，当然也可以把它们看成点集合的集合。反过来，集合也有由一个元素构成的，例如 {3} ，但是在研究具体集合时，一定要区别开元素与集合，3 与集合{3}。

例如 {0，1} 是具有两个元素 0 与 1 的集合，而 { { 0，1} } 是只有一个元素{0，1}的集合；N 是具有无穷多个元素的集合，而{N}是只有一个元素 N 的集合。

$\{2\}\subseteq\{\{2\}，3\}$，但 $\{2\}\subseteq\{\{2\}，3\}$，$\{2，3\}\subseteq\{1，2，3\}$，但 $\{2，3\}\subseteq\{1，2，3\}$。

这样，我们不仅能定义元素的集合，还可以定义集合的集合，集合的集合的集合，等等。对于任何一个集合 A，我们能定义它的所有子集合所构成的集合，这个子集合的集合称为 A

的幂集合，用 P（A）表示，例如

（1）φ只有一个子集合，即φ本身，所以 P（φ）=（φ）；

（2）一个元素的集合有两个子集，即φ和 {a} ，所以 P（ {a} ）= {φ，{a} } 是个两元素的集合；

（3）二个元素的集合有四个子集，φ，{a} ，{b} ，{a，b} ，所以 P（ {a，b} ）={φ，{a} ，{b} ，{a，b} } 是四个元素的集合；

（4）三个元素的集合 {a， b，c}有八个子集；

从这些例子可以看出，一个有 n 个元素的集合，共有 2^n 个子集合。

显然集合的元素数目永远小于子集合的数目。康托尔的主要工作就是把这个定理推广到无穷集合，他现在有两种无穷集合，一个是整数集合，一个是实数集合。整数集合的元素数目是最小的“无穷数”，用 ω_0 表示。他证明实数集合的元素数目c正好是 2^{ω_0}。而且他证明对所有无穷数ω

$$2^{\omega}>\omega$$

当然 c 也有它的幂集合，基数是 2^c，如此下去，我们可以有越来越大的无穷集合，也有越来越大的“无穷数”，在集合论里称为基数或势，而且在集合论中基数的不同是区别不同集合的唯一标准。

140.连续性的矛盾

——芝诺悖论

公元前 5 世纪希腊哲学家芝诺提出 4 个著名的悖论：第一个悖论是说运动不存在，理由是运动物体到达目的地之前必须到达半路，而到达半路之前又必须到达半路的半路……如此下去，它必须通过无限多个点，这在有限长时间之内是无法办到的。第二个是跑得很快的阿希里赶不上在他前面的乌龟。阿希里是希腊跑得最快的人，为了确定起见，假定乌龟的速度是阿希里的$\frac{1}{100}$，阿希里由 A点出发向 B点跑去，乌龟由B点出发向远离A的方向爬。假设 AB 两点各相距 100 米，从经验来看阿希里在很短时间内追上乌龟，但是从芝诺的分析，阿希里永远追不上乌龟并且永远到不了 B 点。这是因为当阿希里到达AB中间C_1时，乌龟已离开B点，走了$\frac{1}{200}$AB的距离。下一步，阿希里到达C_1B的中点C_2，接着C_2B的中点C_3，……同时乌龟也爬了相应的距离的$\frac{1}{100}$，芝诺悖论的关键是阿希里要到达 B，必须经过无穷点列 C_1，C_2，C_3，……，但是有穷的时间不能在无穷多个点上出现，因此，阿希里永远到不了 B 点，从而永远也追不上乌龟。这两个悖论是仅对空间、时间无限可分的观点的。而第三、第四悖论是仅对空间、时间由不可分的间隔组成。第三个悖论是说“飞矢不动，”因为在某一时间间隔，飞矢总是在某个空间间隔中确定的位置上，因而是静止的。第四个悖论是游行队伍悖论，内容大体相似。这说明希腊人已经看到“无穷小”与“很小很小”的矛盾，当然他们无法解决这些矛盾。

近代对芝诺悖论的处理，分成两步，第一步是极限观念的引进，空间及时间的模型是实数直线，实数直线无限可分，并不妨碍在很短的时间通过很短的路程而保持速度一定。无穷多时间片断之和可以存在有限的极限。第二步是认识到对实数域承认阿基米德公理才能解决芝诺悖论。也就是对任何实数 a，b＞0 存在自然数 n 使 a_n＞6，但近代数学的确存在非阿有序域，对于它，芝诺的说法的确成立。

141.实数线上有多少点

——连续统假设

康托尔建立集合论之后，首先碰到两个无穷基数，一个是可数基数ω_0，一个是实数集的基数或连续统的基数 c，而且

$$c=2^{\omega_0}$$

另一方面，从理论上讲，我们可以把所有的无穷基数按大小顺序排个队，最小的当然是ω_0，接着我们依次编号，是

ω_1，ω_2，ω_3，ω_4，……康托尔现在只有一种方法产生更大的基数，也就是造幂集，而且他经过多年研究也没有发现在可数集和连续统之间还存在有不同基数的集合。于是他早在 1878 年就提出连续统假设，

$$\omega_1=C=2^{\omega_0}$$

换句话说，C是最小的不可数基数，或者说在ω_0与C之间不存在其他的基数。康托尔花了毕生精力企图证明它，但是未能成功。希尔伯特把它列入自己著名的 23 个问题的头一个，希尔伯特本人也曾经用了许多精力证明它，并且在 1925—1926 年宣布过证明的大纲，但终究未能成功，这个问题始终悬而未决。

后来数学家更进一步，提出广义连续统假设（GCH）即

$$2^{\aleph_n}=\omega_n+1$$

换句话说，除了幂集造法产生的集之外，不再存在其他的无穷集合。

1938 年，哥德尔证明 ZF（策梅罗的公理集合化）和 GCH 是协调一致的，不过当然要假设 ZF 本身也是协调的，虽然这一点一直没有得到证明。

1963 年 7 月，美国年轻数学家科恩发明力迫法证明连续统假设的否定命题即 CH 也同 ZF 无矛盾，这样一来 CH 和 ZF 中既不能证明也不能否定。他得出结论 ZF 的公理系统还不够强，不够全，不足以证明或否定 CH，甚至加上选择公理仍然不行。

142.能够选出代表吗

——选择公理

一个群体，它的任何人群，只要不是一个人都没有，总可以选出一个代表。这个看起来天经地义之事，对于无穷集合却引起极大争议，这就是选择公理，它来源于康托尔的良序性假说。所谓良序集合，就是一种特殊的全序集合，其中每个（非空的）子集合都有一个起始元素或极小元素，也就是开头的元素。而康托尔则认为每个集合都可以良序化，这就是良序性假设。而1904 年策梅罗就是通过选择公理来证明良序性假设的。

策梅罗的证明思路非常清楚，他利用一集合 M 的所有子集的良序化来使集合 M 本身的良序化，然后他对证明的原则作了明确的阐述：“ 对 M 的每个子集 M ’，都可设想一个对应元素 m ’，它是 M ’ 中的元素，可称为 M ’ 的‘ 代表元素 ’。 ” 这就是后来所谓“ 选择公理 ” 的第一次明确陈述。

选择公理是集合论也是数学中最重要的公理，在数学中也多次被隐含地用到。皮亚诺在 1890 年就隐含地提到它，它比良序性定理要明确得多，这也说明它为什么在数学中到处碰到。

到了 1904 年，策梅罗明确地用选择公理证明良序性假设之后，反对之声来自两方面：有的人根本就反对无穷，更反对把选择公理推广到无穷，也有人对策梅罗的证明有疑问，保莱尔就持有这个观点：他认为由两个问题（A）任意集合 M 的良序化，(B) 从 M 的每个非空子集合中选出一个特殊因素，(A) 和（B）的等价性并不能得出（A）的一般解决，因为如（B）解决，还没有具体的办法使 M 良序化。例如当（A）是实数集合。1905 年以保莱尔、拜尔、勒贝格为一方，以阿达马为另一方进行热烈的讨论。同时，彭加勒也发表文章反对集合论和选择公理。1906 年罗素把选择公理表示为“ 乘法公理 ”，它是讲“ 对于由给定的、互不相交的非空子集所构成的一类，则至少存在一类，此类和每一类恰好有一个公共元素 ”。

143.理发师的两难处境

——罗素悖论

罗素是 20 世纪最伟大的哲学家之一,悖论的发现与悖论的消除是他数学工作的核心。1901 年 5 月，他发现了著名的罗素悖论，并于 1902 年 8 月写信告诉弗雷格：

集合分成两类：一类是集合不是它本身的元素，一类是集合是它本身元素，考虑所有前一种的集合类 A，那么 A 属于哪一类呢？如果 A 是后一种集合，根据定义，A 的元素不该属于 A，即 A 是前一种集合；反过来，如果 A 是前一种集合，根据定义，所有不是它自身元素的集合应该属于 A，即它属于后一种，所以 A 是前一种集合当且仅当 A 是后一种集合，这显然是一个矛盾。

后来他用一个生动的“理发师悖论”来形象地说明自己的悖论。一个乡村理发师，自夸无人可比，他宣称自己当然不给自己刮脸的人刮脸，但却给所有自己不刮脸的人刮脸。有一天他发生了疑问,他是不是应该给自己刮脸？要是他自己给自己刮脸，那么按照他的声明的前一半，他就不应该给自己刮脸，但是要是他自己不给自己刮脸的话，则照他自夸的那样，又必须给自己刮脸，于是这个理发师陷入了逻辑矛盾之中。

在这之前实际上已经发现一些集合论的“悖论”，如大序数悖论及大基数悖论。这些集合论悖论都涉及到一些共同的东西：“所有”或“全体”，

（1）所有集合的集合。

（2）所有基数的集合。

（3）所有序数的集合。

（4）所有不是自身元素的集合。

看到了它们之间的共性，康托尔认为，悖论容易解决，只要你不说“所有”。他 1899 年在给戴德金的一封信中说：“如果把一个总体中所有组分看成一个整体会导致矛盾，就不应该把这一总体看成一个完成了的对象，我称这种总体为绝对无穷或不相容的总体。”这样，他区别开相容集合与不相容集合而把集合论限制在相容集合的范围，这样就把集合论的悖论轻易地排除在集合论之外了。

罗素悖论实质上同理发师悖论意思差不多，但是它涉及的是最基本的集合论概念：元素，属于，集合。

这个矛盾是如此简单明了，用的概念是如此基本，因此产生了极大的震动。弗雷格得到罗素的信后，认为他的“算术的基础动摇了”，戴德金认为他的实数理论也存在问题。庞加莱等人开始否定无穷集合的理论，打算从根本上解决悖论。当然，这样一来，悖论是消除了，但是真正可靠的数学也大部分化为乌有。怎么办？这是摆在 20 世纪初数学家面前的必须解决的问题！

144.一连串的悖论

——集合论悖论与语义悖论

罗素的悖论发表之后，接着又发现一系列悖论：

（1）理夏尔悖论：法国第戎中学教师理夏尔在 1905 年发表了一个悖论，大意如下：法语中某些片语表示实数，比如“一个圆的圆周与直径之比”就表示实数π。法语字母也像英语字母一样，有一定的顺序，所以我们可以把所有片语按照字母顺序排列，然后按照片语中字母的多少排列，少的在前，多的在后。这样我们把能用片语表达的实数排成一个序列，a_1，a_2，a_3…… 这样我们就得到了所有能用有限多字（字母）定义的数了。它们构成了一个可数集合 E，现在我们规定一个规则把这个序列改变一下，造出一个数来：“设 E 中第 n 个数的第 n 位为 P，我们造一个实数如下：其整数部分为 0，如果 P 不是 8 或 9，第 n 位小数为 p＋1，要是 p 是 8 或 9 的话，第 n 位变成 1。”这个实数显然不属于 E，因为它和 E 中每个数都不一样，要是同 am 相同，应该两数的第 m 位相同，可是根据我们的造法它们不一样，所以同 E 中每个数都不一样。但是它们却可以上面有限多个字组成的话来表示，因此应该属于 E，这就出现矛盾。

（2）培里悖论：培里是英国的图书馆管理员，1906 年有一天他告诉罗素下面的悖论，原来的悖论是用英文写的，下面用中文改写一下：考虑下面表达式“不能够由少于二十二个字而命名的最小自然数”，这表达式用了二十一个字而确定了一个自然数，但按照定义，这个自然数是不能由少于二十二个字确定的。

（3）格瑞林和纳尔逊悖论：纳尔逊是新康德主义的小流派之一弗瑞斯派的代表，1908 年他和他的学生格瑞林把下面的悖论发表在弗瑞斯派的一个文集上，这个悖论常称为格瑞林悖论。

如果一个形容词所表示的性质适用于这个形容词本身，比如“黑的”两字的确是黑的，那么这个形容词称为自适用的。反之，一个形容词如果不具有自适用的性质，就叫作非自适用的。例如汉语形容词“英语的”就是非适用的。现在我们来考虑“非自适用的”这个形容词，它是自适用的还是非自适用的呢？如果“非自适用的”是非自适用的，那么它就是自适用的；如果“非自适用的”是自适用的，那么按照这词的意思，则它是非自适用的，这就导出矛盾。

145.为消除悖论而战

——公理集合论

德国数学家策梅罗采取希尔伯特的公理化方法回避悖论，他把集合论变成一个完全抽象的公理化理论。在这样一个公理化理论中，集合这个概念一直不加定义，而它的性质就由公理反映出来。他不说什么是集合，而只讲从数学上怎样来处理它们，他引进七条公理：

（1）决定性公理（外延公理）：假如一个集合 M 的元素同时是一个集合 N 的元素，反过来，集合 N 的每个元素也是 M 的元素，则 M＝N。简单来讲，每一个集合由它的元素所决定。

（2）初等集合公理（空集合公理、单元素集合公理、对集合公理）；

①空集合公理：存在一个集合“空集”，完全不包括任何元素，用φ表示。

②单元素集合公理：如果 a 是任何一个对象，那么就存在一个集合（a），它的元素包含 a 且只包含 a。

③对集合公理：假如 a 和 b 是两个对象，且存在一个集合（a，b），它只包含 a 和 b 为它的元素，而不包含不同于 a 也不同于 b 的任何其他对象。

（3）分离公理组：假如谓词 E（x）对于一个集合 M 的所有元素都有定义，则 M 有一个子集，M 包含且仅仅包含 M 中那些使 E（x）为真的元素。

（4）幂集合公理：每一个集合 T 都对应另外一个集合 P（T）（T 的幂集），它的元素包含且仅仅包含 T 的所有子集合。

（5）并集合公理：对于集合 S 和 T 都对应一个 S、T 的并集即集合 SUT，它包含且仅仅包含作为 S 的元素和 T 的元素为元素。

（6）选择公理。

（7）无穷公理。

实际上策梅罗的公理系统 Z（公理 1 至 7）把集合限制得使之不要太大，从而回避了比如说所有“对象”，所有序数等等，从而消除罗素悖论产生的条件。

策梅罗不把集合只简单看成一些集团或集体，它是满足七条公理的条件的对象，这样排除了某些不适当的“集合”，特别是产生悖论的原因是定义集合的所谓内涵公理组，如今已换成弱得多的分离公理组。

策梅罗的形式集合论，经过许多人的修改和补充形成了一个适当的体系，这个体系对于开展全部数学分析是够用的也是适当的。另一方面，它也很好地避开悖论。从理论上讲，对于任何公理系统都应该证明公理系统是无矛盾的、独立的、完备的以及范畴的，至少为了不推出矛盾的东西。无矛盾性首先要保证，这点至今还没有能够证明。人们只能满足于至今还没有在其中发现什么矛盾。

146.一场大论战

——逻辑主义、直觉主义与形式主义

由于悖论的出现，产生了数学基础论的危机，其后 30 年间，围绕着各种问题进行热烈的争论，出现相互对立的逻辑主义、直觉主义、形式主义三大学派。

罗素是逻辑主义的代表人物，他们解决悖论的办法是分支类型论。为了避免悖论，他们规定集合自身不能作为它本身的元素。这样他们必须对命题加以区分，不同类型的命题不能等量齐观，从而造成极大的复杂性。为了避免繁琐复杂，他们又引进可化归性公理，即所有命题都可以化归为等价的 0 型命题。但这个公理是完全任意的，遭到许多人反对。特别是罗素等人企图从逻辑中推出全部数学。推导过程极为繁琐，以至花上几百页才能把数 1 定义出来，无怪乎法国大数学家彭加勒挖苦他们说：“ 这是一个可钦佩的定义，它献给从来没有听说过 1 的人。 ”从哲学上来讲，也很难在这个体系中补充直观和经验的概念，从而使数学成为“ 不结果实的 ”、纯粹形式的演绎科学。不过，罗素等人以符号形式实现逻辑数学化，从而极大地推动数理逻辑的发展，这种功绩是不可抹杀的。

直觉主义者走向另一个极端，他们否定实在的无限。克罗内克甚至否定无理数的存在，连圆周率π都认为不存在，因为从整数出发，无法造出π来。对于这种极端的看法，当然支持的人多。到 20 世纪初，这种思想又重新抬头，其代表人物是布劳威尔。他在 1907 年发表《论数学基础》，正式建立起直觉主义数学。他们否定排中律，也就是认为存在着既不能证明也不能否证的命题。他们坚持所有定义和命题都必须通过构造来实现，从而根本不承认无穷集合论。虽然他们因此消除了悖论，但是也因此否定了大部分经典数学。到 30 年代，他们又花了很大力气具体实现他们的纲领，把数学建立在构造的基础上，尽管他们取得相当的成就，但也无法与整个数学的大洋相提并论。

反对直觉主义最有力的是希尔伯特。希尔伯特认为数学的真理所在就是没有矛盾，而不在于能否构造出来。他提出的形式主义的主要论点是：数学本身是形式系统的集合，每个形式系统都包含自己的逻辑、概念、公理及推理规则；数学的任务就是发展出每一个这样的演绎系统，在每一个系统中，定理的证明通过一系列程序得到，只要这种推演过程不产生矛盾出来。为此，希尔伯特以数学的证明为研究对象，提出所谓希尔伯特纲领，成为后来证明论（元数学）的来源。在希尔伯特纲领中，希尔伯特要求无矛盾性的证明通过有限的构造步骤达到，而且力图首先在初等算术的系统中实现。但是 1931 年奥地利数学家哥德尔证明这个要求是达不到的。

147.数学的一揽子计划

——希尔伯特纲领

由于数学基础的危机和直觉主义的猖獗，希尔伯特感到整个数学的生存和发展受到极大的威胁，从 1917 年起 20 多年时间里，他为了挽救古典数学竭尽全力。

1917 年他在苏黎世发表一篇演说，题目是“ 公理思想 ”，这篇文章全面叙述了一些与认识论有关的问题。如数论和集合论的无矛盾性，每个数学问题的原则上可解性，找出数学证明的单纯性的标准，数学中内容与形式表示的关系，数学问题通过有限步骤的可判定性问题。这些问题预示着后来数理逻辑的发展。他认为，要想深入研究，就必须对数学证明的概念进行深入的研究。既然逻辑推理可以符号化，进行数学的研究，为什么证明不行呢？他提出了证明论的一般思想和目标，但是没有具体化。

他第一篇证明论的工作是 1922 年发表的，在《数学的新基础：第一篇》中，他论述如何把数论用有限方法讨论，而数学本身却一般须用超穷方法。他指出用符号逻辑方法可以把命题和证明加以形式化，而把这些形式化的公式及证明直接当作研究对象。1922 年在德国自然科学家和医师协会莱比锡会议上，他做了《数学的逻辑基础》的报告，更进一步提出了证明方法。要求有限主义，经过有限步不推出矛盾 0＝1 来即为证明可靠，这称为希尔伯特计划。

希尔伯特走得更远，他提出这样一种明显理论本身也作为一种数学研究的对象，且应用适当的方法来判定它是否无矛盾，这种做法一般称为元数学或证明论。

他建议两条最基本的原则：

（1）形式主义原则：所有符号完全看作没有意义及内容。

（2）有限主义原则，总能在有限机械步骤之内验证形式理论之内一串公式是否一个证明。

148.行得通与行不通

——哥德尔不完全定理

1930 年左右，整个数学基础问题集中于希尔伯特的公理化计划是否行得通，而这个问题的关键是是否能证明数学公理系统，特别是初等数论的公理系统的无矛盾性能否在公理系统内完成。

哥德尔证明不完全性定理是从考虑希尔伯特计划中数学分析的无矛盾性问题开始的。1930 年秋在哥尼斯堡会议上他宣布了第一不完全性定理：一个包括初等数论的形式系统，如果是无矛盾的，那就是不完全的。或者说，包含初等数论在内的协调的或无矛盾的形式系统 F 中，都存在有不可判定的命题，也就是存在命题 S，S 和 S 的否定在 F 中都不能证明。不久之后他又宣布：如果初等算术系统是无矛盾的，则在算术系统内不可证明。

哥德尔的证明使用了“ 算术化 ”的方法。哥德尔说 :“ 一个系统的公式……从外观上看是原始符号的有穷序列……不难严格地陈述，哪些原始符号的序列是合适公式，哪些不是。类似地，从形式观点来看，证明也只不过是（具有某种确定性质的）一串公式的有穷序列。 ” 因此，研究一个形式系统实际上就是研究可数个对象的集合。我们给每个对象配上一个数，这种把每一个对象配上一个数的方法称为 “ 哥德尔配数法 ” 。哥德尔通过这些数反过来看原来形式系统的性质。

哥德尔的论文在 1931 年发表之后，立即引起逻辑学家的莫大兴趣，它开始虽然使人感到惊异不解，不久即得到广泛承认，并且产生巨大的影响。

（1）哥德尔的证明对希尔伯特原来的计划是一个巨大的打击。把整个数学形式化的打算是注定要失败的。形式主义的原则是不能贯彻到底的。

（2）“ 希尔伯特计划 ” 中证明论的有限主义观点必须修正。从而使证明论的要求稍稍放宽。1936 年甘岑在容许超穷归纳的条件下证明了算术的无矛盾性。

（3）哥德尔的工具递归函数促进了递归函数论的系统研究，同时推动不可判定问题的研究，开始出现递归论的新分支。

这样，哥德尔不完全定理的证明结束了关于数学基础的争论不休的时期，走向了一个新的时代。

149.计算行得通吗

——可计算性

数学开始于数和它的计算。小学就学会加法及乘法，从这两个运算可以算出许多函数，例如阶乘。但是，哪些是可以计算的函数呢？这时就要求分析一下最简单的运算，例如加法和乘法。例如 a＋n，实际上是 a 加 1 得 a＋l 再加 1，一直加 n 个 1。所以加法是一个重复加 1 的过程。而乘法 a·n 无非是把 a 加上 n 次，也是个重复过程。这启发人想到，看起来比较复杂的运算，实际上是不断重复简单的运算得到的。什么是最简单的运算呢？可以说，它们开始于三个函数：

（1）零函数 0（x），对所有 x 值它都等于 0

（2）后继函数 S（x），实际上是加 1 的函数即 S（x）＝x＋1

（3）射影函数 P_i（x_1，x_2，…，x_i，…x_m）=x_i，i=1，……，m 也就是多元函数中只取第 i 个自变量的值。

这三个函数可以说根本不用算，但是由它们通过迭代及原始递归两次操作，组成许许多多复杂的函数，这些函数称为原始递归函数。这里原始递归操作是这样由两个已知原始递归函数 f（X_2，……X_n）和 g（X，y，X_2…X_n）定义新的函数 h（X_1，……X_n）的。

h（0，X_2，…X_n）＝f（X_2，…X_n）

h（X_1＋1，X_2，…X_n）＝g（X_1，h（X_1，…X_n），X_2，…X_n）这种方法非常简单，但却产生了日常计算中几乎所有用到的函数。那么自然要问，有没有能递归计算的函数不是原始递归函数呢？遗憾的是有！因此，数学家还得考虑更一般的递归函数。

一般递归函数只是在上面两项操作的基础上再添上第三个操作——极小化操作，即对于已知的递归函数 f（y，x_1，X_2…x_n），如对每一组变元值（a_1，…a_2），存在 y 使 f（y，a_1，a_2，…a_n）＝0

这时定义的新函数 h（x_1，…x_n）为满足它的最小的 y，即 h（X_1，…X_n）＝μy[f（y，X_1，…X_n）＝0]

因此我们得到一大类的能行的、可计算的函数，也就是对于 f，存在一个确定的、机械的程序，通过对已知数据进行有限次的计算，求出确定的函数值来。

1936 年图灵设计出图灵机，它是现代计算机的理想模型，在图灵机上可计算的函数称为图灵可计算函数。显然图灵可计算函数更具体地实现机械可计算性和能行性。至于图灵可计算函数和一般递归函数的关系则由图灵论点所概括：函数的可计算性与图灵可计算性是等价值的。这样，可计算性问题得到完满的解决。

150.无穷无尽的探索

——数学基础与数理逻辑

19 世纪末到 20 世纪初，数学经历一个翻天覆地的变化，其原因来源于对于数学基础的探究。过去对数学的基本概念，人们是不加追究的，除了哲学家之外，没有人去讨论：什么是数？什么是曲线？什么是无穷？数应该是什么？公理应该如何？空间应该如何？直观是否靠得住？逻辑是不是数学的基础？等等。 19 世纪 70 年代起由于几何及分析的发展，有的数学家开始关心起这类事情来。这些问题提起来容易，解决却很困难，而且都不是作为研究论文发表出来，而是在讲课中，在教科书中，在通信中慢慢传播开来的。这个过程很慢，实际上阻力也很大，许多数学家到死都反对没有切线的曲线，克洛内克连π的存在都反对就是典型的例子。正是由于外尔斯特拉斯、戴德金、康托尔和希尔伯特等人的努力，基础研究在康托尔的晚年才为广大数学家所重视，形成一个基础研究的热潮，几乎所有数学家都参加进来。

数学基础的研究大致分成四大支脉，彼此之间也并非完全独立，而是彼此交叉，形成现代数学的新方向：

（1）数的理论以及空间或其他对象的理论。说到数学分析，就要搞清楚什么是无理数？什么是实数？在大家对实数定义了以后，大家又该追问自然数是从哪儿来的？这个方向领头的是外尔斯特拉斯、戴德金、康托尔及皮亚诺。

（2）集合及无穷或超穷数的理论，连续统的理论，这是康托尔一个人的独创。

（3）公理化理论。1899 年希尔伯特的《几何学基础》的出版是这方向研究的里程碑，从此之后，形成一个公理化热潮，集合论也不例外。

（4）逻辑的数学理论。从 19 世纪中叶起，这是在孤立的情况下进行，可以说与数学主流脱离很远,一直到 19 世纪末才汇入整个数学研究的大河当中，领先的是布尔，其后弗雷格及皮亚诺做出决定性贡献，其后罗素及怀特海又集其大成。

康托尔集合论产生出悖论，悖论引发了 20 世纪初的基础大战，因此数学家、哲学家很快就分成两大阵营，各有当时数学界的领袖人物为代表：尊重传统、反对无穷集合论的一派以庞加莱为代表，维护集合论的一派以希尔伯特为代表，他们各有自己的同盟军，面对面地展开基础大战。

希尔伯待把他自己的观点称为公理主义者，他认为只有公理方法才能解决所有的困难，他的目的是建立算术的公理及其无矛盾性，他接受实在无穷并且称赞康托尔的工作。为此，他提出一个明确的方案—一希尔伯特计划。

1930 年以后，数学逻辑开始成为一个专门学科，并得到了蓬勃发展。哥德尔的两个定理证明之后，希尔伯特有限主义纲领行不通，证明论出现新的情况，主要有两方面：通过放宽有限主义的限制来证明算术无矛盾性以及把证明形式化、标准化。这些主要在 30 年代完成。同时哥德尔引进“ 逆归函数，发展成递归论的新分支，开始研究判定问题。而哥德尔本人转向公理集合论的研究，从此出现公理集合论的黄金时代。50 年代模型论应运而生，它与数学有着密切联系，并逐步产生积极的作用。正是这些研究的发展形成当代数理逻辑的四大分支：模型论、公理集合论、递归论、证明论，基础研究形成

独立的数学体系。