目　　录

第一章　瞎扯伽罗华群论思想

基本信息

1. 原文链接 https://www.douban.com/group/topic/95244972/.
2. 本文作者 wxmang.

1.1　序

先声明一句，这篇帖子为了使非数学专业的人能够阅读下去，对主要概念的来源和主要定理的证明进行了一些简化，可能导致不严谨。所以只能归类于瞎扯范畴。专业的数学工作者不要过于苛责。我为了简明扼要说清楚，不得不在严谨上做妥协，甚至有的地方可能是错的。

这篇帖子目的是介绍数学是如何从研究计算进化到研究结构的。

伽罗华是数学从计算转向结构的关键人物，或者说是数学从古代转向近代的关键人物。在伽罗华之前，数学本质是靠计算来解决问题，伽罗华以超凡的洞察力，构建了从数学结构来研究数学本质问题的框架。这时从具体到抽象的一步巨大跨越。

我想用一个具体例子说明人类是如何从具体事物进化到抽象概念的。

为了非数学系的人能够知道我说的内容，我用了大量描述性语言，所以不够严谨。在通俗和严谨之间，只能做此取舍。

人类第一个真正的抽象学科是抽象代数，抽象代数是从伽罗华群论发展起来。为了理解抽象代数，我们介绍一下伽罗华群论的来历，这样便于以后有兴趣看抽象代数，进入本质更快一点。

由于篇幅和豆瓣对符号的限制，一般抽象代数就无法介绍了，有兴趣的自己去看书。这里只做点科普。

我们已经在以前讨论数学基础的帖子里知道，现代数学主要研究从现实世界中抽象出的空间形式和数量关系，也即结构及结构之间的关系，而结构进入数学只有 100 年的历史，是由群的概念引进而开始的。群的概念的引入就是伽罗华，他也是第一位在有意识地以结构的研究代替计算的人。群论彻底解决了代数方程的根式求解问题此发展了一整套关于群和域的理论。

但是群的概念并不是伽罗华发明的，而是产生于拉格朗日研究代数方程的解过程中：拉格朗日已经意识到一元 n 次方程的根是一个置换群，而且也猜想一般五次以上方程无根式解，但是拉格朗日没能证明这个猜想，后来鲁菲力和阿贝尔都企图证明这个猜想，其中鲁菲力的论文有 560 多页，阿贝尔有几页，不过证明被验证后都是错的或逻辑不完备的。

而置换群的性质，柯西在 1815 年就已经发现了，可是柯西没能把其与一元 n 次方程的解结合起来，错过了这一数学史上最伟大的发现。

伽罗华的重大发现不是发明了群的概念，而是发现每个一元 n 次方程的解都与一个置换群对应，而置换群的群结构决定了解的特性。所以不需要计算解，只需要研究置换群的结构，就能了解解的性质。也即把数学计算改为研究数学结构。

抽象代数是研究数学结构的，代数结构＝集合上按照公理体系定义的运算规则（集合包括包括实数、复数、向量（vector）、矩阵（matrix）、变换（transformation）等集合，运算规则包括加法，乘法等等）。

按照教科书定义，抽象代数是研究各种抽象的公理化代数结构的学科，如群（group）、环（ring）、域（domain）等等。

下面先说说目前抽象代数对其研究的主要代数结构的抽象定义，不过这些定义不是我们的重点，看不看无所谓。只是想表明一下抽象的格式是什么，不是我想讲的。

1.2 基本定义

群的定义是：

定义 1.1: 群的定义

假设一个非空集合，上面有一个二元运算，如果满足以下条件：
(1) 封闭性：若 $a,b \in G$，则存在唯一确定的 $c \in G$，使得 $a \cdot b = c$；
(2) 结合律成立，即对 G 中任意元素都有 $(a \cdot b) \cdot c = a \cdot (b \cdot c)$；
(3) 单位元存在：存在 $e \in G$，对任意 a，满足 $a \cdot e = e \cdot a = a$。- e 称为单位元，也称幺元；
(4) 逆元存在：任意 $a \in G$，存在 b，$a \cdot b = b \cdot a = e$（$e$ 为单位元），则称 a 与 b 互为逆元素，简称逆元。记作 a^{-1}；
则称 G 对 $\cdot$ 构成一个群。

环的定义是：

定义 1.2: 环的定义

R 是一个非空集合，若定义了两种代数运算 $+$ 和 $\cdot$（不一定是我们常识的加与乘，是一种抽象运算规则），且满足：
(*1*)、集合 R 在 $+$ 运算下构成阿贝尔群 *(Abel group*，交换群，也即对任意的 $a,b \in G$, 有 $(a \cdot b) \cdot (a \cdot b) = (a \cdot a) \cdot (b \cdot b)$，并不是所有群都是阿贝尔群，比如矩阵的乘法不满足交换律，所以 n 阶可逆方阵关于乘法组成的群不是交换群)；
(*2*)、关于有结合律，即，
$\cdot R$ 对 $\cdot$ 构成一个半群；
(*3*)、分配律与结合律对成立，即，有:；
称代数系统是一个环 *(Ring)*。

域的定义有两种方式：

定义 1.3: 域的定义

第一种定义是 D 是一个有单位元 $e(\neq 0)$ 的交换环（即对于乘法运算可交换），如果 D 中每个非零元都可逆，称 D 是一个域。（比如有理数域，剩余类域，典型域，有理函数域，半纯函数域等等）。

定义 1.4: 域的定义

第二种定义，设 $< R, +, * >$ 是环，如果 $< R, + >$ 和 $< R-0, * >$ 都是交换群（0 为 $< R, + >$ 的幺元）且满足分配律，则称 $< R, +, * >$ 是域。比如 D 是一个含有非零数的数集，如果 D 对于数的四则运算都封闭，那么称代数系统 $(D, +, -, \times, \div)$ 为一个域。

有理数域 $(Q, +, *)$，实数域 $(R, +, *)$，复数域 $(C, +, *)$，连续函数域 $(R^R, +, \cdot)$ 都是域。但整数集 Z 不是域，因为 $\frac{1}{x}$ 不是整数。（整数集 Z 是一个环，是整环）。

线性代数就是域的一个特例。

抽象代数与数学其它分支相结合产生了代数几何、代数数论、代数拓扑、拓扑群等新的数学学科。抽象代数已经成了当代大部分数学的通用语言。

在抽象代数研究的代数结构中，最简单的是群（Group）。它只有一种符合结合率的可逆运算，通常叫“乘法”。如果这种运算也符合交换率，那么就叫阿贝尔群（Abelian Group）。

群论是伽罗华（Galois）在研究多项式方程根式求解过程中提出的，是抽象代数的起点。

所以想理解抽象代数，就得先理解群论，想理解群论，就得先理解伽罗华理论，想理解伽罗华理论，就得先了解拉格朗日的代数方程工作。

1.3　代数方程的历史

我们在初中就知道的一元一次和一元二次方程的求解方法其实在古巴比伦时代就存在了，但是一元三次方程解的公式直到十六世纪初才由意大利人塔塔

里亚发现。

三次方程被解出来后，一般的四次方程很快就被意大利人的费拉里解出。

先补充介绍一下一元一次方程到一元四次方程的解法，这个与后来的群的思想有关。

一次方程：$ax+b=0$，只要是学过初等代数的都会解：$x=-\frac{b}{a}$。

二次方程：$ax^2+bx+c=0$, 解是:$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$，这个用因式分解很容易。

在公元前巴比伦人已能解这种形式的方程。

三次方程：$ax^3+bx^2+cx+d=0$ 和四次方程 $ax^4+bx^3+cx^2+dx+e=0$ 的解法比解一次，二次的方程难得多了。

对一般三次方程 $ax^3+bx^2+cx+d=0$，先除掉 a，令 $b\div a=a$，$c\div a=b$，$d\div a=c$，原方程变成：

$x^3+ax^2+bx+c=0$，

令 $y=x+\frac{a}{3}$，得：$y^3+py+q=0$。（1）

其中 $p=b-\frac{a^2}{3}$，$q=\frac{2a^3}{27}-\frac{ab}{3}+c$，考虑等式 $(u+v)^3=u^3+v^3+3(u+v)uv$.

即 $(u+v)^3-3(u+v)uv-(u^3+v^3)=0$。（2）

比较（1）和（2），令 $y=u+v$，

则方程（2）变为：$(u+v)^3+p(u+v)+q=0$，其中 $p=-3uv$,$q=-(u^3+v^3)$。

即 $u^3v^3=-\frac{p^3}{27}$，$u^3+v^3=-q$。（3）

则得到 $v^6+qv^3-\frac{p^3}{27}=0$

把 v^3 当成 x, 则是一个二次函数，易解得，$u^3=-\frac{q}{2}+((\frac{q}{2})^2+\frac{p^3}{27})^{\frac{1}{2}}$，$v^3=-\frac{q}{2}-((\frac{q}{2})^2+\frac{p^3}{27})^{\frac{1}{2}}$;

由于 u,v 对称，所以也有 $v^3=-\frac{q}{2}+((\frac{q}{2})^2+\frac{p^3}{27})^{\frac{1}{2}}$，$u^3=-\frac{q}{2}-((\frac{q}{2})^2+\frac{p^3}{27})^{\frac{1}{2}}$ 同时成立;

所以可得到:

$y=(-(\frac{q}{2})+(\frac{p^3}{27}+\frac{q^2}{4})^{\frac{1}{2}})^{\frac{1}{3}}+(-(\frac{q}{2})-(\frac{p^3}{27}+\frac{q^2}{4})^{\frac{1}{2}})^{\frac{1}{3}}$，进而可得到原方程根 x 的值。

同理整理四次方程，对于 $x^4+ax^3+bx^2+cx+d=0$，令 $y=x+\frac{a}{4}$，则原方程可变为：

$y^4+py^2+qy+r=0$。（4）

其中 $p=b-6(\frac{a}{4})^2$，$q=c-(\frac{a}{c})b+(\frac{a}{2})^3$，$r=d-(\frac{a}{4})c+(\frac{a}{4})^2b-3(\frac{a}{4})^4$ （4）

移项，得：$y^4+py^2=-qy-r$。（5）

（5）等式左边配方，得：$(y^2+\frac{p}{2})^2=-qy-r+(\frac{p}{2})^2$

在左端括号内加 u 得：$(y^2+\frac{p}{2}+u)^2=-qy-r+(\frac{p}{2})^2+2uy^2+pu+u^2$。（6）

则右端应为完全平方数，故有：$\Delta=q^2-4\times 2u(\frac{p^2}{4}+pu+u^2-r)=0$。（二次方程可以分解为 $(x-\frac{(-b+(b^2-4ac)^{\frac{1}{2}})}{2a})(x-\frac{(-b-(b^2-4ac)^{\frac{1}{2}}}{2a})$，如果 Δ 不等于零，就无法满足右端完全平方数条件。（$\Delta=b^2-4ac$）。

即：$8u^3+8pu^2+(2p^2-8r)u-q^2=0$。（7）

（7）显然为可解的三次方程，解答该方程就可得到 u 的值。

而且（6）就变为 $(y^2+\frac{p}{2}+u)^2=((2uy)^{\frac{1}{2}}-(\frac{q}{2}(2u)^{\frac{1}{2}}))^2$。

因此有 $y^2+\frac{p}{2}+u=(2uy)^{\frac{1}{2}}-((\frac{q}{2})(2u)^{\frac{1}{2}})$

由于 u 已经解出（按照三次方程解法，有两组，每组三个值），p,q,r 都是已知的方程系数（见（4））所以这个二次方程很容易得到 y 的值，进而得到原方程的根 x 的值。

上面工作都是初等数学，学过初中一年级因式分解，理解毫无问题。

注意，数学家的大招马上就来，一步从初中跨入大学。当然后面内容也是检验一个人抽象思维能力的试金石，看不懂的话，也就没法从事数学工作了。

1.4　拉格朗日工作

在介绍拉格朗日工作前，我们先得介绍韦达定理。

定理 1.1: 韦达定理

设 $x_1, x_2, \cdots, x_n$ 是方程 $x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_{n-1}x + a_n$ 的 n 个根，则：

$$x_1 + x_2 + \cdots + x_n = -a_1$$

$$x_1x_2 + x_1x_3 + \cdots + x_1x_n + x_2x_3 + \cdots + x_{n-1}x_n = a_2$$

$$\cdots$$

$$x_1x_2 \cdots x_n = (-1)^n a_n$$

韦达定理很容易用数学归纳法证明。

下面介绍一下简单置换的记号：

$x_1, x_2, x_3, \cdots x_n$, 如果进行置换，例如 x_1 置换成 x_2,x_2 置换成 x_3,$\cdots$ x_n 置换成 x_1, 记成 $(123\cdots n)$，置换不变，记为 1。显然 n 元素所有的置换是一个 $n!$ 元素的集合。

先介绍拉格朗日的发现，然后介绍其发现过程。

拉格朗日发现：

□ 解一元三次方程需要预解二次辅助方程，解一元四次方程需要预解三次辅助方程。

□ 要解高次方程主要是解它的辅助方程。

□ 辅助方程的次数必须小于原方程的次数，不然原方程一般不可解。（当然解三次方程时的辅助方程是六次的，是因为可按二次方程求，所以本质还是降阶了）

□ 由于辅助方程解的表达式可以任意交换其系数 a, b, c 的位置（因为对称），即 3 次方程的解的表达式有 $3! = 3 \times 2 = 6$ 个。（Lagrange 原话是：方程的解其实不依赖 a, b, c 的值，而是依赖辅助方程结构在原方程根下置换出的不同值的个数）。

至此，解代数方程必有置换的想法已正式形成（也即 n 次方程的 n 个根的排列顺序有 $n!$ 个，或者说这 $n!$ 个排列组合的根，都是方程的解，也即方程根时对称的）。

这是一个很重要的发现：也即方程解必须满足置换条件，这也就是伽罗华从研究求解转为研究代数方程结构的起点，他通过研究根组成的集合（置换群）的性质，证明了：大于五次的方程的根组成的置换群其性质导致其不可通过辅助方程降阶，也即不可以用有理运算和方根求解）。

□ 辅助方程的关键是找到根的表达式——预解式（为原方程根的函数），解方程只需要找到预解式。

所以解代数方程实际是要解辅助方程，因此要寻找一个预解式，此预解式在原方程根的置换下取不同值的个数即为辅助方程的次数，找到了合适的预解式就得到了辅助方程（辅助方程的系数可由原方程的系数表示），解答了辅助方程就可以顺利的得到原方程的根。

因为只要有了预解式，就很容易得到它在原方程根下置换出不同值的个数，那么辅助方程的次数就确定了。

下面介绍拉格朗日的工作过程。

先用二次方程来解释他的思考过程。

考虑二次方程 $x^2+px+q=0$，设 x_1,x_2 是其两个解，构造预解式 $r_1=x_1-x_2$，$r_2=x_2-x_1$，显然 r_1,r_2 在置换 S(2)（包括置换 1 和（12），其实 1 和（12）就是一个 2 阶置换群）下，有 $r_1\to r_1$，$r_2\to r_2$；$r_1\to r_2$，$r_2\to r_1$。

构造以 r_1,r_2 为根的辅助方程（也称预解方程）

$\Phi(X)=(X-r1)(X-r2)$，这个方程显然在 S（2）下不变，

根据韦达定理，$r_1+r_2=-p$,$r_1r_2=q$，能够得到 $\Phi(X)=X^2-(p^2-4q)$

对一般 n 次方程，由于 $\Phi(X)$ 是根组成的置换群，是对称的，可以按照高中代数学过的牛顿多项式定理，得到：$\Phi(X)=$ 原方程系数构成的初等多项式表达，也即 $\Phi(X)$ 可以用原来方程的系数表达出来。

显然 X 有两个解：r_1,r_2,$(p^2-4q)^{\frac{1}{2}}$ 和 $-(p^2-4q)^{\frac{1}{2}}$，（当然 r_1,r_2 具体取值，是一个 2! 的排列组合）

$r_1=x_1-x_2$,$r_2=x_2-x_1$，原方程满足 $x_1+x_2=-p$，那么原方程解得到：$x_1,x_2=\dfrac{-p\pm(p^2-4q)^{\frac{1}{2}}}{2}$，具体 x_1,x_2 取值，也是 2! 个组合。

对一般三次方程 $x^3+px+q=0$(参见前面介绍，任意三次方程总能整理成

这个模式），假设 x_1, x_2, x_3 是其根，引入预解式：

$r_1 = x_1 + wx_2 + w^2x_3$,（其中 $x^3 = 1$ 的三个根表达为 1，w，w^2）

$x^n = 1$ 的解可以这样考虑，令 $x = r(isin\theta + cos\theta)$, 由于 $x^n = 1$，所以 $r = 1$，$n\theta = 2\pi k, k = 1, 2, \cdots, n-1$，方程的 n 个解分别为 $1, w, w^2, w^3 \cdots w^{n-1}$，其中 $w = \mathrm{e}^{\mathrm{i}2\pi/n}, e$ 是欧拉常数。（这是大学微积分常识）

用 $S(3)$ 做置换计算得到（$S(3)$ 包括：$1, (132), (321), (213), (231), (312)$ 等六个置换，这是一个六阶置换群）

$1 = x_1 + wx2 + w^2x3$

$r_2 = wx_1 + w^2x_2 + x_3$

$r_3 = w^2x_1 + x_2 + wx_3$

$r_4 = x_1 + w^2x_2 + wx_3$

$r_5 = wx_1 + x_2 + w^2x_3$

$r_6 = w^2x_1 + wx_2 + x_3$

做这种置换，是要用韦达定理把根与系数的关系建立起来.

定义预解方程 $\Phi(X) = (X - r_1)(X - r_2)\cdots(X - r_6)$

显然 $1 + w + w^2 = 0$，$w^3 = 1$（w 是 $x^3 = 1$ 的三个根之一）。

得到：$\Phi(X) = (X^3 - r_1^3)(X^3 - r_2^3) = 0$

令 $r_1^3 = u$,$r_2^3 = v$,$x^3 = t$, 则转化为一个二次方程，也即形如 $x^3 + px + q = 0$ 的一元三次方程的求根公式的形式应该为 $x = A^{\frac{1}{3}} + B^{\frac{1}{3}}$ 型，即为两个开立方之和。

（1）将 $x = A^{\frac{1}{3}} + B^{\frac{1}{3}}$ 两边同时立方可以得到

（2）$x^3 = (A + B) + 3(AB)^{\frac{1}{3}}(A^{\frac{1}{3}} + B^{\frac{1}{3}})$

（3）由于 $x = A^{\frac{1}{3}} + B^{\frac{1}{3}}$，

所以（2）可化为 $x^3 = (A + B) + 3(AB)^{\frac{1}{3}}x$，

移项可得

（4）$x^3 - 3(AB)^{\frac{1}{3}}x - (A + B) = 0$，

和一元三次方程和特殊型 $x^3 + px + q = 0$ 作比较，可知

(5)$-3(AB)^{\frac{1}{3}} = p, -(A + B) = q$，化简得

（6）$A+B=-q$，$AB=-(\frac{p}{3})^3$

（7）这样其实就将一元三次方程的求根公式化为了一元二次方程的求根公式问题，因为 A 和 B 可以看作是一元二次方程的两个根，而 (6) 则是关于形如 $ay^2+by+c=0$ 的一元二次方程两个根的韦达定理，即

（8）$y_1+y_2=-\frac{b}{a}, y_1y_2=\frac{c}{a}$

(9) 对比（6）和（8），可令 $A=y_1, B=y_2, q=\frac{b}{a}, -(\frac{p}{3})^3=\frac{c}{a}$

(10) 由于型为 $ay^2+by+c=0$ 的一元二次方程求根公式为 $y_1=-\frac{(b+(b^2-4ac)^{\frac{1}{2}}}{2a}; y_2=-\frac{(b-(b^2-4ac)^{\frac{1}{2}}}{2a}$ 可化为

(11)$y_1=-\frac{b}{2a}-((\frac{b}{2a})^2-(\frac{c}{a}))^{\frac{1}{2}}, y_2=-\frac{b}{2a}+((\frac{b}{2a})^2-(\frac{c}{a}))^{\frac{1}{2}}$

将 (9) 中的 $A=y_1, B=y_2, q=b/a, -(\frac{p}{3})^3=c/a$ 代入（11）可得

(12)$A=-(\frac{q}{2})-((\frac{q}{2})^2+(\frac{p}{3})^3)^{\frac{1}{2}}, B=-(\frac{q}{2})-((\frac{q}{2})^2+(\frac{p}{3})^3)^{\frac{1}{2}}$

(13) 将 A，B 代入 $x=A^{1/3}+B^{1/3}$ 得

(14)$x=(-(\frac{q}{2})-((\frac{q}{2})^2+(\frac{p}{3})^3)^{\frac{1}{2}}))^{(\frac{1}{3})}+((-(\frac{q}{2})+((\frac{q}{2})^2+(\frac{p}{3})^3)^{\frac{1}{2}}))^{\frac{1}{3}}$ 式

(14) 只是一元三方程的一个实根解，按韦达定理一元三次方程应该有三个根，不过按韦达定理一元三次方程只要求出了其中一个根，另两个根就容易求出了。

对于一般一元四次方程，$ax^4+bx^3+cx^2+dx+e=0$，拉格朗日照样构造预解方程 $r_1=x_1+ix_2-x_3-ix_4$, 然后经 S（4）置换，出来 24 组，

然后构造 $\Phi(X)=(X-r_1)(X-r_2)\cdots(X-r_24)$

根据上面解三次方程方法，拉格朗日解决了四次方程求根办法（具体写起来太复杂，纯粹是力气活，没创意，省略）。

按照拉格朗日方法，一元四次求根公式这种方法来解一元四次方程，只需求解一个一元三次方程即可。

根据上述结果，拉格朗日认为：解三次方程方法，辅助方程为二次的，解四次方程，辅助方程为 3 次的，那么解 n 次方程，只要找到 n-1 辅助方程，就可以解。

为这个目标，Lagrange 利用 1 的任意 n 次单位根 $(x^n = 1)$, 引进了预解式 $1+x+x^2+x^3+\ldots+x^{n-1}$ 来试图找到 n 次方程解法，但是用这种方法，Lagrange 进行五次及五次以上方程的尝试都失败了。

因为按照他的方法，解一元五次方程需要预解二十四次的辅助方程（Tschirnaus、Bezut、Euler 也得到同样的结果）。由此，他开始怀疑五次以上方程是无根式解的。

1771 年，拉格朗日发表长篇论文《关于方程的代数解法的思考》提出了这个怀疑。（不过德国数学家高斯在 1801 年，他解决了分圆方程 $x^p - 1 = 0$（p 为质数）可用根式求解，这表明并非所有高次方程不能用根式求解。因此，可用根式求解的是所有高次方程还是部分高次方程的问题需进一步查明）。

根据拉格朗日的判断，鲁菲力（Ruffini）1813 年，从反面论证高于五次的方程可能没有一般代数解，不过他的证明不严谨。

1826 年，阿贝尔严格证明：如果一个方程可以根式求解，则出现在根的表达式中的每个根式都可表示成方程的根和某些单位根的有理数。并且利用这个定理又证明出了阿贝尔定理：

一般高于四次的方程不可能代数地求解，这些方程的根不能用方程的系数通过加、减、乘、除、乘方、开方这些代数运算表示出来。但是阿贝尔没有回答每一个具体的方程是否可以用代数方法求解的问题。

阿贝尔还在在高斯分圆方程可解性基础上，证明了：

任意次的一类特殊方程的可解充分必要条件是全部根都是其中一个根（假设为 x）的有理函数，并且任意两个根 $q1(x)$ 与 $q2(x)$ 满足 $q1q2(x) = q2q1(x)$，$q1, q2$ 为有理函数。（现在称这种方程为阿贝尔方程）。

其实这就是群，只是阿贝尔没能意识到，也没有明确地构造方程根的置换集合（因为若方程所有的根都用根 x_1 来表示成有理函数 $q_j(x_1), j = 1, 2, 3, \cdots, n$，当用另一个根 x_i 代替 x_1 时，其中 $i \leq n$，那么 $qj(xi)$ 是以不同顺序排列的原方程的根，$j = 1, 2, \cdots, n$。也即根 $x_i = q_1(x_i), q_2(x_i), \cdots q_n(x_i)$ 是根 $x1, x2, \cdots, xn$ 的一个置换），阿贝尔仅仅考虑了根的可交换性：$q1q2(x) = q2q1(x)$，并证明方程只要满足这种性质，便可简化为低次的辅助方程，辅助方程可依次用根式求解。

所以阿贝尔解决了构造任意次数的代数可解的方程的问题，却没能解决判定已知方程是否可用根式求解的问题。

1.5　伽罗华出场了

伽罗华的思想来自于拉格朗日用置换的思想进行代数方程求解。

（1）、伽罗华从拉格朗日方程根的置换思想入手

为了介绍伽罗华从拉格朗日工作飞跃，我们用一个简单例子来解释。

拉格朗日已经意识到，如果一元 n 次方程能够变成 $[x-x(1)][x-x(2)]\cdots[x-x(n)]$ 这样彻底地分解因式，那么方程的解就得到了。但往往不能，必须扩张系数的数域才行。例如：

$f(x)=x^2+1$

这个多项式在实数范围内是不能分解的，如果允许把虚数单位 i 作为系数的话，这个式子可以分解成：

$f(x)=(x+\mathrm{i})(x-\mathrm{i})$

也即：当域的范围越大，在这个域中进行的因式分解就越彻底，当一个 n 次多项式可以被分解为 n 个一次多项式的乘积时，方程的 n 个解就找出来了。这个域叫做 $f(x)$ 的分裂域。

通过一系列的扩域就能把多项式的系数域扩张到多项式的分裂域，方程就找到解了。可是这里有一个核心问题：系数域可扩张为分裂域的充分必要条件是什么，或者是不是分裂域都是存在的（也即等价于一元 n 次方程都是有解的）。

由于域定义了四种运算（例如四则运算），拉格朗日发现域是一种非常难以把握的集合。而且一元 n 次方程涉及的大部分域都是无限域（有无限多的元素，比如实数域，有理数域），要准确地给出系数域可以扩张为分裂域的充分必要条件是非常困难的。

（2）、伽罗华的工作

伽罗华首先是对一元 n 次多项式方程可解的定义进行改进：

简单说是指经过有限次加、减、乘、除、乘方和开方运算可以表示出方程的

根。（这个定义的严格表达是：如果一个集合包含方程的系数，且对加、减、乘、除、乘方和开方封闭，那么求根公式的存在性等价于根在这个集合中的存在性。这个结论是显然的，多想一下就明白）。

所以一个代数方程是否有解，要看我们对于解所加的限制条件而定，例如如果允许 x 可以是负数的话，$x+5=3$ 是可解的，但是如果限定 x 不能是负数，那么这个方程就无解了。

同样，假使 x 表示有理数，方程 $2x+3=10$ 是可解的。如果 x 表示整数，这方程就无解了，因为 x=3.5 在整数里面没有意义。

再例如，要三等分任意一角，若只准用直尺与圆规，这是不可能的，但是若许用别的仪器，就可能了。

所以关键的一个要点来了：一个多项式是可以因式分解的或不可因式分解的，要看在什么数域中分解而定。例如 x^2+1 在实数域中就是不可分解的，但是在复数域中却是可分解的，因为 $x^2+1=(x+\mathrm{i})(x-\mathrm{i}), \mathrm{i}=(-1)^{\frac{1}{2}}$。所以，单说一个多项式是不是可因式分解的，而不说出在什么数域内，这其实是废话。

同理，一个命题在什么范围中是对的，在什么范围中是错的，甚而至于在什么范围中是绝对没有意义的也是这个道理。

伽罗华要解决的问题是：一般高于四次的方程不能用根式解。

不能用根式解，就是说方程的根不能用方程的系数通过有限次的有理运算（加，减，乘，除）和开方得到（或者说等价于方程的根不能表达成方程系数通过有理运算形成的函数）。

例如一次方程 $ax+b=0$，方程的根是 $x=-\dfrac{b}{a}$，也即 x 的值可以用 a 除 b 而得，这是一个有理运算。

二次方程 $ax^2+bx+c=0$，两根是 $x=\dfrac{-b\pm(b^2-4ac)^{\frac{1}{2}}}{2a}$，这也可以由方程系数经过有限次的有理运算和开方而得。

同样，一般的三次，四次方程的根也表达成用有限次的有理运算和开方方程系数的函数。

显然乘方是乘法的特例（反复乘法）。开方显然不是四则运算（域中被定义的运算只有加减乘除四种），所以必须把开方通过扩域的方式被加入到求根公式

允许的运算方式中。

所以伽罗华发展出了第一个重要的概念：扩域。也即伽罗华发现了：从包含方程系数的最小的域出发，通过域的扩张逐渐添加元素，直到把方程的所有解包含在某个扩域中为止：如果我们能这样做到，方程就是有解的，否则，方程就没有一般的求根公式。

（3）、伽罗华定理

基于上述发现，伽罗华继续努力。

对有理系数的 n 次方程 $x^n + a_1x^n - 1 + a_2x^{n-2} + \cdots + a_{n-1}x + a_n = 0$，

假设它的 n 个根是 $x_1, x_2, \ldots, x_n$，伽罗华证明：

每个方程对应一个域（由方程系数和全部根组成，这个域定义为伽罗华域），每一个域与一个伽罗华群对应。

这也就意味着，伽罗华发现了研究一元 n 次方程解结构问题，可以转为研究伽罗华群结构性质。

伽罗华群定义：某个数域上任意一个一元 n 次多项式方程, 它的根的置换群里面某些置换所构成的一个子群，满足如下条件就被定义作该方程的伽罗华群：

对任意一个取有理数值关于根的多项式函数，伽罗华群中的每一个置换都使函数的值不变。同时，如果伽罗华群中每一个置换都使根的一个多项式函数的值不变，则这个多项式函数的值是有理的。

伽罗华域定义：对有理系数的 n 次方程 $x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_{n-1}x + a_n = 0$，假设它的 n 个根是 $x_1, x_2, \cdots, x_n$，方程系数生成的域是 F，E 是把 n 个根添加到 F 上生成的域，又叫伽罗华扩域或伽罗华扩张。

伽罗华定理：假定 G 为这个方程的伽罗华群，一元 n 次方程是否有根式解的充分必要条件是：假定 F 是含有这个方程的系数及 $x^n = 1$ 的各次方根的最小域，那么 F 是否可以经过有限次添加根式扩张，成为 E。

也即是否存在有限多个中间域 $Fi(i = 1, 2, \cdots, k)$, 使 $F < F_1 < F_2 < \cdots < F_k < E$，其中每个 F_i 都是由 F_{i-1} 添加 F_{i-1} 数的根式产生的扩域。F 是方程系数和 1 的 n 次方根组成的最小域。

那么如何把伽罗华域伽罗华群联系起来呢？

伽罗华定义了域上自同构群。域上的自同构群概念的引入，使域与群发生

了联系，即建立了伽罗华域的子域与伽罗华群的子群之间的一一对应关系：保持 F 元素不动的的每个自同构决定方程根的一个置换, 它属于伽罗瓦群 G, 反之,G 中每个置换引起的一个自同构, 它使 F 的元素不动。

这样就建立了 E 的自同构群和方程的伽罗瓦群之间的同构。由此建立 E 的子域 (包含 F) 和 G 的子群之间的一一对应：保持子域 F_i 不动的 G 中全部置换构成的一个子群 G_i, 让 G_i 与 F_i 对应, 而且反过来也可用 G_i 来刻划 F_i, 即 F_i 是 E 中被 G_i 的每个置换保持不动的元素全体。也即 F_i 和 G_i 存在一一对应关系。

这就是伽罗瓦基本定理。显然利用这种一一对应关系，就可由群的性质刻划域的性质, 反之亦然。因此, 伽罗华的理论是群与域这两种代数基本结构综合的结果。

那么怎么用伽罗华群性质证明方程是否可解呢？

伽罗华在拉格朗日方法基础上，认为解方程，必须从预解式开始，当他构造二项方程作为预解方程时，发现其相应的置换子群应是正规子群且指数为素数才行。利用正规子群概念可以区分合成群与单群的概念，利用它的性质就可以判别已知方程能否转化为低次方程的可解性问题。这是伽罗华的第二个重要发现。

伽罗华的的思想是：

首先定义正规子群的概念：群 G 的子群 N 是 G 的正规子群，是指对每个 $g \in G$，$g^{-1}Ng = N$；

其次是寻找极大正规子群列，确定极大正规子群列的一系列合成因子。

伽罗华证明：伽罗华域 F，如果每次所添加的根式均为素数次根，那么, 那么 F 可以经过有限次添加根式扩张，成为 E（也即方程有根式解）。这时中间域 Fi 的结构等价于使 F_{i-1} 保持不变的的 Fi 自同构置换群的结构。这样的自同构群是素数阶的循环群, 且阶数为 $(F_i : F_{i-1})$。

伽罗华因此定义：如果一个群所生成的全部合成因子都是素数，则称这个群为可解的。

这样就利用可解群的概念全面刻划了用根式解方程的特性，给出了一个方程可用根式解的判别准则是：一个方程可用根式解的充要条件是这个方程的伽罗华群是可解群。

这样伽罗华证明了：一元 n 次多项式方程能用根式求解的一个充分必要条件是该方程的伽罗华群为可解群。这时是 1832 年。

由于高于四次的一般方程的伽罗华群不是可解群，也就直接推论出高于四次的一般方程的不可解性。

也即伽罗华发现本质就是：域的无数种扩张方式其实就是有限阶的群。n 阶对称群对应着 n 次一元方程，而 5 阶和 5 阶以上的对称群不是可解群，也就是五次和五次以上的代数方程没有求根公式。

（4）、伽罗华的创造

□ 先不忙考虑求解方法，先证明解是不是存在，不然就是无用功，也即伽罗华把存在性证明与数的计算相分离，这是人类伟大的一步。

□ 通过研究根式扩张和根对称性得出代数方程是否可解。也即发现方程解的对称性和解的结构，决定是否可以根式解。

具体说就是伽罗华发现：一个多项式方程有根式解的话，各个根的对称性要满足一定关系：出现在根的表达式中的每个根式，一定可以表成方程诸根及某些单位根的有理函数。五次以上的方程这个关系不一定满足。

伽罗华的发现证明：计算不如结构重要。

伽罗华定义的群本质就是方程根形成的集合必须具有对称性质。

如果解集上定义某种两两映射（同构），如果能保持解集不变，解集就是对那个自同构对称的。

事实上，如果解集存在，保持解集不变的自同构一定是存在的（很容易证明）。因为至少有一个恒等自同构，即从自身映射到自身。再比如一元二次方程有两个根 x_1 和 x_2，那么 x_1 到 x_2 的映射也是一个自同构。

这就是说，如果某个扩张域是存在的，扩张所对应的自同构也一定存在，这两个存在性是等价的。所以扩域的研究自然而然地变成了对自同构的研究。

至此为止，我们把伽罗华的基本思想介绍完了。至于细节，我们在下面简单介绍一下，对不想麻烦的人，不看也可以。

附：五次以上方程不可解性的严格证明（给一个抽象代数教科书典型的证明定理的例子）

证：若 $S5$（五阶置换群）是可解的，则存在正规子群 N 使 $S5/N$ 可交换。

设 f 为 $S5$ 到 $S5/N$ 的自然同态，考察三项循环 $(a,b,c) \in S5$，再取另两元 d,e。令 $x=(d,b,a)$，$y=(a,e,c)$。$x^{-1}y^{-1}xy$ 的 f 像为 $x'^{-1}y'^{-1}x'y' \in S5/N$，由 $S5/N$ 可交换知 $x'^{-1}y'^{-1}x'y'=1$，即有 $x^{-1}y^{-1}xy=(a,b,c) \in N$。故 N 包含所有三轮换，同理其正规群列均包含三轮换，所以不可能结束于 1。这就是 5 次以上一元方程不可解的证明。

1.6 说点细节

其实伽罗华关键工作我们已经介绍完了。下面说点细节。

伽罗华定义的的群并不是现在抽象代数定义的群（最前面介绍的），伽罗华定义群是方程根的置换。从直觉来看，方程的解显然和它们的顺序无关，所以当置换作用于方程的解集合时，方程对这种变换而言是对称的。

伽罗华发现满足这些条件的集合（群）的结构是非常固定的。举个最简单的例子：包含三个元素的群的结构一定是 $(0,1,-1)$，其中 0 是恒等元，-1 是 1 的逆元。（但是 5 阶以上的对称群不一定是可解群，所以 5 次以上的代数方程没有一般的求根公式）。

在 1831 年的论文中，伽罗华首次提出了群这一术语，把具有封闭性的置换的集合称为群，首次定义了置换群的概念。他发现置换群是解方程的关键，方程的根是一个置换群。他从此开始把解方程问题转化为置换群结构问题 (其实群这个概念不是伽罗华原创，柯西在 1813 年就提出了，只是没能进一步发现：群的基本性质对称结构对一元 n 次多项式方程解的关系）。

（1）、群的定义

□（封闭性）集合中任意两个元素用规定的运算时，所得的结果还是系统中的一个元素。也即集合 G，任意 x,y 属于 G，集合 G 上定义的运算为 $*$，$x*y$ 也一定属于 G。（这个运算 * 的定义是广义的，既可以是加减乘除等运算，也可以是旋转，置换等等一切行为）。

例如：一个整数加到另一个整数上去的结果还是一个整数；两个有理数相乘的结果还是一个有理数；一个置换将 x_1 变成 x_2,x_2 变成 x_3,x_3 变成 x_1, 另外一

个置换是将 x_2 变成 x_1,x_3 变成 x_2,x_1 变成 x_3，那末这两个置换结合仍然是一个置换；平面一个 60 度的旋转（逆时针方向）之后跟着一个 120 度的旋转 (逆时针方向)，结果是一个 180 度的旋转 (逆时针方向)，仍然是一个旋转等等。

□ 结合律必须成立。也即任意 x, y, z 属于 G，$(x * y) * z = x * (y * z)$。

□ 集合中必须含有单位元，也即与集合中任意另一个元素运算的结果仍是那另一个元素。也即集合 G 存在单位元 e，任意一个 x 属于 G，$e * x = x$。

例如，在定义加法的整数中，单位元是 0，因为 0 与任何整数相加的结果还是那个整数；在定义乘法的有理数中，单位元是 1，因为任意一个有理数用 1 乘了之后的积还是那个有理数；在置换中，单位元就是那个将 x1 变成 x1，x2 变成 x2，x3 变成 x3 的置换，因为任意一个置换和这个置换结合的结果还是那个置换；在平面旋转中，单位元就是那个 360 度的旋转，因为集合中任意一个旋转和这个旋转结合的结果还是那个旋转等等。

□ 每个元素必须有一个逆元素：一个元素和他的逆元素用集合上定义的运算结合的结果是单位元。也即任意 x 属于 G，存在 x^-,$x^- * x = e$。

例如，在整数集合中，定义加法，3 的逆元素就是-3，因为 3 加上-3 的和是 0；在有理数集合中定义乘法，则 a/b 的逆元素是 b/a, 因为 a/b 和 b/a 相乘的积是 1；在置换中，将 x_1 变成 x_2,x_2 变成 x_3,x_3 变成 x_1 的置换的逆元素是那个将 x_2 变成 x_1,x_3 变成 x_2，x_1 变成 x_3 的置换，因为这两个置换结合的结果是那个将 x_2 变成 x_2,x_3 变成 x_3,x_1 变成 x_1 的置换；在平面旋转中，那个 60 度的旋转（逆时针方向）的逆元素是一个一个顺时针方向的 60 度的旋转，因为这两个旋转结合的结果和那个 360 度的旋转一样。

满足上述的四条性质，就是一个群。

如果在整数上定义加法，但是若把 0 去掉，就不成为群了，因为没有单位元；一切整数用乘法作集合中的运算不是群，例如 3 的逆元素 $\frac{1}{3}$ 不在整数集合中。

所以一个集合是否是群，不但与元素有关，也与运算有关。

前面已经说了，群的元素不必一定是数，可以是一种运动（如平面旋转），也可以是一种动作（例如置换）；运算不必一定要是加法或乘法，或寻常算术，抽象代数中所称为的运算，可以是任何定义，例如乘法可以是一个置换跟着另一

个置换，也可以说是一个置换乘另一个置换。这个乘法与普通算术或代数中乘法不是一个概念，千万不要蒙，而且群定义的广义的乘法的性质可以和普通乘法的性质大异，例如，在普通的乘法中，$2*3=3*2$（普通的乘法是适合交换律的），也即普通乘法中因子的次序可以交换，结果相同。可是，置换中的“乘法”，交换律就不成立了，例如将 x_1 变成 x_3，x_3 变成 x_1，x_2 变成 x_2 的置换和一个将 x_1 变成 x_2，x_2 变成 x_3，x_3 变成 x_1 的置换就没有交换律，如果先进行第一个置换然后进行第二个置换于式子 $x_1x_2+x_3$，那末，这式子先变成 $x_3x_2+x_1$，再变成 $x_1x_3+x_2$；如果将置换的次序交换一下，那末，原来的式子先变成 $x_2x_3+x_1$，再变成 $x_2x_1+x_3$，这个结果显与前面一个不同。所以群里面定义的“乘法”是不需要适合交换律的，因此，相乘时元素的次序很重要；两个元素用运算结合时当照一定的次序结合。

（2）、置换群

伽罗华用来解方程的是置换群 (SubstitutionGroup)，下面先介绍一下记号。

一个将 x_1 变成 x_2,x_2 变成 x_3,x_3 变成 x_1 的置换，可以用简单记号来表示：x 可以省去，只要用 1,2,3 来代表于是这个置换可以记作 (123)，这记号的意思是说：1 变作 2，2 变作 3，3 变作 1。也即：x_1 变作 x_2，x_2 变作 x_3，x_3 变作 x_1。（每个数变作他后一个数，而最后的一数则变成最先的一数，如此完成一个循环）

同样，一个将 x_2 变成 x_3,x_3 变成 x_1,x_1 变成 x_2 的置换可以记作 (231)；同样 (132) 表示一个将 x_1 变成 x_3,x_3 变成 x_2,x_2 变成 x_1 的置换；又如 (13)(2) 或 (13) 表示一个将 x_1 变成 x_3,x_3 变成 x_1,x_2 变成 x_2 的置换，所以前面讲乘法交换律时所说两个置换相乘的例子，若照第一种次序是 (13)(123)=(23)；若照第二种次序是 (123)(13)=(12)，由这两个式子就知道这种乘法是不适合交换律的，将一个元素右乘或左乘另一个元素，他的结果是完全不同的。

一个群的一部分元素构成一个群，这种群称为子群 (Subgroup)。例如整数集定义加法成为群，单拿偶数集，定义加法，也成一群：因为群的四个性质都能适合：

- 两个偶数的和还是偶数；
- 零是单位元；
- 一个正偶数的逆元素是一个负偶数，而一个负偶数的逆元素是正偶数；

□ 结合律成立。

所以单是偶数全体对于加法而言是一个群，这个群就是是那个由一切整数定义加法而成的群的子群。

再例如，一个置换群（即是以置换作元素的群）也可以有子群。

例如，$1,(12),(123),(132),(13),(23)$ 六个置换构成一个群（1 表示那个不动置换，即是将 x_1 变成 x_1,x_2 变成 x_2,x_3 变成 x_3 的置换)，因为群的四条性质都成立：这六个置换中每两个的积还是这六个中的一个置换，例如 (12)(123)=(13)，(123)(132)=1,(13)(23)=(123)，(123)(123)=(132)，等等）单位元是 1；每个元素的逆元素都在这六个元素之中，比如 (123) 的逆元素是 (132)，(12) 的逆元素是 (12) 等等；结合律成立。

现在从这六个置换中取出 1 和 (12) 两个来，这两个元素也成为一个群，这是原来那个群的子群。

很容易证明：子群的元数 (即集合中元素的个数) 是原来的群的元数的约数（拉格朗日定理）。

（3）、不变子群

最重要的子群是不变子群。

变换的直观定义：群中一个元素若以另一个元素右乘，再用这另一个元素的逆元素左乘，所得结果称为元素应用另一个元素的变换。

例如一个元素 (12), 我们用另一个元素 (123) 去右乘他，再用 (123) 的逆元素 (132) 去左乘他，结果是 (132)(12)(123)=(23)，(23) 就称为 (12) 应用 (123) 的变换。

定义：一个子群中任何元素应用原来的群中任何元素的变换，若仍是子群中的元素，这子群就称为原来那个群的不变子群。

对伽罗华理论来讲，不变子群是很重要的概念。

定义：设 H 是 G 的不变子群，假如 G 中没有包含 H 而且比 H 大的不变真子群存在时，H 就称为 G 的一个极大不变真子群。

定义：假设 G 是一个群，H 是 G 的一个极大不变真子群，K 是 H 的一个极大不变真子群..... 若将 G 的元数用 H 的元数去除，H 的元数用 K 的元数去除，..... 如此所得的系列数，就称为群 G 的组合因数，假设这些组合因数都是素数，就说 G 是一个可解群（可解的含义后面再介绍）。

在有些群中，群中的一切元素都是某一个元素 (不是单位元) 的乘幂，比如在群 1,(123),(132) 中，2(123)=(123)(123)=(132)3(123) =(123)(123)(123)=1，这群中的元素都是 (123) 的乘幂，像这种群，称为循环群。

在一个置换群中，如果每个元素都有一个而且只有一个置换将元素换成其他某一个元素 (这个元素也可以和原来那个元素相同)，那末，这个群就称为正置换群。

例如前面所说的群 1,(123),(132) 在 1 中 x_1 变成 x_1, 在 (123) 中 x_1 变成 x_2, 在 (132) 中 x_1 变成 x_3，..... 所以这是一个循环正置换群。这种群在方程的应用上很重要。

伽罗华证明：对于一个一定的数域，方程 $x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0$ 的根都能构造一个置换群，群的阶数是 $n!$。

例如对三次方程：$ax^3 + bx^2 + cx + d = 0$，假定它的三个根 x1,x2,x3 是不同的，随便取一个这三个根的函数，例如 $x1x2 + x3$，在这函数中，我们若将这些 x 互相替换，那末，一共有多少种置换呢？

显然只有 1,(12),(13),(23),(123),(132) 六个置换（3!）。

(12) 置换，也即将 $x_1x_2 + x_3$ 变成 $x_2x_1 + x_3$；(13) 置换，就是将 $x_1x_2 + x_3$ 变成 $x_3x_2 + x_1$ 等等。

(123) 置换就是把原来的函数变成 $x_2x_3 + x_1$。而 1 就是不动置换了。所以对于这三个 x，一共有 3! 种可能的替换。

同理，对于四个 x 有 4！种可能的置换，一般的情形，n 个 x 就有 n! 可能的替换。

当然一个函数进行一个置换的时候，函数的值可以因此而变，也可以仍旧不变，例如若将 (12) 这个置换施行于函数 x1+x2，这函数的值不变，可是，若将 (12) 施行于函数 x1-x2，函数的值就由 x1-x2 一变而为 x2-x1 了。

计算一个已知 n 次方程的伽罗华群是很困难的，因此伽罗华认为目标不在于计算伽罗华群，而是证明：

对任意 n 次方程，其伽罗华群是方程根的最大置换群 $S(n)$，$S(n)$ 是由 n! 个元素集合构成的，S(n) 中的元素乘积实际上是指两个置换之积。现在把 S(n) 中的元素个数称为阶，S(n) 的阶是 n!。

伽罗华找出方程系数域中的伽罗华群 G 后，找到它的最大真子群 H_1，用有理运算来构造根的一个函数 $\Phi 1(x)$，$\Phi 1(x)$ 的系数属于方程的系数域 R，并且在 H_1 的置换下不改变值，但在 G 的所有别的置换下改变值。

再用上述方法，依次寻找 H_1 的最大子群 H_2，再找到一个函数 $\Phi 1(x)$，$\Phi 1(x)$ 的系数属于方程的系数域 $R1$;再找到 H_2 的最大子群 $H_3, \cdots$ 于是得到 $H_1, H_2, \cdots, H_m$，直到 H_m 里的元素恰好是恒等变换（即 H_m 为单位群 1）。

在得到一系列子群与逐次的预解式的同时，系数域 R 也随之一步步扩大为 $R_1, R_2, \cdots, R_m$，每个 R_i 对应于群 H_i。当 $H_m = 1$ 时，R_m 就是该方程的根域，其余的 $R_1, R_2, \cdots, R_{m-1}$ 是中间域。

我们从拉格朗日工作已经知道一个方程可否根式求解与根域的性质密切相关。

于是，伽罗华引出了根式求解原理，并且还引入了群论中的一个重要概念“正规子群”

（4）、正规子群

正规子群定义：设 H 是 G 的一个子群，如果对 G 中的每个 g 都有 $gH = Hg$，则称 H 为 G 的一个正规子群。

伽罗华证明：当作为约化方程的群 (如由 G 约化到 H_1) 的预解式是一个二项方程 $x^p = A$ (p 为素数) 时，则 H_1 是 G 的一个正规子群。反之，若 H_1 是 G 的正规子群，且指数为素数 p，则相应的预解式一定是 p 次二项方程。

极大正规子群：如果一个有限群有正规子群，则必有一个子群，其阶为这有限群中所有正规子群中的最大的，这个子群称为有限群的极大正规子群。

一个极大正规子群又有它自己的极大正规子群，这种序列可以逐次继续下去。因而任何一个群都可生成一个极大正规子群序列。

把一个群 G 生成的一个极大正规子群序列标记为 $G, H_1, H_2, H_3 \cdots$, 则可以确定一系列的极大正规子群的合成因子 $[G/H1], [H1/H2], [H2/H3] \cdots$。合成因子 [G/H]=G 的阶数/ H 的阶数。例如对上面的四次方程 $x^4 + px^2 + q = 0$，H_1 是 G 的极大正规子群，H_2 是 H1 的极大正规子群，H3 又是 H2 的极大正规子群，即对方程 $x^4 + px^2 + q = 0$ 的群 G 生成了一个极大正规子群的序列 G、H1、H2、H3。

伽罗华在此基础上定义可解群：如果它所生成的全部极大正规合成因子都是素数。也即伽罗华群生成的全部极大正规合成因子都是素数时，方程可用根式求解。若不全为素数，则不可用根式求解。

或者说：当且仅当一个方程系数域上的群是可解群时，该方程才可用根式求解。

可解性的性质在某一意义上是可继承的，如：

若 G 为可解的，且 H 为 G 的子群，则 H 也是可解的。

若 G 是可解的，且 H 为 G 的正规子群，则 G/H 也是可解的。

若 G 是可解的，且存在一 G 满射至 H 的同态，则 H 也是可解的。

若 H 及 G/H 为可解的，则 G 也是可解的。

若 G 及 H 为可解的，则其直积 $G \times H$ 也是可解的。

例如 $x^4 + px^2 + q = 0$，它的 [G/H1]=8/4=2，[H1/H2]=2/1=2，2 为素数，所以 $x^4 + px^2 + q = 0$ 是可用根式解的。

再看一般的 n 次方程，当 n=3 时，有两个二次预解式 $t^2 = A$ 和 $t^3 = B$，合成序列指数为 2 与 3，它们是素数，因此一般三次方程可根式解。同理对 $n = 4$，有四个二次预解式，合成序列指数为 2，3，2，2，于是一般四次方程也可根式求解。

（5）、5 次以上一元方程不可解

一般 n 次方程的伽罗华群是 s(n)，s(n) 的极大正规子群是 A(n) (A(n) 是由 s(n) 中的偶置换构成的一个子群。如果一个置换可表为偶数个这类置换之积，则叫偶置换。)，A(n) 的元素个数为 s(n) 中的一半，且 A(n) 的极大正规子群是单位群 1，因此 [s(n)/A(n)]=n!/(n!/2)=2，[A(n)/I]=(n!/2)/1=n!/2，2 是素数，但当 n ≥5 时，n！/2 不是素数，所以一般的高于四次的方程是不能用根式求解的。

例如，四次方程 $x^4 + px^2 + q = 0$，p 与 q 独立，系数域 R 是添加字母或未知数 p、q 到有理数中而得到的域，先计算出它的伽罗华群 G。

G 是 $S(4)$ 的一个 8 阶子群,$G = E, E1, E2, \cdots E7$，其中 $E = 1, E1 = (1234), E2 = (2134), E3 = (2143), E4 = (3412), E5 = (4312), E6 = (3421), E7 = (4321)$。

要把 R 扩充到 R1，需在 R 中构造一个预解式:$t^2 - (p^2 - 4q) = 0$，

则添加预解式的根 $((p^2-4q))^{\frac{1}{2}}$ 到 R 中得到一个新域 R1，于是可证明原方程 $x^4+px^2+q=0$ 关于域 $R1$ 的群是 $H1$，$H1=E,E1,E2,E3$，并发现预解式的次数等于子群 $H1$ 在母群 G 中的指数 $8\div 4=2$（即指母群的阶除以子群的阶）。

然后构造第二个预解式 $t^2-2(-p-(p^2-4q)^{\frac{1}{2}})$，

解出根 $(2(-p-(p^2-4q)^{\frac{1}{2}}))^{\frac{1}{2}}$ 在域 R1 中添加得到域 R2，同样找出方程 $x^4+px^2+q=0$ 在 R2 中的群 H2，$H2=E,E1.$

此时第二个预解式的次数也等于群 $H2$ 在 $H1$ 中的指数 $4\div 2=2$。

再然后构造第三个预解式 $t^2-2(-p+(p^2-4q)^{\frac{1}{2}})$，得它的根 $2(-p+(p^2-4q)^{\frac{1}{2}}))^{\frac{1}{2}}$，把添加到 $R2$ 中得扩域 $R3$，此时方程 $x^4+px^2+q=0$ 在 R3 中的群为 $H3\square H3=E$，即 H3=1，则 R3 是方程 $x^4+px^2+q=0$ 的根域，且该预解式的次数仍等于群 H3 在 H2 中的指数 $2\div 1=2$。

在这个四次方程中，系数域到根域的扩域过程中每次添加的都是根式，则方程可用根式解。

这种可解理论对于一般的高次方程也同样适用，只要满足系数域到根域的扩域过程中每次都是添加根式，那么一般的高次方程也能用根式求解。

现仍以四次方程 $x^4+px^2+q=0$ 为例，伽罗华从中发现了这些预解式实质上是一个二次的二项方程，既然可解原理对高次方程也适用，那么对于能用根式求解的一般高次方程，它的预解式方程组必定存在，并且所有的预解式都应是一个素数次 p 的二项方程 $x^p=A$。由于高斯早已证明二项方程是可用根式求解的。因此很容易得到：如果任一高次方程所有的逐次预解式都是二项方程，则能用根式求解原方程。

至此，伽罗华完全解决了方程的可解性问题。

（6）、用直尺与圆规的作图

伽罗华解决了用直尺与圆规的作图难题。

伽罗华发现了判别方程能否用根式解的方法后，他还解决了如何求一个能用根式解的方程的根的方法，这方法是利用一组辅助方程，这些辅助方程的次数恰是原来那个方程的群的组合因数。

基本流程如下：先把第一个辅助方程的根加入数域 F 中，将数域扩大了可

以增加 $P(y)$ 分解因数的可能性，也能将 $P(y)$ 的不可约部分减少，因此能将方程的群变小，当然，必须数域扩大了之后的确能继续分解 $P(y)$ 的因数，才会成立。

现在假设数域经第一个辅助方程的根加入而扩大了，而且使分解因数的工作因之可以再继续下去，结果使方程在这扩大了的数域 $F1$ 中的群是 H。

再将第二个辅助方程的根加入 $F1$ 中，使方程的群变为 K，如此持续，直到后来，方程在那个最后扩大成的数域 Fm 中的群是 1。函数 $x1$ 显然不能被群 1 中的置换变更他的值，所以 $x1$ 必在数域 Fm 中。仿此，其余的根也都在 Fm 中。

这样先决定了方程的群和此群的组合因数，才知道辅助方程的次数。由此我们可以知道什么样的数应该加入原来的数域里去，而把方程的群变为 1。于是可以决定方程的根存在于怎样一个数域中。

现在用方程 $x^3-3x+1=0$ 为例,这个方程在有理数域中的群是由 1,(123),(132) 三个置换构成的，其唯一极大不变真子群是 1，所以组合因数是 3, 所以有一个次数是 3 的辅助方程，而这个辅助方程的根含有一个立方根，所以这个立方根必须加入数域中，才能使方程的群变为 1，这样原来的方程的根可以从有理数域中的数及这个立方根用有理数运算得出。

直尺与圆规作图等价于直线和圆作交点图。也即求一次和二次方程的交点，只要解一个二次方程就可以把交点的坐标用有理运算和平方根表作系数的函数。所以凡是能用直尺与圆规作出的图都可以有限次的加，减，乘，除和平方根表出，而且假使给了两线段 a,b 和单位长度，我们可以用直尺与圆规作出他们的和 $a+b$, 差 $a-b$, 积 ab, 商 $\frac{a}{b}$, 以及这些量的平方根如 $(ab)^{\frac{1}{2}}$,$b^{\frac{1}{2}}$ 之类，这种运算当然可以重复应用于一切已经作出的线段。

一个作图单用直尺，圆规是否可能时，必须作出一个表示这作图的代数方程：假使这方程在数域中可以分解成单是一次和二次的代数式，那么，一切实数根当然都能用直尺与圆规作出。即使方程不能分解成上述的样子，只要方程的实数根能用有限次的有理运算与平方根作已知的几何量的函数，那末这作图单用直尺，圆规还是可能的，否则这作图就不可能了。

也即立方根是无法靠直尺和圆规作出的。

如果能够找到一个三等分角的方程是不能用直尺与圆规三等分，那末用直

尺和圆规三等分任意角的作图就不可能了。

取 120 度角来三等分。假定这角位于一个半径是单位长的圆中心。假使能作出 $cos40$ 度来，那末，只要取 $OA = cos40$, 于是 a 就是一只 40 度的角，而三等分 120 度的作图就完成了。

应用三角恒等式 $2cos(3\alpha) = 8cos(\alpha^3) - 6cos(\alpha)$，令 $x = 2cos\alpha$, 证明：

$$
\begin{aligned}
cos3\alpha &= cos(2\alpha + \alpha) \\
&= cos2\alpha cos\alpha - sin2\alpha sin\alpha \\
&= (2cos^2(\alpha) - 1)cos\alpha - (2sin\alpha cos\alpha)sin\alpha \\
&= 2cos^3(\alpha) - cos(\alpha) - 2sin^2(\alpha)cos\alpha \\
&= 2cos^3(\alpha) - cos\alpha - 2(1 - cos^2(\alpha))cos\alpha \\
&= 2cos^3(\alpha) - cos\alpha - 2cos\alpha + 2cos^3(\alpha) \\
&= 4cos^3(\alpha) - 3cos(\alpha)
\end{aligned}
$$

则有：$2cos3\alpha = x^3 - 3x$

因为 $3\alpha = 120$ 度，$cos3\alpha = -1/2$，所以上面的方程可以写作 $x^3 - 3x + 1 = 0$

这正是以前讨论过的方程。

现在作一个半径是单位长的圆，而且可以作 $OB = 1/2$，于是角 $AOC = 120$ 度。因为所给的只有单位长，所以数域限定在有理数域。

所以要解这个方程，必须将一个立方根加入于有理数域中，然而一个立方根是不能用直尺与圆规作出的，这样，我们可以知道：用直尺与圆规三等分任意角是不可能的。

以相似的方法，不难证明用直尺，圆规解决立方倍积问题也是不可能的，对于这个问题，方程是 $x^3 = 2$

数域是有理数域，这方程在这个数域中的群含有六个置换。可以当证明须加入一个平方根和一个立方根于有理数域中，方程的群才会变成 1。又因一个立方根是不能用直尺，圆规作出的，所以我们这个立方倍积问题是不可能的。

类似的，也可以可以应用群论去探讨正多边形作图的问题。

附：伽罗华的关键定理的思想脉络：

问题：要证明一个方程若有一个伽罗华可解群，这方程就可用根式解。

伽罗华的思想脉络如下：在二次方程 $x^2+bx+c=0$ 的两个根 $x1,x2$ 中，用韦达定理有 $x_1+x_2=-b$ 与 $x_1x_2=c$ 的关系，那么为什么不从这两个方程中去解 x_1,x_2 呢？因为这条路是走不通的，因为经过计算的结果是与原来的二次方程丝毫也没分别。

但是，如果能得到一对都是一次的方程，x_1 和 x_2 就可以求得了。

假设方程 $f(x)=0$ 有 n 个相异的根，而且由方程的系数及 $x^n=1$ 的 n 次根决定的数域中，此方程的群是一个元数为素数的循环正置换群。

为什么伽罗华要先引进这个 1 的 n 个 n 次根呢？

先看看 1 有三个立方根：$1,-\frac{1}{2}+\frac{1}{2}(-3)^{\frac{1}{2}},-\frac{1}{2}-\frac{1}{2}(-3)^{\frac{1}{2}}$,(通常都记作 1，$\omega$，$\omega^2$)（在一般的情形，1 有 n 个 n 次根，这 n 个 n 次根记作 $1,\rho,\rho^2,\cdots,\rho^{n-1}$）

1 的三个立方根只包含有理数和有理数的根数，同样 1 的 n 个 n 次根也只包含有理数和有理数的根数，所以这种数加入数域中去时并不影响到方程是能用根式解的命题。

因为前面假定这个方程的群是一个元数为素数的循环正置换群，群中元素都是置换群，群中的元素都是置换 $(123\cdots n)$ 的乘幂，这个置换的 n 次乘幂就是不动置换。

现在构造一组一次方程（n 个）：

$x_1+(\rho^k)x_2+(\rho^{2k})x_3+.....+(\rho^{(n-1)k})x_n=\gamma k$

此处 k 的值为 0 与 $n-1$ 间之任何整数。

例如当 $k=0$ 时，上式就成为：

$x_1+x_2+x_3+.....+x_n=\gamma 0$

当 $k=1$ 时，上式就成为：

$x_1+\rho x_2+\rho^2x_3+.....+\rho^{n-1}x_n=\gamma 1$ 等等。

因为一个方程的最高次项系数是 1，则诸根之和等于方程中第二项的系数的负值，所以 $\gamma 0$ 之值可以直接从方程的系数中求得。

现在要将置换 $(123\cdots n)$ 作用于上面方程组的左端，左端就成为：

$x_2+(\rho^k)x_3+(\rho^{2k})x_4+.....+(\rho^{(n-1)k})x_1$ 等等，

若将左端用 $\rho^(-k)$ 一乘，也可得出同样的结果，这是因为 $\rho^n=1$ 的缘故。所以置换 (1234.....n) 将 γk 之值变为 $\rho^{-k}\gamma k$。

又因 $\rho^n = 1$，所以 $\gamma k^n = (\rho^{-k}\gamma k)^n$

所以置换 (1234.....n) 不变更 γk^n 的值。同样，群中其他的置换也不变 γk^n。

这样群中一切置换都不变更 γk^n 之值，γk^n 之值必在数域中。因此，γk^n 是数域中某一个数的 n 次根，这就是说：所有 γ 的值都可由根式得到（对于定义的数域而言）。而上面方程组中可以将 x 用 ρ 与 γ 表出，于是这组方程是可以用根式解的。这些 x 就是方程 $f(x) = 0$ 根。

所以已经证明：如果方程在一数域中的群是元数为素数的循环正置换群，则此方程必可用根式解。

举例来说：方程 $x^3 - 3x + 1 = 0$ 在有理数域中的群是 1，(123),(132)；这是一个元数为素数的循环正置换群，所以可以从

$$x1 + x2 + x3 = 0$$

$$x1 + \omega x2 + \omega^2 x3 = \gamma 1$$

$$x1 + \omega^2 x2 + \omega x3 = \gamma 2$$

这三个一次方程求解。此处 ω 表示 1 的一个虚立方根，$\gamma 1$ 与 $\gamma 2$ 可以由数域中的数的根数而得。换句话说，如果把根加入到数域中去，则 x 都存在于扩大的数域中。

假使方程的群是一个可解群时，由于组合因数都是素数，这方程还是能用根式解的，因为这时候每个辅助方程在那个用前几个辅助方程的根扩大成的数域中的群是一个元数为素数的循环正置换群，这些辅助方程都能用根式解。因为这些加入原来的数域去的辅助方程的根，都只不过是原来的数域中的数的根数而已。所以只要方程的群是可解群，方程就是能用根式解的。

在一般的情形，取：

$y^2 = (x(1) - x(2))^2(x(1) - x(3))^2 \cdots (x(n-1) - x(n))^2$ 作第一个辅助方程，此式右端是所有每两个根之差的平方之积。假若方程的第一项系数是 1 的话，则上式的右端正是方程的判别式，例如二次方程 $x^2 + bx + c = 0$ 的两个根 x_1,x_2 之差之平方是 $(x_1 - x_2)^2 = (x_1 + x_2)^2 - 4x_1x_2 = b^2 - 4c$，这恰是方程的判别式。同样，高次方程的判别式也可从系数求得。现在第一个辅助方程的两个根就是这判别式的两个平方根，将这两个平方根加入数域中，方程式在这新的数域 F1 中的群是 H，再照同样方法用其余的辅助方程进行下去。

设若所要解的方程是一个一般的三次方程，将第一个辅助方程的根加入原来的数域之后，方程的群变为 H，在这情形，H 是一个元数为素数的循环正置换群，所以我们可以利用

$x_1 + x_2 + x_3 = 0$

$x_1 + \omega_2^x + \omega^2 x_3 = \gamma 1$

$x_1 + \omega^2 x_2 + \omega_3^x = \gamma 2$

这三个一次方程来解原来的三次方程，此中的 $\gamma 1$，$\gamma 2$ 可由数域 (由三次方程的系数以及第一个辅助方程式的根决定）中的数的根数求得。换句话说，假使把 $\gamma 1$，$\gamma 2$ 的值也加入数域中，则方程的群变为 1，这也就是说，x1,x2,x3 存在于这个最后经 $\gamma 1$，$\gamma 2$ 之加入而扩大成的数域中。

如此就已经证明：方程在一个由其系数与 1 之 n 个 n 次根而决定的数域中的群若是一个可解群，则此方程是可以用根式解的。

当然，如果方程在一个含有其系数的数域中的群是可解群，则对于这数域而言，此方程是可以解的。

至此伽罗华解决了为何五次以上之方程式没有公式解，而四次以下有公式解。

他也解决了古代三大作图问题中的两个：“不能任意三等分角”，“倍立方不可能”。

对上述思想再举一个简单例子：

二次方程 $x^2 + 3x + 1 = 0$, 有两个根 x_1, x_2, 因为只有两个根，所以可能的置换只有 1 和 (12) 两种（也即是 $S(2)$ 置换群），所以这方程的伽罗华群或者含有这两个置换，或者只有 1 一个，至于是什么，这就要凭在什么数域中而决定了。

现在取函数 $x_1 - x_2$，从韦达定理中我们知道：二次方程 $x^2 + bx + c = 0$ 的两个根之差是 $x_1 - x_2 = (b^2 - 4c)^{\frac{1}{2}}$，$b = 3, c = 1$, 所以 $x_1 - x_2 = 5^{\frac{1}{2}}$ ，如果所讨论的数域是有理数域，这个函数的值不在数域中，所以群中必有一个置换，他能变更这函数的值。而 1 和 (12) 两个置换中只有 (12) 变更函数 x1-x2 的值。所以伽罗华群中必含有 (12), 因此，这方程在有理数域中的伽罗华群是由 1,(12) 两个置换构成的。

如果所讨论的数域是实数域，显然 $5^{\frac{1}{2}}$ 在其中，所以 S（2）群中一切置换都

不改变函数 $x_1 - x_2$ 的值。所以 (12) 不能在伽罗华群中，这方程在实数域中的伽罗华群是由 1 一个置换构成的。

再以方程 $x^3 - 3x + 1 = 0$ 为例，假设三个根为 x_1, x_2, x_3，所以至多有六种可能的置换，即是 1,(12),(13),(23),(123),(132)（即 $S(3)$ 置换群）。

求这方程在有理数域中的伽罗华群，我们应用 $(x_1-x_2)(x_1-x_3)(x_2-x_3)$ 这个函数，根据韦达定理，$(x_1-x_2)(x_1-x_3)(x_2-x_3)$ 之值是 $\pm(-4c^3-27d^2)^{\frac{1}{2}}$。现在 $c=-3, d=1$, 所以 $(x_1-x_2)(x_1-x_3)(x_2-x_3) = \pm 9$，$\pm 9$ 是有理数，在有理数域中，伽罗华群中一切置换都不能变更函函数的值。但在上列六个置换中，只有 1,(123),(132) 不变更这数的值，所以这个三次方程在有理数域中的伽罗华群的元素或者就是这三个置换，或者只是 1 一个，所以单利用函数 $(x_1-x_2)(x_1-x_3)(x_2-x_3)$ 还不能决定这个方程在有理数域中的伽罗华群。我们再应用另外一个函数 x1，如果群中只有 1 一个元素，那么，1 不会变更函数 x_1 的值，所以 x_1，必在有理数域中，换句话说，这个三次方程的根 x_1 必须是有理数，同样的道理，x_2, x_3 也须是有理数，但是，这个三次方程没有一个根是有理数，所以，他在有理数域中的伽罗华群不能单含 1 一个元素，个伽罗华群必定是由 1,(123),(132) 三个元素构成的。

如此，我们利用 $(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$ 和 x_1 两个函数而决定了这个方程在有理数域中的伽罗华群。

上面讨论的这个三次方程也是讨论直尺圆规三等分任意角问题的基本方程。

至此，我们已经知道什么叫做一个方程在一个数域中的伽罗华群，而且知道如何去求。

根据前面介绍的伽罗华定理，我们知道一个方程在一个含有他的系数的数域中的群若是可解群，则此方程就能用根式求解，而且仅满足这个条件的方程才能用根式解。

例如对一般的二次方程:$ax^2 + bx + c = 0$，设两个根是 x_1, x_2, 在一个含有他的系数的数域中的置换群的元素是 1 和 (12), 这个置换群的唯一的极大不变真子群是 1，所以此群的组合因数是 2/1=2 是一个素数，因此，根据伽罗华定理，二次方程都可用根式解。

再例如，一般的三次方程 $ax^3 + bx^2 + cx + d = 0$，三个根 x_1, x_2, x_3，在一个

含有他的系数的数域中，他的群含有 1,(12),(13),(23),(123),(132) 六个置换，此群的唯一极大不变真子群 H 含有 1,(123),(132) 三个置换，而 H 的唯一极大不变真子群是 1，所以组合因数是 6/3=2, 与/1=3, 两个都是素数，所以三次方程都是可用根式求解。

再例如四次方程 $ax^4+bx^3+cx^2+dx+e=0$，在一个含有其系数的数域中的群的元数是 $4!=24$, 按照前面计算，能够得到这个群的组合因数是 2,3,2,2, 这些都是素数，所以四次方程也都可以用根式解。

对于一般的五次方程，G 含有 5! 个置换，G 的极大不变真子群 H 含有 5!/2 个置换，而 H 的唯一极大不变真子群是 1，所以组合因数是 2 与 5!/2，5!/2 当然不是素数，所以一般的五次方程是不能用根式解的。

其实，对于一般的 n 次方程，n 若是大于 4，组合因数便是 2 与 $n!/2$ 而后者当然不是素数。

这样就得到了用方程结构来决定一个方程是否能用根式求解。

但是如果方程的群是一个元数为素数的循环正置换群，这方程的确可以通过辅助方程降阶简化，也即可以根式解。

1.7 感想

一般说来，一个抽象的集合不过是一组元素而已，无所谓结构，没有结构的集合，是没有意义的。数学研究的集合是定义了运算或者变换的（系统科学研究的集合，上面定义的可能是正反馈，负反馈，延迟，发散，收敛等等），这些运算或变换，就形成了集合的结构，也即定义了集合中元素的关系。伽罗华群结构思想是人类第一次将对象按照结构进行研究，并不管研究对象和运算具体是什么（群上面定义的运算，可以是加法，也可以是变换等等）。

伽罗华思想的价值是开启了现代数学的大门，使数学从运算转向研究运算性质，也即集合的结构。所以伽罗华思想是开启后来法国布尔巴基学派以数学结构观念统一数学的先导（布尔巴基学派简介在上一篇介绍数学基础时介绍过），布尔巴基实现了伽罗华没有实现的理想，他们把康托尔的集合论及希尔伯特的

公理化方法作为统一数学的基础，从抽象群的公理理论出发，通过分析抽象群结构，搞清楚了人类的抽象结构概念是如何产生的，例如他们通过群就是在某一集合中定义有结合性，么元和逆元的一个运算（这三个性质叫群结构公理），就抽象出结构概念的一般特点是：满足一定条件公理的关系集合。所以布尔巴基认为集合上的关系是数学至关重要的概念，因为关系是各种运算的抽象，是构成一个结构的基础，不同的关系可以构成不同的结构。布尔巴基学派就从结构观点出发，选出三种基本结构：代数结构、序结构、拓扑结构，作为元结构，通过从简单到复杂，从一般到特殊的层次概念，构造出各种不同的结构，如复合结构、多重结构、混合结构，建立了各种公理理论，在此基础上，统一了人类目前所有的数学学科。这其实是伽罗华思想的进一步发展：数学本质就是研究结构的。布尔巴基所做的只是把伽罗华已经确定的观念推广而已。

群的抽象定义是凯莱提出的，到 20 世纪初，已经成为数学的核心概念，几乎所有大数学家都认为其是数学的中心概念和统一数学的基础概念。如外尔就说过：没有群就不可能理解近代数学。庞加莱也曾说过：可以说群论就是那摒弃其内容化为纯粹形式的整个数学。

总之，群论是十九世纪最杰出的数学成就，是人类摆脱幼年思维的标志，也即从此摆脱了依赖直观 + 计算来理解世界。群论以结构研究代替计算，把人类从偏重计算研究的思维方式转变为用结构观念研究的思维方式，是物理学和化学发展的重要推动。

抽象的力量是巨大的，Feynman 认为用代数角度而不是偏微分方程来理解量子力学要容易得多，因为代数的抽象恰好可以避免望文生义和误入歧途。而且偏重计算的偏微分方程会导致学生舍本求末，陷入细节而难以抓住本质。

代数虽然没几何直观，但是面对 N 维空间时，其实几何直观优势已经荡然无存（所以克莱因才要用抽象代数统一几何）。面对直观以外的世界，我们唯一可以依靠的只有抽象 + 逻辑。

几何与代数的特点很像以现象研究为对象的初等物理和以本质研究（不变量研究）为主的理论物理。

代数通过不断的抽象来提炼更加基本的概念，例如两个群，不论它们的元素真实背景是什么（这些元素不管描述的是膨胀、收缩、转动、反演、振动、声

音、流体、电磁波等等)，只要运算性质相同，彼此就是同构的，并且可以因此认为是相同的代数结构而不加区别。

代数的每一次抽象都是学科升级的过程。

例如克莱因用群论来把几何中的许多互不相干的分支之间建立了内在的联系。

克莱因对几何学的定义：几何学是当集合 S 的元素经受某变换群 T 中所包含的变换时集合 S 保持不变的那些性质的研究，为方便起见, 这种几何学以符号 G(S,T) 表示。

也即任何一种几何学可以用公理化方法来构建，也可以把变换群和几何学联系起来。

例如集合 S 叫做空间，S 的元素叫做点，S 的子集 A 和 B 叫做图形，凡是等价的图形都属于同一类（图形等价类）。

同一类里的一切图形所具有的几何性质必是变换群 G 下的不变量，因而可用变换群来研究几何学（Erlangen 纲领），例如在正交变换群下保持几何性质不变的便是欧式几何，在仿射变换群下保持不变的便是仿射几何，在射影变换群下保持不变的便是射影几何，在微分同胚群下保持不变的便是微分几何。

稍微具体一点，平面欧几里得度量几何为设 S 为通常平面上所有点的集合, 考虑由平移、旋转和线上的反射组成的所有 S 的变换的集合 T。因为任何两个这样的变换的乘积和任何这样的变换的逆变换还是这样的变换, 所以,T 是一个变换群。长度、面积、全等、平行、垂直、图形的相似性, 点的共线性和线的共点性这样的一些性质在群 T 下是不变的, 而这些性质正是平面欧几里得度量几何所研究的。

仿射几何就是把平面欧几里得度量几何的变换群 T 扩大, 除了平移、旋转和线上的反射外, 再加上仿射变换（换句话说，就是从欧几里得空间的距离概念抽象化出单比的概念，就从欧式几何中舍弃距离不变而保留更普遍的单比不变，就从欧氏几何升级到仿射几何）。在此扩大的群下, 像长度面积和全等这类性质不再保持不变, 因而不再作为研究的课题。但平行垂直图形的相似性, 点的共线性, 线的共点性仍然是不变的性质, 因而仍然是这种几何中要研究的课题.

射影几何所研究的是平面上的点经受所谓射影变换时仍然保持不变的性质

从单比抽象到交比概念（换句话说，从仿射几何中舍弃单比不变而保留更普遍的交比不变，升级射影几何）。在前面讲的那些性质中, 点的共线性和线的共点性仍然保持不变, 因而是这种几何所要研究的课题

在上述的几何中, 使某变换群的变换起作用的基本元素是点, 因此, 上述几何均为点几何的例子。还有线几何, 圆几何, 球几何和其他几何的例子。

在建立一种几何时, 人们首先是不受拘束地选择几何的基本元素，其次是自由选择这些元素的空间或流形，自由选择作用于这些基本元素的变换群, 这样, 新几何的建立就成为相当简单的事了。也即从欧式空间（长度，夹角）到内积空间（模，不严格的夹角）再到赋范空间（范，完全抛弃夹角），不断的抽象，最后甚至连范数（最不愿抛弃的度量或度规）也抛弃了，从不严格的距离发展到不确定的距离，也即由欧式空间的连续函数抽象出度量空间的连续映射，一直到抽象出拓扑空间中的同胚映射，最后得到了拓扑空间的概念，这是人类目前为止在抽象上最深刻的极限。可以说克莱因用群论来研究几何学是人类思想的突破。

总之，群是数学中最有影响的概念，不了解群，就不可能了解现代数学。群论直接推动了代数数论、代数几何、函数论、微分方程与特殊函数论和代数拓扑的产生和发展，甚至很多经典数学领域，因为群论的引入而现代化。

其实数学上这种抽象过程，也推动了理论物理学的发展，例如狭义相对论发展就是要摆脱坐标而直接度量时空的过程，而广义相对论发展就是摆脱时空度量概念，走向空间同胚概念的过程。

目前群论已经是现代物理的主要工具。群论广泛用于基本粒子、核结构、原子结构和晶体结构等，因为对称性是物质世界最普遍的性质，例如各种物体（分子、晶体或图形）都可以用特定的对称性群来描述其结构（晶体的空间对称性可以用点群描述，其实晶体 X 射线衍射的图案直接与其点群相关）；再例如时空存在对称性，可以用彭加莱群描述不同表示对应不同自旋的粒子，例如标量粒子、旋量粒子、矢量粒子；再例如量子力学里的全同粒子就是对称性的（基本粒子的规范对称性可以由李群描述，其实李群的结构常数直接决定了规范玻色子，比如胶子、W、Z 玻色子的自相互作用）。

现代化学也离不开群论。例如化学中分子的性质受到分子对称性的影响（因

为分子的对称性反映出分子中原子核和电子云的分布情况），所以可以根据分子对称性判断该分子的一些基本性质，例如判断是否具有旋光性（判别分子是否具有旋光性的常用的方法是比较实物和它的镜像，看它们能否完全重合，凡不能和镜像重合的分子都具有旋光性；反之，如果两者能够重合，则分子就没有旋光性），所以可以用分子的对称元素和所属对称群来判断其是否具有旋光性。

同样，根据分子的对称性，也可以判断分子有无偶极距。分子偶极距大小决定于分子正负电重心间的距离与电荷量，其方向规定为从正至负。因为分子所具有的对称性是分子中原子核和电子云对称分布的反映，分子正负电重心一定处于分子的对称元素上。所以分子的永久偶极矩是分子的静态性质，静态性质的特点就是它在分子所属点群每一对称操作下必须保持不变，为此 μ 向量必须落在每一元素上，因此可以根据“分子对称元素是否只交于一点”来预测分子有无永久 μ。如果分子有对重心落在同一点上，因而无偶极距。若不存在上述的对称元素时，则分子的正负电重心不落在同一点上，就有偶极矩。

如果分子具有对称中心，那么分子的所有对称元素都交于此点，此点亦即分子正负电荷的重心。因此，具有对称中心的分子没有偶极矩。如果分子有两个对称元素交于一点，比如有一个对称面和垂直于此面的对称轴，或者有两个以上不相重合的对称轴，那么分子的正负电荷中心必重合于此交点，因而也没有偶极矩。分支虽有对称面和对称轴，但他们若不相较于一点，而且对称轴为对称面所包含，则他们具有偶极矩。

按照这一判据，可将分子所属点群和它是否具有偶极矩的关系总结为：对于具有偶极矩的分子可以进一步推断：当分子有 C2 轴时，偶极矩必沿着此轴；当分子有对称面时，偶极矩必位于此面上；当分子有几个对称面时则偶极矩必沿着他们的交线。

再例如化学位移等价性的判别质子或其他的原子核，在一定的交变磁场的作用下，由于分子中所处的化学环境不同，从而将在不同的共振磁场下显示吸收峰。这一现象就叫做化学位移。化学位移是核磁共振波普中反映化合物结构特征最重要的信息之一。

氢气 (H1) 谱亦即质子谱，在核磁共振波普中应用最为广泛。氢谱中的各个峰与分子中的不同环境的质子相对应。这样便可根据分子对称性识别等价院子

或基团，进而可以判别氢谱中化学位移的等价性。全同质子（通过旋转操作课互换的质子）在任何化学环境中都是化学位移等价的。对映异位质子（存在对称操作使分子中两个质子互换的质子）在非手性溶剂中具有相同的化学性质，也是化学位移等价的，但在光学活性或酶产生的手性环境中就不再是化学等价的，在核磁共振波普中可以显示偶合现象。此外，非对映异位质子（不能通过操作达到互换的质子）在任何化学环境中都是化学位移不等价的。分子中化学位移等价的核构成一个核组，相互作用的许多核组构成一个自旋系统。考虑分子的对称性，有利于对它们进行分类，因而群论就是最基础的。

群论也广泛用于分子结构判断，因为分子外形的对称性通过分子波函数与分子结构联系，而分子波函数可以作为分子所属点群的不可约表示的基。

杂化轨道理论主要是研究分子的几何构型，而构型和杂化的原子轨道在空间的分布和方向有密切的联系。由于在微观世界中，分子都具有一定的对称性，而对称性不同时，则其分子构型也必然不同，因此分子对称性就与其杂化轨道有内在的联系。群论的方法可以告诉我们：在具有一定形状的分子的化学成键中，中心原子可能采用什么样的杂化方式。运用群论的知识还可以知道中心原子提供哪些原子轨道去构成合乎对称性要求的杂化轨道，而且还可以进一步求出杂化轨道的数学表达式。

当然群论还有实际工程应用，例如先进陶瓷材料研发。我们都知道，先进陶瓷材料现在用途极为广泛，例如涡扇发动机用的陶瓷涂层材料，或陶瓷基复合叶片，甚至在尾喷管，燃烧室等等都开始使用陶瓷复合材料，以及在导弹某些关键部位的应用。

而大家不知道的是：群论在先进（陶瓷）材料的结构筛选中是基本工具。

因为晶粒对陶瓷的性能起着关键性的作用，所以研究晶粒是获得新材料性能的关键，例如由晶体的各向异性性，可以通过控制外界工艺条件使晶粒在某个晶向优先生长，从而可能具有某些前所未有的性能，在力学上使结构陶瓷得到更好的晶须增韧效果，在物理性能上或者在力学性质上增强，使功能陶瓷获得更好的韧性，刚性，抗切变性，或者是在电学性能上增强，例如获得更好的压电性能、热释电性、倍频效应，或者使人工晶体获得更好的旋光性等光学性能等等。

目前通常做法是通过对这些具有一定力学性能、物理性能的材料的微观本质的分析，可以利用对称群分析计算，筛选出掺杂物质和优化结构构造方式等等，来改变晶体的晶格，以获得性能更佳，物理效应更显著的晶体。

用对称群为工具也可以研究非晶态材料和非平衡态材料结构。非晶体与晶体相比有着大量的缺陷，原子或离子间的结合也不如晶体那般整齐有序，所以比同类晶体具有更大的内能，因此当非晶态向晶态转变或者反过来晶态向非晶态转变时将吸收或放出大量的能量，选择适当的材料显然在某些场合可以考虑由此而用来存储能量。

1.8 小结

本篇帖子由于豆瓣不能上数学公式，所以用一直奇特的模式写，痛苦不堪，可能错误比较多，发现错误请指出来。

再顺便感概一下，现在大学老师基本都是照本宣科，把教科书在课堂朗读一遍拉倒，纯粹是误人子弟。

我在中国科大数学系上学时，任意一门数学课的老师教课都是这个模式：任何一个重要概念的实际背景（包括但不限于工程，物理，军事等等问题），来龙去脉，要解决什么问题，结果解决什么问题，这些抽象概念的基本思想和原型是什么等等，都要让学生知其然，也知其所以然。

一个蒙查查的老师，一般不太可能教出什么明白学生。大学之间的水平差距，其实在老师之间的差距。上大学，如果不想被教成蒙查查，最好上最好的大学，不然人家勇猛奋进的四年光景，你不过是混日子的四年。

第二章　瞎扯贝叶斯理论的基本思想

基本信息

1. 原文链接 https://www.douban.com/group/topic/82509566/.
2. 本文作者 wxmang.

2.1　序

既然是瞎扯，就不是很严谨的，因为要简单明了，可能有的地方细节删除过多了，导致说不通。所以只能大概齐。真的要弄懂，还是请看教科书为好。写这种文章，华罗庚先生写得最好。

2.2　倒向问题

自从人类有自我意识，可能就在讨论一个至今没有结论的问题：机遇（或者运气，或者机会）到底是什么？怎么把握，怎么预测，怎么估计或计算大小等等。这是人类的一个核心问题。其实这也是一个典型的倒向问题。

人类思考问题有两个方向，一个是正向，也即知道结果找原因（例如现在我们经常讨论的明朝灭亡的原因）；一个是倒向，也即根据一些现象判断结果（例如如果我们事先并不知道黑箱里面黑白球的比例，而是闭着眼睛摸出一个或好

几个球，观察这些取出来的球的颜色之后，就此对黑箱里面的黑白球的比例进行推测。现实需要大量的倒向计算，例如现在某些现象出现，企业会不会破产，现在应该怎么办等等）。

用正向思维方式研究问题的，我们叫他们事后诸葛亮，历史学家便是。这种研究只能用于经验总结和知识储备上。

用倒向思维方式研究问题的，我们叫他们预测大师，见微知著的真人，管蠡窥测或以蠡测海的超人等等。实际上有价值的问题多数都是倒向问题，例如：股市上，通过那几点征兆就能判断是一次多或空的机会；医院中，通过那几个症状就能判断是什么病；科学研究上，通过几个实验数据，就能构造什么理论解释模型等等。一般说来，数学家，物理学家等等都是研究倒向问题的，或者说，他们不能通过很少征兆或现象来预测或判断结果，就没存在价值（顺便说一句，不知道倒向问题思维方式的人，没法再金融市场或股市搏击，目前投机市场最前沿的研究几乎就是倒向随机过程和鞅论为主，中国的彭实戈院士是倒向随机过程领域的其中一个领军人物）。

如何用倒向思维方式研究机遇？也即如何从一些征兆或现象，判断机遇，或者把问题进行等价推广：如何用一些已知的信息或经验，判断或预测未知。

1763 年，英国的长老会牧师贝叶斯发表了一篇论文“论有关机遇问题的求解”，提出了解决的框架：那就是用不断增加的信息和经验，可以逐步逼近未知的真相或理解未知。并给出了算法（其实贝叶斯由于是一个牧师，他关心的原始问题本来的表述是：人能不能根据凡人世界的经验和现实世界的证据，证明上帝的存在，因为宗教人士的逻辑是机遇就是上帝存在的主要证据，能够认识机遇的规律，几乎等同于证明上帝存在）。

后来经拉格朗日等数学家进一步努力，获得了大突破，贝叶斯理论成为现代统计学两大支柱之一。

由于我们不讨论数学，所以不进一步讨论贝叶斯思想的各种复杂数学表达，我们只讨论其基本思想。

2.3　贝叶斯基本思想

下面我们用一个例子来介绍贝叶斯思想。

假定甲乙两人做一个游戏：甲蒙上眼睛，乙随手在一张纸上画一条线段，两个端点分别 A 点和 B 点，再随机在线段上画一点 C，现在游戏是：通过一些信息，让甲判断 C 点位置。

信息 1：乙随机在线段上划点，并告诉每一个点是离 A 端点近还是离 B 端点近；

信息 2：乙必须告诉每一个点是位于 AC 之间还是 CB 之间。

显然只需要有限次划点，甲就能基本确定 C 的大致位置。

为简单说明，我们假设乙划点总是在中位点，也即第一个点是 A+B 的中点，根据上面游戏规则，必须告诉甲的两个信息，甲这样就能判断 C 靠 A 点近还是 B 点近；如果 C 靠 A 点近，那么第二个点划在 $A+C$ 的中点，同样重复上述步骤，甲就能判断 C 点是靠 A 点近还是 $\frac{(A+C)}{2}$ 近，继续循环。我们知道中位点划分会形成一个 $\frac{1}{2^N}$ 的收敛级数，可以在有限的 N 次内，就能获得 C 点位置范围，误差不大于 $\frac{1}{2^N}$。

这就是用逐步增加的信息确定未知的例子。这也就是贝叶斯思想的基本模型。

当然贝叶斯考虑的逐步增加信息是人的经验和知识构成的先验信息，而先验信息会因为个人的经验偏见，视野局限，测量误差，外部环境变化，数据丢失等等导致一定的不准确，也即只有概率意义上的正确。所以他能够得到的对未知的判断也只能是概率意义上的。

简单总结一下贝叶斯的基本观点是：

（1）、由于经验或知识是否正确带有不确定性因素，所以基于以前经验和知识（也即先验），根据一些随机出现或观察到的现象来判断事物真相，原因或未知，就有一定的不确定性。（其实这个观点有一个强大的前提假设：未知世界本质是随机的，所以任何未知都具有不确定性，这是现代量子力学和复杂系统证明了的假设）。

（2）、由于我们依靠已知的经验或知识, 来分析观察到的现象，以此推测或断未知，所以任何未知的判断或推测都是是不确定的，能用一个概率分布去描述，也即未知的不确定性程度由先验概率分布和现象出现的概率分布决定。

上面这句话的本质意思是：是不是能够通过一些现象正确判断或推测未知，取决于我们经验多少和掌握的现象多少。例如，医生能否通过症状判断疾病，取决于医生知识（这也是一种经验，是从前人身上学习的）和经验积累，也取决于掌握的症状多少（掌握症状一般通过各种检查实现，例如为了掌握症状而进行的量体温，照 X 光，核磁共振，CT 检查，超声波检查等等）。

但是光掌握症状并不能完全判断病症（否则就不需要医生，只需要检测工程师了），还需要医生知识和经验。

因为任何检查都不可能完备，再加上任何医生的经验和知识也是不完备的（例如对一种新疾病原有的经验和知识就无能为力），所以任何判断都有一定的对错概率。

（3）、由于基于 N 个现象（或症状）和个人主观经验来判断未知带有一定不确定，我们往往需要做多次检测和由具有不同个人主观经验的人来判断，然后按照极大似然原则选择结果（例如医院诊断重大疾病往往要请不同医生会诊，也要进行不同系列，不同类型的医疗检查）。

（4）、先验信息可以通过的收集、挖掘和加工而数量化，形成先验分布（也即所谓专家知识库可以提高判断精度）。

上面这些观点，综合起来就是：

现实世界本身是不确定的，人类的观察能力是有局限性的（例如如果人类能够直接观察到电子的运行，还需要假设什么模型），人类所观察到的只是事物表面上的结果（若干症状或现象），例如往往只能知道从黑箱里面取出来的球是什么颜色，而并不能直接看到黑箱里面实际的情况。

贝叶斯思想给我们提供了一个猜测黑箱里面情况的方法。当然这种方法得到的结果是不确定的（因为世界本质就是不确定的，而且这种方法依赖于人的主观经验，本身是否正确也是不确定的）。

贝叶斯的方法其实是一个算法：第一步，算出各种不同猜测的可能性大小；第二步算出最可能的猜测是什么。

第一步就是计算特定猜测的后验概率（对于连续的猜测空间就是计算猜测的概率密度函数），第二步则是极大似然方法。

定义 2.1: 极大似然的定义是

事件 A 与参数 $\theta \in \Theta$ 有关，θ 取值不同，则 $P(A)$ 也不同，若 A 发生了，则认为此时的 θ 值就是 Θ 的估计值。这就是极大似然。

例如两人一起打猎，只响一枪，就打中一猎物，那么按正常逻辑，一枪命中，肯定是枪法好的那个，这个推断就体现了极大似然法的基本思想。

再例如，袋中装有许多黑、白球，不同颜色球的数量比为 3:1，试设计一种方法，估计任取一球为黑球的概率 P。

显然 P 的值无非是 $\frac{1}{4}$ 或 $\frac{3}{4}$，需要通过抽样来决定分布中参数究竟是 $\frac{1}{4}$ 还是 $\frac{3}{4}$。现从袋中有放回地任取 3 只球，显然白球出现次数多的话，P 就是 $\frac{1}{4}$，黑球出现次数多的话，P 就是 $\frac{3}{4}$)。

也即贝叶斯思想就是：对于给定观测数据，一个猜测是否正确，取决于这个猜测本身独立的可能性大小（先验概率）和这个猜测生成我们观测到的数据的可能性大小（似然）的乘积。（也即最可能的猜测 = 先验概率 * 似然最大的这个值）。

2.4 贝叶斯公式

先用一个 wikipedia 上的例子介绍贝叶斯公式的原理。

假设一所学校里面有 60% 的男生，40% 的女生。男生总是穿长裤，女生则一半穿长裤一半穿裙子。问题：随机在校园乱走，看到在校园里穿长裤的人里面有多少女生的概率？

假设学校里面人的总数是 N 个。60% 的男生都穿长裤，于是我们得到了 $N \cdot P(\text{Boy})P(\text{Pants}|\text{Boy})$ 个穿长裤的（男生）（其中 $P(\text{Boy})$是男生的概率 $= 60\%$，这里可以简单的理解为男生的比例；$P(\text{Pants}|\text{Boy})$ 是条件概率，即在 Boy 这个条件下穿长裤的概率是多大，这里是 100%，因为所有男生都穿长裤）。40% 的女生

里面又有一半（50%）是穿长裤的，于是我们又得到了 $N \cdot P(\text{Girl})P(\text{Pants}|\text{Girl})$ 个穿长裤的（女生）。加起来一共是 $N \cdot P(\text{Boy})P(\text{Pants}|\text{Boy}) + N \cdot P(\text{Girl})P(\text{Pants}|\text{Girl})$ 个穿长裤的。

整理一下：

$$P(\text{Girl}|\text{Pants})(\text{穿长裤的人里面有多少女生的概率}) = \frac{N \cdot P(\text{Girl})P(\text{Pants}|\text{Girl})}{[N \cdot P(\text{Boy})P(\text{Pants}|\text{Boy}) + N \cdot P(\text{Girl})P(\text{Pants}|\text{Girl})]}$$

很容易发现公式与校园内人的总数是无关的，可以消去。于是得到

$$P(\text{Girl}|\text{Pants}) = \frac{P(\text{Girl})P(\text{Pants}|\text{Girl})}{[P(\text{Boy})P(\text{Pants}|\text{Boy}) + P(\text{Girl})P(\text{Pants}|\text{Girl})]}$$

上式中的 Pants 和 Boy/Girl 可以指代一切东西，所以其一般形式就是：

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\ B)P(\ B)}$$

即已知 $P(A|B)$、$P(A)$ 和 $P(B)$，可以计算出 $P(B|A)$。

这就是著名的贝叶斯公式。所以拉普拉斯说概率论只是把常识用数学公式表达了出来。

2.5 贝叶斯思想在决策中

当然贝叶斯思想用的最多的领域是决策。贝叶斯决策就是在不完全情报下，对部分未知的状态用主观概率估计，然后用贝叶斯公式对发生概率进行修正，最后再利用期望值和修正概率做出最优决策。

贝叶斯决策属于风险型决策，决策者虽不能控制客观因素的变化，但却掌握其变化的可能状况及各状况的分布概率，并利用期望值即未来可能出现的平均状况作为决策准则。

贝叶斯思想能够用于决策的原因是：除样本（例如决策调查获得的一些事件或资料）提供的基础信息外，人类的经验的先验信息也是决策判断的重要依据。

例如以对神童出现的概率 p 的估计为例。按经典统计的做法，完全由样本提供的信息 (即基础抽样信息) 来估计，认为参数 p 是一个“值”。而人类经验其实对 p 已经有了一个了解，如 p 可能取 p_l 与 p_2，且取 p_1 的机会很大，取 p_2 机会很小。

先验信息关于参数 p 的信息是一个“分布”，如 $P(p=p_1)=0.9, P(p=p_2)=0.1$，即在抽样之前已知道 (先验的)$p$ 取 p_1 的可能性为 0.9。若不去抽样便要作出推断，自然会取 $p=p_1$。但若抽样后，除非采样信息 (即样本提供的信息) 包含十分有利于“$p=p_2$”的支持论据，否则采纳先验的看法“$p=p_1$”。

下面用一个例子解释这种决策。

假设患者有肺炎为事件 B，$P(B)$ 是医生在检测前，基于经验对当时当地的肺炎肺病率的概率判断，例如 1%，也即如果没有任何检查，医生基于经验，会判断得肺炎概率很小，$P(\ B)$ 是不是肺炎的概率，例如是 99%；

$P(A)$ 是全体人群（包括肺炎和不是肺炎人群）检查出现各种指标异常的概率，这可以是一个函数曲线，自变量是各种检测指标，总和等于 1（可以经过规范整理得到）；

$P(A|B)$ 是当病人是肺炎时，各种检测指标出现异常的概率，是医院经过大量临床试验和检测数据分析得到的；

$P(A|\ B)$ 是当病人不是肺炎时，各种检测指标出现异常的概率，是医院经过大量临床试验和检测数据分析得到的；

那么根据贝叶斯公式，就可以计算出 $P(B|A)$，也即当某一个异常指标出现时，确诊是肺炎的概率。

例如假设我们知道检测肺炎有 12 个指标，根据多年来的经验和大量临床数据，如果是肺炎，可以得到这 12 个检测指标的出现异常的概率分布情况，也即知道 $P(A_i|B)$；对不是肺炎的，这 12 个指标的异常概率分布情况我们也知道，也即 $P(A_i|\ B)$。

在疹疗过程中，医生要根据临床经验对各种病症状 Ai 进行权衡。当然由于经验数据误差和知识不完备，存在误诊概率。

举个例子来说明，假设有一台肺炎诊断仪，通过对它以往的诊断记录的分析，如果患者确实患有肺炎它的确诊率为 90%，若果患者没有癌肺炎，被诊断成肺炎的概率为 10%。

那么如果一个人被这台诊断仪确诊成肺炎（这是现象），这个人患有肺炎的概率是多少（这是判断）?

设 A：肺炎诊断仪给出肺炎诊断。B1：病人是肺炎患者。B2 病人不是肺炎

患者。

根据贝叶斯公式：$P(A|B1) = 90\%$；$P(A) = 90\% \cdot P(B1) + 10\% \cdot P(B2)$；则 $P(B1|A) = P(B1) \cdot 90\%/(90\% \cdot P(B1) + 10\% \cdot P(B2))$；

我们知道当时当地人群中肺炎患者的比重很小，假设为 1%（这就是经验，有一定的主观性），则 $P(B1) = 1\%$；$P(B2) = 99\%$；

可以算出：$P(B1|A) = 8\%$，也即一个人被这台诊断仪确诊成肺炎，而这个人真正患有肺炎的概率只有 8%。显然这是一个不能饶恕的结果。

但是实际上医生经常这么做，只是他们不是用肺炎诊断仪，而是用一些简单诊断手段就随便下结论了。

所以医院需要加入更多的检查项目，增加更多的检测设备，加入更多医生判断的判断，通过极大似然原则，筛选最可能的结果，逐步逼近真相，解决掉判断的错误风险。

2.6 先验和后验

先验就是指在抽样前就有的经验信息的概率表述，先验不必有客观的依据，它可以部分地或完全地基于主观信念。

再以肺炎为例子，某人认为自己得肺炎，去看病，医生在没有检测之前，基于经验，认为本地在此时此地，得肺炎情况很小，认为他极可能只是感冒，这就是先验。

当然医生不会完全相信自己先验，必然会给某人进行完整的检查，例如测体温，量血压，甚至照 X 光等等，这些检测结果数据就构成检测样本 X。

医生诊断时, 必然不会只使用 X（检测数据）提供的信息进行诊断，而必须先验（否则就不需要医生，只需要工程师就能治病了）。

假定 $\theta = 0$ 是没病，$\theta = 1$ 是有病，贝叶斯理论认为 X 的分布取决于 θ 是 θ 还是 1, 同时根据贝叶斯公式，我们知道了 X，就有助于推断 θ 是否为 1。

通过使用贝叶斯公式计算后验分布，也即根据抽样样本（检测数据）X 的分布 p（x）及 θ 的先验分布 p(θ)（当时当地某种疾病的分布的先验概率，由医

生经验和知识积累决定），用贝叶斯公式就可计算出在已知 X=x 的条件下,θ 的条件分布 p(θ|x)（这就是后验概率，也即经取样和先验概率计算后的出来的）。

显然这个分布综合了样本 X（检测数据）及先验分布 p(θ)(医生经验）所提供的有关的信息。假定设 p(θ=1)=0.001, 经计算，p(θ=1|x)=0.86, 则含义为：在某人的检测指标量出之前, 根据医生经验（p(θ=1)），他患病的可能性定为 0.001，而在得到 X 后, 认识发生了变化：其患病的可能性提高为 0.86。这一判断既与 X 有关，但也也离不开先验分布。也即若当时当地肺炎的发病率很小, 医生将倾向于只有在样本 X 显示出很强的证据时, 才诊断某人有肺炎，这样就可以避免误诊。

2.7　奥卡姆剃刀对决策的筛选

贝叶斯思想本质上是一个经验归纳推理的计算公式，解决了逐步逼近真实的过程，但是显然，这种逼近有不确定性，因为即便一个判断与经验数据非常符合，也并不代表这个判断就是正确的判断，因为经验数据总是会有各种各样的误差，比如观测误差，记忆误差等等，再加上有时无法分别背景噪音，把背景噪声纳入经验数据，来加以判断（背景噪声可以看成与事件完全无关的因素）。

这时，我们就必须应用奥卡姆剃刀：如果两个理论具有相似的解释力度，那么优先选择那个更简单的（往往也正是更平凡的，更少繁复的，更常见的）。

所以现实中，我们建立解释问题的模型往往只提取出几个与结果相关度很高，很重要的因素，而不会面面俱到，太复杂的模型不但成本高，而且会因为无法辨别背景噪声因素，导致失真。

例如我们知道现在各种计算健康体重的模型都是用身高和体重近似于一个二阶多项式的关系（对人群随机抽取了 N 个样本，用最小二乘法拟合出一个二阶多项式），但大家都知道并不是只有身高才会对体重产生影响。但是总体上来说，绝大多数人的身高与体重有关，成正态分布，这个分布就保证了身高体重相关模型能够在大多数情况下做出靠谱的预测。当然人有胖瘦，密度也有大小，所以完美符合身高体重的二阶多项式关系的人是不存在的，只是近似。这就是人类对事物理解的模式。

2.8 简单总结

根据上述描述，我们可以得出几个结论：

（1）、当假定世界本质是随机的，那么我们认识世界的结论也是不确定的，只能通过不断累积的经验去逼近；

（2）、所以我们认识世界本质基于经验积累；

（3）、我们判断事物是什么的准确概率，往往基于经验积累程度多少。或者说，我们经验积累，能够使我们逐步减少判断事件错误的概率。

所以贝叶斯对先验概率的指定既是主观的，又是理性的，而且随经验积累逐渐优化。

从哲学角度看，贝叶斯思想是一种逻辑 + 历史的方法，是归纳推理方法的一次革命（归纳推理就是根据过去的经验预测未来的推理），把经验数据量化，并直接带入预测判断。解决了休谟对归纳推理的合理性提出的质疑。

现在贝叶斯的思想已经成为归纳逻辑的核心, 并且逐步发展为一套一般性的科学推理理论和方法。贝叶斯思想现在是机器学习[1]的核心方法之一。

不用数学公式介绍数学思想，真的很费力，也不知道讲清楚没有。

[1]比较好的书籍参见, 周志华著. 机器学习, 北京: 清华大学出版社,2016. 南京大学周志华研究组链接 http://cs.nju.edu.cn/zhouzh/

第三章　瞎扯数学分析 1、微积分

基本信息

1. 原文链接 https://www.douban.com/group/topic/96542437/.
2. 本文作者 wxmang.

3.1　序

这一篇帖子主要介绍人类如何从一个基于几何直观或直觉的计算技巧或计算方法，进化到逻辑基础严密的公理体系的例子，想说明人类抽象的另外一个方向：语言抽象（结构抽象已经在介绍伽罗华群论时介绍过）。

为了让非数学专业的人能够看下去，采用了大量描述性语言，所以严谨是谈不上的，只能算瞎扯。

现代数学基础有三大分支：分析，代数和几何。这篇帖子以尽量通俗的白话介绍数学分析。数学分析是现代数学的第一座高峰。

最后为了说明在数学中，证明解的存在性比如何计算解本身要重要得多，用了两个理论经济学中著名的存在性定理（阿罗的一般均衡存在性定理和阿罗的公平不可能存在定理）为例子来说明数学家认识世界和理解问题的思维方式，以及存在性的重要性：阿罗的一般均衡存在性，奠定了整个微观经济学的逻辑基础–微观经济学因此成为科学而不是幻想或民科；阿罗的公平不可能存在定理，

摧毁了西方经济学界上百年努力发展，并是整个应用经济学三大支柱之一的福利经济学的逻辑基础，使其一切理论成果和政策结论成为泡影。

3.2 微积分

数学分析是微积分基础上发展起来的，所以先说说微积分。

微积分的基本思想是以直为曲，也即用直线来逼近曲线，在中国古代，刘徽，祖冲之计算圆周率用的割圆术就是典型的微积分方法，三国时期的刘徽在他的割圆术中提到的“割之弥细，所失弥小，割之又割，以至于不可割，则与圆周和体而无所失矣。”魏晋南北朝时期的祖冲之说的更简单：以曲为直逼近。在古代巴比伦，希腊都用这种方法来处理曲线计算问题，有史可查的记录是公元前三世纪，古希腊的阿基米德计算抛物弓形的面积、球和球冠面积、螺线下面积和旋转双曲体的体积时，就用了直线逼近。

所以在牛顿（Newton）和莱布尼茨（Leibniz）发明微积分之前，很多实际上的微积分的工具已经开始运用在科学和工程之中。例如法国的费尔玛、笛卡尔、罗伯瓦、笛沙格；英国的巴罗、瓦里士；德国的开普勒；意大利的卡瓦列利等人都用这种以直为曲的逼近方法计算工程问题。

但是微积分为什么说是十七世纪牛顿和莱布尼茨发明的呢，我觉得主要是两点：第一点是引入了函数概念来描绘变量；第二点是发明了一套符号体系，可以计算各种初等函数微分（初等函数简单说就是多项式函数、幂函数、指数函数、对数函数、三角函数、反三角函数，以及由这些函数经过有限次四则运算或函数的复合而得的所有函数）。

牛顿和莱布尼茨发明的最原始的微积分可以解决以下问题：

求即时速度的问题；求曲线的切线；求函数的最大值和最小值；求曲线长、曲线围成的面积、曲面围成的体积、物体的重心、一个体积相当大的物体作用于另一物体上的引力等等。

牛顿和莱布尼兹最本质的贡献是把求切线问题（微分学的中心问题）和求积问题（积分学的中心问题）变成一个问题。这就是著名的牛顿–莱布尼兹公式。

牛顿和莱布尼茨建立微积分的基本思想是以曲为直，逐步逼近，其中创造是引入了无穷小量 Δ，因此微积分也称为无穷小分析。

不过他们两个有区别，牛顿从运动角度入手，莱布尼茨从几何角度路入手。

牛顿在 1671 年写了《流数法和无穷级数》，这本书直到 1736 年才出版，它在这本书里指出，变量是由点、线、面的连续运动产生的，否定了以前自己认为的变量是无穷小元素的静止集合。他把连续变量叫做流动量，把这些流动量的导数叫做流数。牛顿在流数术中所提出的中心问题是：已知连续运动的路径，求给定时刻的速度（微分法）；已知运动的速度求给定时间内经过的路程（积分法）。

莱布尼茨 1684 年发表世界上最早的微积分文章：《一种求极大极小和切线的新方法，它也适用于分式和无理量，以及这种新方法的奇妙类型的计算》，创立了现代的微分符号和基本微分法则（远远优于牛顿的符号，现在使用的微积分通用符号就是当时莱布尼茨创造的），1686 年，莱布尼茨发表了人类第一篇积分学的文章。

微积分的创立，极大地推动了数学的发展，过去很多初等数学束手无策的问题，运用微积分，往往迎刃而解。例如牛顿应用微积分及微分方程从万有引力定律推导出了开普勒行星运动三定律。

微积分也极大的推动天文学、力学、物理学、化学、生物学、工程学等的发展。

由于争抢微积分发明权，欧洲大陆的数学家和英国数学家的长期对立，英国数学陷入牛顿的“流数术”中停步不前，英国数学后来比欧洲整整落后了一百年。

虽然原始微积分是一种强大计算工具，但是从逻辑上讲，牛顿和莱布尼茨的工作都是很不完善的，他们为了计算微分，引入的在无穷和无穷小量概念，其实没有说清楚是个什么东西，例如牛顿的无穷小量，有时候是零，有时候不是零而是有限的小量；莱布尼茨干脆回避解释。无穷小的逻辑基础存在的问题导致了第二次数学危机的产生（这个在介绍现代数学基础的帖子里已经介绍了，不重复）。

19 世纪初，法国的柯西对微积分的理论进行了认真研究，建立了极限理论，

后来德国的魏尔斯特拉斯进一步的严格化，使极限理论成为了微积分的坚定基础。才使微积分在逻辑上站住脚，而不仅仅是一种计算工具。

微积分的基础概念是函数和极限。前者是微积分的工作对象，后者是微积分的基本工作技巧。

3.2.1 函数

函数概念是人类一个很伟大的发现，价值不下于对于数的发现，也是高度抽象的产物。

不过函数的思想却很早，至少在公元前就有了：因果关系，也即有因必有果，一个因对应一个或多个果，或者一个果对应多个因。

这在中国《易经》中已经有成熟的体现（其实《易经》就是 64 变量的函数论），正因为有了这种因果关系概念，中国远古时代我们先人就有了成熟精妙的辩证法（比黑格尔的辩证法高级多了，精细多了）。西方辩证法也是在有了成熟的函数概念后才成熟的。恩格斯就说过：“数学中的转折点是笛卡儿的变数，有了变数，运动进入了数学；有了变数，辩证法进入了数学”。

不过近代函数概念直接来源于代数方程中对不定方程的求解。

笛卡儿在 1637 年出版的《几何学》中，引入了现代函数的思想。英国人格雷果里在 1667 年论文《论圆和双曲线的求积》给出了函数的定义：从一些其他量经过一系列代数运算或任何其他可以想象的运算而得到的一个量。这里的运算指的是加减乘除开方五种代数运算以及求极限运算。

不过现在我们看到的函数定义来自于德国人莱布尼兹，他在 1673 年论文中，把任何一个随着曲线上的点变动而变动的几何量，如切线、法线、点的纵坐标都称为函数；并且强调这条曲线是由一个方程式给出的。直接定义了：函数表示依赖于一个变量的量。

紧接着函数概念被不断改进，第一个重要改进是瑞士人约翰. 伯努利于 1698 年给出的：由变量和常量用任何方式构成的量都可以叫做的函数。这里的任何方式包括了代数式和超越式。

第二个重要改进是 1748 年欧拉在《无穷小分析引论》中给出的函数定义：

变量的函数是一个解析表达式，它是由这个变量和一些常量以任何方式组成的。现代函数的符号就是欧拉发明的。欧拉还区分了显函数和隐函数、单值函数和多值函数、一元函数和多元函数等。

1775 年，欧拉在《微分学》一书中，给出了函数的另一定义：如果某些变量，以这样一种方式依赖于另一些变量，即当后者变化时，前者也随之变化，则称前面的变量为后面变量的函数。这个定义，为辩证法数学化打开了大门。

第三次重要改进是从函数的几何特性开始的，是 1746 年达朗贝尔给出的，把曲线称为函数（因为解析表达式在几何上表示为曲线）。但是后来欧拉发现有些曲线不一定是由单个解析式给出的，因此提出了一个新的定义：平面上随手画出来的曲线所表示的 x 与 y 的关系。即把函数定义为由单个解析式表达出的连续函数，也包括由若干个解析式表达出的不连续函数（不连续函数的名称是由欧拉提出的）。

在整个十八世纪，函数定义本质就是一个解析表达式（有限或无限）。

第四次最重要的改进是 1821 年柯西在《解析教程》中，给出了如下函数定义：在某些变量间存在着一定的关系，当一经给定其中某一变量的值，其他变量的值也随之确定，则将最初的变量称为自变量，其他各个变量称为函数。这个定义把函数概念与曲线、连续、解析式等纠缠不清的关系给予了澄清，也避免了数学意义欠严格的变化一词。函数是用一个式子或多个式子表示，甚至是否通过式子表示都无关要紧。

不过函数精确定义是德国人狄利克里于 1837 年给出的：若对 x（$a\le x\le b$）的每一个值，y 总有完全确定的值与之对应，不管建立起这种对应的法则的方式如何，都称 y 是 x 的函数。这一定义彻底地抛弃了前面一些定义中解析式的束缚，强调和突出函数概念的本质，即对应思想。

对应思想是人类伟大的发现，后来的映射，同构，同态等等概念来源于此，这是这个概念最伟大的地方。

当然我们知道狄利克里伟大，主要不是他给出函数的科学定义，而是他给出了著名的狄利克里函数，这个函数是难以用简单的包含自变量 x 的解析式表达的，但按照上述定义的确是一个函数。

为使函数概念适用范围更加广泛，人们对函数定义作了如下补充：“函数

$y = f(x)$ 的自变量，可以不必取 $[a, b]$ 中的一切值，而可以仅取其任一部分”，换句话说就是 x 的取值可以是任意数集，这个集合中可以有有限个数、也可以有无限多个数，可以是连续的、也可以是离散的。这样就使函数成了一个非常广泛的概念。但是，自变量及函数仍然仅限于数的范围，而且也没有意识到“函数”应当指对应法则本身。

最后，我们要说说现代数学理解的函数（来自于美国人维布伦）：设集合 X, Y，如果 X 中每一个元素 x 都有 Y 中唯一确定的元素 y 与之对应，那么我们就把此对应叫做从集合 X 到集合 Y 的映射，记作 $f : X \to Y, y = f(x)$。

不过从布尔巴基以后，基于数学结构的函数概念更进一步抽象，从函数、映射进化到关系：

1939 年布尔巴基用集合之间的关系定义了函数：设 E 和 F 是两个集合，E 中的每一个元素 x 和 F 中的每一个元素 y 之间的一个关系 f 称为函数，如果对每一个 $x \in E$，都存在唯一的 $y \in F$，它们满足给定的关系。记作 $f : E \to F$。在布尔巴基的定义中，E 和 F 不一定是数的集合，函数是集合之间的一个关系。也即设集合 E 和 F，定义 E 与 F 的积集 $E*F$ 如下：$E*F = \{(x, y)|x \in E, y \in Y\}$。积集 $E * F$ 中的一个子集 f 称为 E 与 F 的一个关系，若 $(x, y) \in f$，则称 x 与 y 有关系 f，记为 xfy, 若 (x, y) 不属于 f，则称 x 与 y 无关系 f。设 f 是 x 与 y 的关系，即 $f \in X * Y$，如果 $(x, y) \in f$,$(x, z) \in f$，必有 $y = z$，那么称 f 为 X 到 Y 的映射或函数。

这个定义回避了对应这种模糊不清的描述语言，而且把函数从单纯的数的概念推广到一切对象，例如结构，图像，集合等等。

不过微积分要处理的函数概念还是原始的，甚至只能处理初等函数。特点就是函数自变量的变化范围是数域，也即函数定义域与因变量的变化范围值域都是数域。这就是微积分的工作对象。这个对象可以描述一部分基于初等函数规律描述的变量跟结果的因果关系，通过对这种因果关系的分析和计算，人类就能预测或控制符合相应初等函数规律描述的事件或事物的因果关系，例如各种工程设备，武器系统等等，就能建立工业文明。

3.2.2　极限

极限是微积分的主要工作技巧。整个数学分析就是建立在极限概念上（包括级数）来处理初等函数因果关系的一门学科。

极限技巧一般是：对无法把握的连续变量，用可以计算的序列（例如数列，时间序列，多项式序列等等）逐步逼近变量，并能够证明这些序列可以无限逼近所求的未知量，然后计算这个序列的极限就可得到变量。

极限思想是微积分的基本思想，函数的连续性，导数以及定积分等等都是借助于极限来定义的。

所以可以说：数学分析就是用极限思想来研究函数的一门学科。

极限的思想在刘徽割圆术就有了，但是仅仅是一种计算方法，而不是一个思维方式。真正的现代极限思想来自于 16 世纪荷兰人斯泰文计算三角形重心过程中，用逐步逼近方式逼近重心。

牛顿和莱布尼茨最早并不是用极限思想来建立微积分的，他们的概念基础是无穷小，但是由于无穷小是个逻辑上有瑕疵的概念，导致微积分的逻辑基础无法自洽。例如牛顿用路程的改变量 ΔS 与时间的改变量 Δt 之比 $\dfrac{\Delta S}{\Delta t}$ 表示运动物体的平均速度，让 Δt 无穷小，得到物体的瞬时速度，并由此引出导数概念和微分，他并没有极限概念，他说：“两个量和量之比，如果在有限时间内不断趋于相等，且在这一时间终止前互相靠近，使得其差小于任意给定的差，则最终就成为相等”。这是一种几何直观而不是逻辑，就像小孩在纸上顺便划一下圆，就说是太阳。所以牛顿说不清楚他理解的无穷小到底是是什么。其实牛顿的说法如果用极限概念，很容易在逻辑上说清楚：如果当变量（例如时间 t) 无限增大或变量的差无限接近 0 时（$\Delta t \to 0$)，则 $\dfrac{\Delta S}{\Delta t}$ 无限地接近于常数 A，那么就说 $\dfrac{\Delta S}{\Delta t}$ 以 A 为极限，这个极限就是 s（路径函数）在 t_0 时的导数。

不过上述无限的概念仍然是几何直观的，并没有用逻辑描述出无限这个过程是什么，也没有定量地给出 ΔS 和 Δt 两个无限过程之间的数量联系，所以在逻辑上仍然有漏洞。

所以牛顿和莱布尼兹的微积分不断收到怀疑和攻击，例如最常见的质疑是贝克莱大主教的：在瞬时速度概念中，究竟 Δt 是否等于零？如果说是零，怎么

能用它去作除法呢？如果它不是零，又怎么能把包含着它的那些项去掉呢？这就是数学史上所说的无穷小悖论。

牛顿由于没有极限概念，无法回答这种质疑，只能混战。主要原因是微积分起源于人类计算需要从常量扩展到变量，但是牛顿采用处理常量的传统思想来处理变量。

18 世纪，罗宾斯、达朗贝尔与罗依里埃等人明确表示极限是微积分严格化的基础。其中最接近现代定义的是达朗贝尔的极限定义：一个量是另一个量的极限，假如第二个量比任意给定的值更为接近第一个量。但是这些定义都无法摆脱对几何直观的依赖。例如什么叫“接近”，逻辑上的含义是什么，其实还是几何直观。

现代极限概念来自于柯西，19 世纪，柯西出版的《分析教程》定义：当一个变量逐次所取的值无限趋于一个定值，最终使变量的值和该定值之差要多小就多小，这个定值就叫做所有其他值的极限值，特别地，当一个变量的数值（绝对值）无限地减小使之收敛到极限 0，就说这个变量成为无穷小。

柯西把无穷小视为以 0 为极限的变量，也即无穷小不是似零非零，无穷小非零，只是其极限为零。

魏尔斯特拉斯把柯西的语言翻译成 $\epsilon-\delta$ 语言，给微积分提供了严格的理论基础。所谓 $\lim_{n\to\infty} a_n$，是指：如果对任何 $\epsilon>0$，总存在自然数 ，使得当 $>$ 时，不等式 $|a_n-A|<\epsilon$ 恒成立。

这个定义，借助不等式而不是几何直观，通过 ϵ 和 之间的关系，定量刻划了两个无限过程之间的联系。这个定义中，涉及到的仅仅是数及其大小关系，此外只是给定、存在、任取等词语，已经摆脱了“趋近”一词，不再求助于运动的直观。

这个定义，本质揭示了无限与有限有本质的不同：无限个数的和不是一般的代数和，它是部分和的极限，是动态过程，而非静态计算结果。举例来讲，用任何静态计算，都无法计算出变速直线运动的瞬时速度，因为速度是变量。这其实就是量变和质变的一个例子：量变能引起质变。例如对任何一个圆内接正多边形来说，当它边数加倍后，得到的还是内接正多边形，是量变而不是质变；但是，不断地让边数加倍，经过无限过程之后，多边形就变成圆，多边形面积便

转化为圆面积，这就是量变到质变，这就是极限概念的本质。极限是区分初等数学和高等数学的分界线，初等数学处理静态问题，高等数学可以处理非静态问题了，例如求瞬时速度、曲线弧长、曲边形面积、曲面体体积等问题。

极限概念中，最重要的定理，非魏尔斯特拉斯的多项式逼近连续函数定理莫属，这个定理的简单表述是：闭区间上的连续函数可由多项式一致逼近。

这个定理意味着任何连续函数，都能构造一个多项式函数来逼近它，而多项式函数的导数，微分，积分的计算，简单易行，也即这个定理解决了连续函数的近似计算的逻辑基础问题：存在性。

这个定理最著名的证明是苏联数学家伯恩斯坦构造的著名的伯恩斯坦多项式，这个方法开启了函数构造法这一研究领域（当然对周期性的函数，还可以用三角级数，也即傅利叶级数逼近）。用多项式函数或三角级数逼近连续函数，是现代工程解决问题的主要方法，例如通信领域，如果不懂傅利叶级数，基本寸步难行，在流体力学、结构力学和弹性力学领域，不用多项式函数逼近，也基本无法计算海量的变量函数。函数构造方法其实是计算数学算法的基础（伯恩斯坦多项式符号太多，无法介绍，有兴趣可以上网搜索：伯恩斯坦多项式即可，有魏尔斯特拉斯定理用伯恩斯坦多项式证明的全过程）。

魏尔斯特拉斯本人最初的证明，是使用的核函数（正态核），并将核函数展开成一致收敛的幂级数，截取前面有限部分就构造出了逼近多项式。现在教材上选取的核函数是 Landau 核，这个核函数本身就是多项式，因此相比原证明减少了一步，但本质没有改变。魏尔斯特拉斯本人最初的证明不如伯恩斯坦的证明那么直截了当，那么优美（可以翻教科书参考，如果想详细了解过程，可以看菲赫金哥尔茨的《微积分学教程》，这是经典微积分教材）。当然这个定理最直观的证明是勒贝格的折线逼近法：闭区间上的连续函数可以用折线逼近（可以查书）。

极限是微积分的核心概念，微积分处理初等函数变化，一般都涉及无穷概念，无穷概念只有从极限角度理解，才能正确描述和把握，其实描述极限的语言体系是 $\epsilon-\delta$ 语言是一个相当于公理体系的定义，$\epsilon-\delta$ 意义下的极限是一种公理定义下的逼近，这种逼近不是几何描述的，所以没有逻辑悖论的可能。

逼近的常见技巧是放缩和夹逼，也即不等式是极限的主要技巧。

微积分中讨论的连续函数、导数、定积分、级数的敛散性、多元函数的偏导数，广义积分的敛散性、重积分和曲线积分与曲面积分等等概念都是基于极限的思想方法给出。

3.2.3 连续

前面说过，微积分主要对象是初等函数，初等函数的本质性质就是连续，就像一元 n 次方程的根的本直性质的是对称一样，这是很本质的核心问题，当然微积分必须抓住。

所以换句话说，微积分主要工作对象就是连续函数。其实人类在直到牛顿莱布尼兹时代，并不知道还有非连续的函数概念。预先假定都是连续的，而且他们对连续函数理解仅仅是几何直观，把能一笔画成的曲线所对应的函数叫做连续函数。例如伽利略所研究的落体运动，开普勒所研究的绕日运转的行星所扫描的扇形面积，牛顿所研究的流等都是连续变化的量。

所谓连续，直观解释就是运动变化的过程连绵不断，连续函数就是刻画变量连续变化的数学模型。

微积分是以直为曲的，所以对连续函数也要进行这种处理，例如柯西和魏尔斯特拉斯就用离散的多项式来逼近连续函数，这就是极限理论的由来，有了极限，才开始真的能够把握连续函数的性质。

最早人类理解连续函数，就是当 x 逐渐改变时，函数 f(x) 的相应变动也是逐渐的，不会有任何突增或突减的跳跃式振荡。但这种理解毫无用处，因为既不能计算，也不能控制。

定义 3.1: 函数连续的精确表述

函数连续的精确表述：设函数 $f(x)$ 在点 x_0 的某一去心邻域内有定义，任给 ϵ 大于零，存在 δ 大于零，当 $|x-x_0|<\delta$ 时，有 $|f(x)-f(x_0)|<\epsilon$, 则称函数 $f(x)$ 在 x_0 点连续。

这就是数学分析的基本语言：$\epsilon-\delta$ 语言，不熟悉这套语言体系，无法学会数学分析。

用 $\epsilon-\delta$ 语言定义的连续函数，就能计算其极限问题，这是微积分的重要内容，因为微分本质就是计算极限。

而连续函数求极限这种复杂问题本质是可以转化为求函数值的问题的，这就可以大大简化求极限难度。

我们知道，函数的连续性是一个局部性质，对区间也不例外。但如果是闭区间上的连续函数，却能把局部性质转化为整体性质，象闭区间上连续函数的有界性、最大最小值性、介值性、根的存在性、一致连续性等。

用 $\epsilon-\delta$ 语言，我们就能把握连续函数的性质：

连续函数的局部性质：若函数 f 在点 x_0 连续，则 f 在点 x_0 有极限，且极限值等于函数值 $f(x_0)$。根据这个性质，可以容易证明下述定理：

定理 3.1: 局部有界性定理

部有界性定理：若函数 f 在点 x_0 连续，则 f 在 x_0 的某邻域 $U(x_0)$ 内有界。

定理 3.2: 局部保号定理

部保号定理：若函数 f 在点 x_0 连续，且 $f(x_0)>0$(或 <0)，则对任何正数 $r<f(x_0)$(或$r<-f(x_0)$)，存在某 $U(x_0)$，使得对一切 $x\in U(x_0)$ 有 $r<f(x)$(或$r<-f(x)$)。

定理 3.3: 四则运算定理

则运算定理：若函数 f 和 g 在点 x_0 连续，则$f\pm g$, $f*g$, f/g(这里 $g(x_0)\neq 0$)也都在点 x_0 连续。

定理 3.4: 复合函数定理

合函数定理：若函数 f 在点 x_0 连续，g 在点 u_0 连续，$u_0=f(x_0)$，则 $\lim g(f(x))(x\to x_0)=g(\lim f(x))(x\to x_0)=g(f(x_0))$

定理 3.5: 海涅（Heine) 定理

涅（*Heine)* 定理：$lim_{x\to x_0}f(x)$ 存在的充分必要条件是对任给的序列 $\{x_n\}$，若满足 $lim_{n\to\infty}x_n = x_0(x_n \neq x_0)$，则有 $lim_{n\to\infty}f(x_n)$ 存在。

定理 3.6: 最大、最小值定理

大、最小值定理：若函数 f 在闭区间 $[a,b]$ 上连续，则 f 在 $[a,b]$ 上有最大值与最小值；或称函数 f 在 $[a,b]$ 上达到最大值。

定理 3.7: 有界性定理

论（有界性定理）：若函数 f 在闭区间 $[a,b]$ 上连续，则 f 在 $[a,b]$ 上有界。

定理 3.8: 介值性定理

值性定理：设函数 f 在闭区间 $[a,b]$ 上连续，且 $f(a) \neq f(b)$。若 μ 为介于 $f(a)$ 与 $f(b)$ 之间的任何实数（$f(a) < \mu < f(b)$ 或 $f(a) > \mu > f(b)$），则至少存在一点 $x_0 \in (a,b)$ 使得 $f(x_0) = \mu$。

定理 3.9: 根的存在定理

的存在定理：若函数 f 在闭区间 $[a,b]$ 上连续，且 $f(a)$ 与 $f(b)$ 异号，则至少存在一点 $x_0 \in (a,b)$ 使得 $f(x_0) = 0$。即方程 $f(x) = 0$ 在 (a,b) 内至少有一个根。

定理 3.10: 反函数连续定理

函数连续定理：若函数 f 在 $[a,b]$ 上严格单调并连续，则反函数 f^{-1} 在其定义域 $[f(a),f(b)]$ 或 $[f(b),f(a)]$ 上连续。

定理 3.11: 初等函数的连续定理

等函数的连续定理：任何初等函数在它的定义域上都连续。

3.2.4 导数

导数最初定义是 1823 年柯西在《无穷小分析概论》中定义的：如果函数 y=f(x) 在变量 x 的两个给定的界限之间保持连续，并且我们为这样的变量指定一个包含在这两个不同界限之间的值，那么是使变量得到一个无穷小增量。

现在导数定义是 19 世纪 60 年代魏尔斯特拉斯用 $\epsilon-\delta$ 语言定义的：设函数 $y=f(x)$ 在点 x_0 的某个邻域内有定义，当自变量 x 在 x_0 处有增量 Δx,$(x_0+\Delta x)$ 也在该邻域内时，相应地函数增量 $\Delta y=f(x_0+\Delta x)-f(x_0)$, 如果任意给 $\epsilon>0$, 存在常数 a 和 $\delta>0$, 当 $|\Delta x|<\epsilon$ 时, 使 $|\frac{\Delta y}{\Delta x}-a|<\epsilon$，则称函数 $y=f(x)$ 在点 x_0 处可导，并称这个极限为函数 $y=f(x)$ 在点 x_0 处的导数, 记为 $f^{'}(x_0)$，也记作 $y^{'}|_{x=x_0}$ 或 $\frac{dy}{dx}|_{x=x_0}$。

导数的几何直观就是函数形成的曲线在一点的切线的斜率。

最早导数主要用于求变速运动的瞬时速度（计算弹头的穿透能力或动能必须知道弹头接触目标的瞬时速度）和求曲线上一点的切线。牛顿从第一个问题出发，莱布尼兹从第二个问题出发，分别给出了导数的概念。

牛顿的想法很直观，如一辆汽车在 10 小时内走了 600 公里，它的平均速度是 60 公里/小时。但在实际行驶过程中，是有快慢变化的，不都是 60 公里/小时。设汽车所在位置 s 与时间 t 的关系为：$s=f(t)$, 那么汽车在由时刻 t_0 变到 t_1 这段时间内的平均速度是：

$\frac{f(t_1)-f(t_0)}{t_1-t_0}$，当 t_1 与 t_0 无限趋近于零时，汽车行驶的快慢变化就不会很大，瞬时速度就近似等于平均速度。

自然就把当 $t_1\to t_0$ 时的极限 $\lim\frac{f(t_1)-f(t_0)}{t_1-t_0}\frac{f(t_1)-f(t_0)}{t_1-t_0}$ 作为汽车在时刻 t_0 的瞬时速度，这显然就是导数。

显然根据上述定义，导数是通过极限对函数进行局部的线性逼近，所以导数是函数的局部性质。一个函数在某一点的导数描述了这个函数在这一点附近的变化率。

显然不是所有的函数都有导数（例如产生突变点，奇点的函数就没有导数），一个函数也不一定在所有的点上都有导数。

若某函数在某一点导数存在，则称其在这一点可导，否则称为不可导。显

然很容易证明：可导的函数一定连续；不连续的函数一定不可导。

如果函数 $y = f(x)$ 在开区间内每一点都可导，就称函数 $f(x)$ 在区间内可导。这时函数 $y = f(x)$ 对于区间内的每一个确定的 x 值，都对应着一个确定的导数，这就构成一个新的函数，这个函数为原来函数 $y = f(x)$ 的导函数，记作 y'、$f'(x)$、$\frac{dy}{dx}$ 或 $\frac{df(x)}{dx}$，简称导数。

显然，导数运算满足一下性质：

$$(u \pm v)' = u' \pm v'$$

$$(uv)' = u'v + uv'$$

$$(u/v)' = \frac{(u'v - uv')}{v^2}。$$

根据上导数定义和性质，很容易计算出一些常见函数的导数：

$$y = x^n, y' = nx^{n-1}$$

$$y = a^{bx}, y' = ba^{bx}lna$$

$$y = a^u, y' = u'a^u lna$$

$$y = e^b x, y' = be^{bx}$$

$$y = e^u, y' = u'e^u$$

$$y = loga^x, y' = \frac{1}{xlna}$$

$$y = lnx, y' = \frac{1}{x}$$

$$y = sinx, y' = cosx$$

$$y = cosx, y' = -sinx$$

$$y = tanx, y' = sec^2(x)$$

$$y = cotx, y' = -csc^2(x)$$

$$y = secx, y' = secxtanx$$

$$y = cscx, y' = -cscxcotx$$

$$y = arcsinx, y' = \frac{1}{(1-x^2)^{\frac{1}{2}}}$$

$$y = arccosx, y' = -\frac{1}{(1-x^2)^{\frac{1}{2}}}$$

$$y = arctanx, y' = \frac{1}{1+x^2}$$

$y = arccotx, y' = -\frac{1}{1+x^2}$

$y = shx, y' = chx$

在实际上应用中，大部分常见的函数都上述函数的和、差、积、商或相互复合的结果。所以一般情况下，函数的导函数计算是简单容易的。

导数的几个用途：

判别单调性：若导数大于零，则单调递增；若导数小于零，则单调递减；导数等于零为函数驻点，不一定为极值点。

求极值：如果存在一点，使得导数在之前区间上都大于等于零，而在之后区间上都小于等于零，那么是一个极大值点，反之则为极小值点。

自然推论：若已知函数为递增函数，则导数大于等于零；若已知函数为递减函数，则导数小于等于零。

判断函数凹凸性：如果函数的导函数在某个区间上单调递增，那么这个区间上函数是向下凹的，反之则是向上凸的。如果二阶导函数存在，如果在某个区间上二阶导数恒大于零，则这个区间上函数是向下凹的，反之这个区间上函数是向上凸的。曲线的凹凸分界点称为曲线的拐点。

导数的最著名应用是中值定理和洛必达法则。

中值定理应包括罗尔中值定理、拉格朗日中值定理、柯西中值定理、泰勒中值定理。

罗尔中值定理：如果函数 f(x) 满足：在闭区间 [a,b] 上连续；在开区间 (a,b) 内可导；在区间端点处的函数值相等，即 f(a)=f(b)，那么在 (a,b) 内至少有一点 $\xi(a<\xi<b)$，使得 $f'(\xi)=0$。

几何上，罗尔定理含义是一条连续的曲线弧，如果除端点外处处有不垂直于 x 轴的切线，且两端点的纵坐标相等，则弧上至少有一点的切线是水平的。

拉格朗日定理：如果函数 $f(x)$ 满足：在闭区间 $[a,b]$ 上连续；在开区间 (a,b) 内可导，那么在 (a,b) 内至少有一点 $\xi(a<\xi<b)$，使等式 $f(b)-f(a)=f'(\xi)(b-a)$ 成立。

柯西定理：如果函数 f(x) 及 F(x) 满足：在闭区间 [a,b] 上连续；在开区间 (a,b) 内可导；对任一 $x\in(a,b)$，$F'(x)\neq 0$，那么在 (a,b) 内至少有一点 ξ，使等

式 $\frac{[f(b)-f(a)]}{[F(b)-F(a)]}=\frac{f'(\xi)}{F'(\xi)}$ 成立。

泰勒公式：若函数 f(x) 在开区间（a，b）有直到 n+1 阶的导数，则当函数在此区间内时，可以展开为一个关于 $(x-x_0)$ 多项式和一个余项的和：

$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\frac{f'''(x_0)}{3!}(x-x_0)^3+\cdots+\frac{f(n)(x_0)}{n!}(x-x_0)^n+R_n$ 其中 $Rn=\frac{f^{n+1}(\xi)}{(n+1)!}(x-x_0)^{n+1}$，这里 ξ 在 x 和 $x,0$ 之间，该余项称为拉格朗日型的余项。（$f^n(x_0)$ 是 $f(x_0)$ 的 n 阶导数，不是 $f(n)$ 与 x_0 的相乘）

推论：麦克劳林公式：

若函数 f(x) 在开区间（a，b）有直到 n+1 阶的导数，则当函数在此区间内时，可以展开为一个关于 x 多项式和一个余项的和：$f(x)=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\cdots+\frac{f^n(0)}{n!}x^n+R_n$ 其中 $Rn=\frac{f^{n+1}(\theta x)}{(n+1)!}x^{n+1}$, 这里 $0<\theta<1$。

达布定理：若函数 $f(x)$ 在 $[a,b]$ 上可导, 则 $f'(x)$ 在 [a,b] 上可取 $f'(a)$ 和 $f'(b)$ 之间任何值。

推广：若 f(x),g(x) 均在 [a,b] 上可导, 并且在 [a,b] 上,g′(x)≠0, 则 f′(x)/g′(x) 可以取 f′(a)/g′(a) 与 f′(b)/g′(b) 之间任何值。

洛必达法则：设当 $x\to a$ 时，函数 $f(x)$ 及 $F(x)$ 都趋于零；在点 a 的去心邻域内，$f'(x)$ 及 $F'(x)$ 都存在且 $F'(x)\neq 0$；当 $x\to a$ 时 $\lim\frac{f'(x)}{F'(x)}$ 存在 (或为无穷大)，那么 $x\to a$ 时 $\lim\frac{f(x)}{F(x)}=\lim\frac{f'(x)}{F'(x)}$。

又设当 $x\to\infty$ 时，函数 $f(x)$ 及 $F(x)$ 都趋于零；当 $|x|>N$ 时 $f'(x)$ 及 $F'(x)$ 都存在，且 $F'(x)\neq 0$；当 $x\to\infty$ 时 $\lim\frac{f'(x)}{F'(x)}$ 存在 (或为无穷大)，那么 $x\to\infty$ 时 $\lim\frac{f(x)}{F(x)}=\lim\frac{f'(x)}{F'(x)}$。

中值定理经常用于证明方程根的存在性，证明恒等式，证明不等式，研究函数的单调性，求函数极限（用罗必达法则求 $0/0$，∞/∞ 函数极限是常用手段），求函数的极值与最值，讨论函数的凸凹性，求函数的拐点，求函数的渐近线，描

绘函数的图象等等。具体例子可以查教科书。

3.2.5 微分

其实导数和微分概念是一致的，没什么更多可说的。

函数 $y = f(x)$ 的微分 $dy = f'(x)dx$。可导与可微是等价的，若求出了函数在一点的导数，再乘以 dx 即得该点的微分；若求出了函数在一点的微分，再除以 dx 即得该点的导数；因此导数又叫做微商。

需要注意的是函数在 x 点的微分是自变量增量的线性函数，因为微分是对函数的局部变化的一种线性描述。如果一个非线性函数某点可微，其在某点的自变量有一个微小的改变时，函数的变化可以分解为两个部分。一个部分是线性部分：在一维情况下，它正比于自变量的变化量 Δx，可以表示成 Δx 和一个与 Δx 无关，只与函数及有关的量的乘积；在更广泛的情况下，它是一个线性映射作用在 Δx 上的值。另一部分是比 Δx 更高阶的无穷小，也就是说除以 Δx 后仍然会趋于零。当改变量很小时，第二部分可以忽略不计，函数的变化量约等于第一部分，也就是函数在 x 处的微分。

所以微分主要用于计算函数值的近似值。

但是不是所有的函数的变化量都可以分为以上提到的两个部分。若函数在某一点不可微，就无法用线性函数逼近。

在现代微积分中，微分被定义为将自变量的改变量映射到变化量的线性部分的线性映射。这个映射也被称为切映射。给定的函数在一点的微分如果存在，就一定是唯一的。

微分有以下运算法则：

连锁律：$\dfrac{dy}{dx} = \dfrac{dy}{dz}\dfrac{dz}{dx}$；

乘法律：$\dfrac{d(uv)}{dx} = u(\dfrac{dv}{dx}) + v(\dfrac{du}{dx})$；

除法律：$\dfrac{d(u/v)}{dx} = \dfrac{(v(\dfrac{du}{dx}) - u(\dfrac{dv}{dx}))}{v^2}$

$\dfrac{dy}{dx}$ 被称为一阶导数

$\frac{d(\frac{dy}{dx})}{dx} = \frac{d^2y}{dx^2}$ 被称为二阶导数

以此类推，

$\frac{d^ny}{dx^n}$ 被称为 n 阶导数。

稍微多说一句是法线。曲线上一点的法线和那一点的切线互相垂直，微分可以求出切线的斜率，自然也可以求出法线的斜率。函数 $y = f(x)(x_0, y_0)$ 点切线的斜率为 $m = \frac{dy}{dx}$ 在 (x_0, y_0) 的值，那么法线的斜率为 $\frac{-1}{m}$。

3.2.6 积分

积分原始思想的萌芽很早，甚至早于微分思想，主要用于计算物体运动的路程、变力作功以及由曲线围成的面积和由曲面围成的体积等问题，现在资料据说古希腊德莫克利特、阿基米德、中国的刘徽都用积分思想计算过面积和体积，当然这些方法都建立在特殊的技巧之上，不具有一般性，也没有逻辑基础保证其是正确的。再晚一点，开普勒的“同维无穷小方法”、卡瓦列利的“不可分量法”、费马的“分割求和方法”更是典型的积分思想。

不过真正的积分发明者还是牛顿与莱布尼兹，因为他们揭示了微分与积分的内在联系–微积分基本定理，也即牛顿–莱布尼茨公式。

积分是微积分的一个核心概念。通常分为定积分和不定积分两种。

定积分严格的数学定义是黎曼用的方式极限给出的，把曲边梯形设想为一系列矩形组合的极限，也即对于一在区间 $[a, b]$ 上之给定非负函数 $f(x)$，$f(x)$ 所代表的曲线与 Ox 坐标轴所夹图形的面积

$S = \int_b^a f(x)dx = lim_{n\to\infty} \sum_{i=0,1,\cdots,n-1} f(t_i)(x_{i+1} - x_i)$

上述符号定义：在闭区间 $[a, b]$ 中取一个有限的点列 $a = x_0 < x_1 < x_2 < \cdots < x_n = b$。每个闭区间 $[x_i, x_{i+1}]$ 构成一个子区间。定义 λ 为这些子区间长度的最大值：$\lambda = \max(x_{i+1} - x_i)$，其中 $0 \le i \le n - 1$。

一个闭区间 $[a, b]$ 进行分割 $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ 后，在每一个子区间中 $[x_i, x_{i+1}]$ 取出一点 $x_i \le t_i \le x_{i+1}$。

黎曼定义的积分的就是是微分的无限积累，或者说定积分是无限个无穷小

量之和。核心思想就是试图通过无限逼近来确定这个积分值。

所以定积分是一种极限, 这种极限不同于数列的极限也不同于函数的极限。它是和式的极限，对于体现自变过程的变量的每一个值，不仅区间的分法有无穷多种，而且对于每一个分法，介点也有无穷多种取，因而相应的和式一般有无穷多个值。但它仍然有着与数列极限、函数极限的本质上的相同之处，即当 $[x_i, x_{i+1}]$ 无限变小时，相应的一切和式与某一定数的距离能够变得并保持任意的小。

微积分的最初发展中, 定积分即黎曼积分。在实变函数中, 可以利用测度论将黎曼积分推广到更加一般的情况，如勒贝格积分。

显然，黎曼积分定义有一个自然问题就是这个黎曼和式是不是一定有极限，极限与子区间划分方法有无关系。

前者就是所谓的可积问题，后者是极限收敛问题。

决定是否可积一般依赖于四个因素：函数、区间、区间的分法、介值的的取法。

很容易证明，当函数在区间上可积时,，不依赖于区间的分法与介值的取法，函数积分数字只与函数和区间两个因素有关。所以在可积的条件下，当求某函数在指定区间上的定积分时，往往可以取一个特殊的分法（如 n 等分），取介值为划分内的特殊点（如左或右端点）。

可以证明下述结论：

□ 可积函数必有界，有界函数不一定可积，无界函数一定不可积；□ 连续函数一定可积；

□ 有有限个间断点的有界函数一定可积；

□ 有无限多个不连续点的单调函数一定可积；

□ 区间上有无限个不连续点的有界函数（只要间断点的测度为 0）也可积。

定积分的主要应用是求和，例如平面图形的面积，求已知截面面积的立体的体积，求旋转体的体积，求曲线的弧长，求旋转曲面的面积，求变力所作的功，计算运动物体的路程，以及物体之间的万有引力等等。另外，定积分可以作为定义函数的一种新的工具，例如连续函数的变上限积分是函数的一个原函数，又知道某些函数的原函数并不是初等函数。如椭圆积分就不是初等函数，这时

我们就把这个积分本身，作为此函数的定义，此为出发点来研究函数。

微积分最基础的定理是牛顿和莱布尼茨分别独自发现的：

一个可积函数在区间 $[a,b]$ 上的定积分等于它的任意一个原函数在区间 $[a,b]$ 上的增量。也即：如果函数 $f(x)$ 在区间 $[a,b]$ 上可级，并且存在原函数 $F(x)$，则：

$\int_b^a f(x)dx = F(b) - F(a)$

这个发现给定积分提供了一个有效而简便的计算方法，大大简化了定积分的计算过程。

这个定理是微积分存在的基础，但是证明极其简单。

证明 1 证明：因为函数 $f(x)$ 在区间 $[a,b]$ 上可积，任取区间的分割 $a = x_0 < x_1 < x_2 < \cdots < x_n = b$，在区间 $[x_i, x_{i+1}](i = 1, 2, \cdots, n-1)$ 上任取一点 ζ_i，则有 $lim_{n\to\infty} \sum_{i=1,\cdots,n-1} f(\zeta_i)(x_{i+1} - x_i) = \int_b^a f(x)dx$，

由于函数 f 是函数 F 的导函数，所以根据拉格朗日中值定理得 $F(x_i) - F(x_{i-1}) = f(\zeta_i)(x_{i+1} - x_i)$ 其中 $\zeta_i \in (x_{i-1}, x_i)$，因此有

$F(b)-F(a) = \lim_{n\to\infty} \sum_{i=1,\cdots,n-1} F(x_i)-F(x_{i-1}) = lim_{n\to\infty} \sum_{i=1,\cdots,n-1} f(\zeta_i)(x_{i+1}-x_i) = \int_b^a f(x)dx$。

牛顿-莱布尼茨公式简化了定积分的计算，只要知道被积函数的原函数，总可以求出定积分的精确值或一定精度的近似值。

牛顿-莱布尼茨公式是联系微分学与积分学的桥梁，它是微积分中最基本的公式，它证明了微分与积分是可逆运算，同时在理论上标志着微积分完整体系的形成，从此微积分成为一门真正的学科。

利用牛顿一莱布尼茨公式可以证明定积分换元公式，积分第一中值定理和积分型余项的泰勒公式。牛顿-莱布尼茨公式还可以推广到二重积分与曲线积分，从一维推广到多维。牛顿-莱布尼茨公式促进了其他数学分支的发展，在微分方程，傅里叶变换，概率论，复变函数等数学分支中都要用到。

下面说说不定积分。不定积分是是已知导数求原函数，用公式表示是：$\int f'(x)dx = f(x) + c$ 而前面已经说了，定积分是求面积（Riemann 和的极限），不定积分只是求导数的逆运算，所以不定积分与定积分是完全不同的两个概念。但是，牛顿莱布尼兹公式把它们连接在一起。

不过，函数在所讨论区间上的 Riemann 和的极限的存在性不取决于该函数的不定积分的存在性，函数在所讨论区间上的不定积分的存在性也不取决于该函数的 Riemann 和的极限的存在性。

因为容易证明：

函数可积不一定该函数存在原函数：因为 f(x) 在区间 [a,b] 上连续，在区间 [a,b] 上有界，且只有有限个第一类间断点，和在区间 [a,b] 上单调有界, 则 f(x) 都在在 [a,b] 上可积，由牛顿莱布尼兹公式知道，一个函数如果可导，那么它的导函数是不可能存在第一类间断点的，所以说一个函数如果存在第一类间断点，那么它是不会有原函数的，也即可积并不能保证有原函数。

函数连续只是可导的必要条件，而非充分条件（如果一个函数可导，其必然连续。如果一个函数连续，则不一定可导，如 $Y = |X|$)。

同时，也容易证明，函数有原函数但该函数不一定可积。例如，函数 $y = x^{\frac{3}{2}} sin(\frac{1}{x})$ 各点可导，但由于在闭区间 $[-1, 1]$ 上有无界点，故在 $[-1, 1]$ 上上不可积。

所以函数可积问题，是传统微积分没能解决的一个问题（有些函数是连续的但处处不可微，有的函数的有限导数并不黎曼可积，一个黎曼可积的函数列收敛到的那个函数不一定是黎曼可积的等等），直到实变函数发展起来，扩展了可积的概念，例如勒贝格积分，也扩展了基于勒贝格测度理解的连续函数的概念，这个问题才圆满解决。

显然，因为牛顿-莱布尼兹微积分基本公式，导数的公式逆向就是初等函数的积分公式，不必多说。

积分计算有非常多的技巧，换元，变量替换，逼近，因式分解等等，可以看教科书，里面有非常多的计算技巧例子。多做习题。华罗庚是世界现代数学家中计算能力名列前茅的变态，他的很多发现或定理证明都是算出来的，晚年的华罗庚为保证自己思维状态，每天没事干就是算积分玩，而且是极难的积分。这个不是传说，是亲眼所见。原来的科大数学系学生（77，78,79 三级），计算积分和矩阵，能力在中国所有大学中，无人能及，科大学生不能把华罗庚的线性代数的打洞公式和积分的变换技巧用得风生水起，都不算合格学生。

3.2.7 多元微积分

传统多元微积分的基本概念都是一元微分与积分的基本推广，1687 年牛顿就提出了偏导数和重积分的思想，欧拉在 1769 年给出了二重积分及其累次积分与换元计算方法，拉格朗日在 1773 年给出了三重积分及其累次积分与换元计算方法，雅可比在 1833 年给出了变量替换中的雅可比矩阵表达。不过当自变量是多元变量时，导数的概念已经不适用了（尽管可以定义对某个分量的偏导数），但仍然有微分的概念。

下面我们只介绍把一元函数微积分二元函数微积分情况，因为扩展到多元函数是类似的。

3.2.7.1 二元函数连续性的定义

多元函数微积分的推广，最初是从几何角度开始的。

二元函数 $u = f(x, y)$ 的变量 (x, y) 在一个平面直角坐标系中代表一个动点 P，它的全部可能的位置形成一个平面点集 S。从而函数关系 f 便把动点 P 的每一个位置 (x, y) 对应到变量 u 的一个惟一确定的数值（函数值）$f(x, y) = f(P)$。于是整个函数便表现为变量 u 按照这个对应关系随着动点 P 在定义域 S 上变化而变化，这样，二元函数的概念便同一元函数的一致。

当动点 P 由一个位置 $P(x, y)$ 变到另一个位置 $P_1(x_1, y_1)$ 时，这变化由它的位移向量 $\vec{PP_1} = \{x_1 - x, y_1 - y\} = \{\Delta x, \Delta y\} = \Delta P$ 来刻画，这变化的大小便由这向量的长度 $\left|\Delta P\right| = ((\Delta x)^2 + (\Delta y)^2)^{\frac{1}{2}}$ 来度量。相应的 u 的变化 Δu 其大小由 $|\Delta u|$ 来度量。

$$\Delta u = f(p_1) - f(p) = f(x + \Delta x, y + \Delta y) = f(x, y)$$

于是多元函数在一点 P 处的连续性也可以用一元函数连续型定义，也即，在 P1 无限趋近于 P 的过程中，$|\Delta u|$ 随着 $|\Delta P|$ 而无限变小。这就是说，多元函数 u 连续，就是任意给定 $\epsilon > 0$，都存在一个 $\delta > 0$，使得只要 $|\Delta P| < \delta$，就有 $|\Delta u| < \epsilon$。

所以多元连续函数的基本性质也同一元连续函数的一样：

□ 多元函数在一有界闭集 S 上定义，其在 S 上处处连续，则至少在某一点处达到最小值 m, 又至少在某一点处达到最大值 M；

□ 多元函数连续性在整个集合 S 上是一致的（即 δ 不依赖于 P 而对于 S 上的每个点 P 都有效）；

□ 如果 S 是连通的（即 S 上每两点都能够用完全位于 S 上的一条折线连接起来）, 则每一个中间值 $\mu(m \leq \mu \leq M)$ 都是某一点处的函数值；

□ 多元函数如果连续，它在 S 的每个内点处都可以分解成一元的情形：函数 u 在一点 P 的某个领域 (δ) 内处处连续, 则必定在其内部的一个方邻域 [δ] 上一致连续，而在这个方邻域上的变化量具有向量分解式：

$\Delta u = \Delta_x u + \Delta_y u$

式中

$\Delta_x u = f(x+\Delta x, y) - f(x,y), \Delta_y u = f(x+\Delta x, y+\Delta y) - f(x+\Delta x, y)$

分别作为一元函数 $g(x) = f(x,y), h(y) = (x+\Delta_x, y)$，显然其连续性分别关于 y 或 x+Δx 是一致的（即相应的 δ 不依赖于 y 或 $x+\Delta x$ ）。

3.2.7.2 偏导数

定义了多元函数连续性，就能定义导数了。显然用 $\Delta u = \Delta_x u + \Delta_y u$ 和 $g(x) = f(x,y), h(y) = (x+\Delta x, y)$ 能够证明在 $|\Delta P|$ 趋向 0 的过程中，变化量 Δu 随 $\Delta x, \Delta y$ 趋向 0 的依赖关系。

这就要用到一元函数 g,h 变化率，即导数 g(x)、h(y)。假定 g,h 的导数它们在 $P(x,y)$ 的附近都存在，并分别记为 $f(x,y), f(x+\Delta x, y)$,

$$\frac{\partial u}{\partial x} = f_x^{'}(x,y) = \lim_{\Delta x \to 0} \frac{(f(x+\Delta x, y) - f(x,y))}{\Delta x},$$

$$\frac{\partial u}{\partial y} = f_y^{'}(x,y) = \lim_{\Delta x \to 0} \frac{(f(x, y+\Delta y) - f(x,y))}{\Delta x}。$$

这种对自变量之一 (其余作为参变量) 的导数称为偏导数。利用这些偏导数的存在和一元微分学的中值定理，可以得到：$\Delta u = \Delta_x u + \Delta_y u = f_x^{'}(x,y)\Delta x + f_y^{'}(x+\Delta x, y+\theta\Delta y)\Delta y + \alpha\Delta x)$

式中 θ 介于 0 到 1 之间,α 为无限小量。当偏导数连续时，可以进一步写成：

$\Delta u = \dfrac{\partial u}{\partial x}\Delta x + \dfrac{\partial u}{\partial y}\Delta y + \alpha(\Delta x) + \beta(\Delta y)$，$\alpha, \beta$ 为无限小量。

3.2.7.3 全微分

函数 u 在点 P 处是可微的定义：

$\Delta u = \Delta_x u + \Delta_y u = f'_x(x,y)\Delta x + f'_y(x+\Delta x, y+\theta\Delta y)\Delta y + \alpha(\Delta x)$ 表明，在点 P 处，变化量 Δu 随着 $\Delta x, \Delta y$ 趋向 0 的过程中，存在着近似线性的依赖关系：$\Delta u = A\Delta x + B\Delta y + \alpha\Delta x + \beta\Delta y$,

式中主要部分的系数 A、B 不依赖于 $\Delta x, \Delta y$，而余项部分的系数 α, β 是无限小量。

并把这个线性主要部分为 u 的一个 (全) 微分，记为 $du = A\Delta x + B\Delta y$ 令 $\Delta x \to 0, \Delta y = 0$ 或 $\Delta x = 0, \Delta y \to 0$，即可推出:$A = \partial u/\partial x, B = \partial u/\partial y$,

所以只要微分存在，它的系数就必然是偏导数，因而是惟一的。

在某些特殊情形, 这些偏导数都存在，$du = \partial u/\partial x\Delta x + \partial u/\partial y\Delta y$ 关系却不成立；所以不同于一元函数的情形：只有偏导数的存在还不能保证微分存在。

不过偏导数的连续性可以保证微分存在。也即函数是连续可微的，所以这时 u 的微分可以写成 $du = \partial u/\partial x dx + \partial u/\partial y dy$。具体证明可以查教科书，这里不啰嗦，因为很简单（因为 x,y 是动点 P 的连续函数）。

3.2.7.4 变量替换

变量 x、y 既然当作动点 P 的函数，也就可以表达为：动点 P 在任一别的坐标系 (r,s) 中的坐标的函数：$x = \phi(r,s), y = \psi(r,s)$

假定这些坐标函数也在其定义域 S 上是处处连续可微的，也就是说，出现在下列微分等式中的系数都是连续的：

$dx = \dfrac{\partial x}{\partial r}dr + \dfrac{\partial x}{\partial s}ds, dy = \dfrac{\partial y}{\partial r}dr + \dfrac{\partial y}{\partial s}ds,$

既然 u 关于 (x,y) 连续可微，那么根据微分教计算规则，得到：$\dfrac{\partial u}{\partial r} = \dfrac{\partial u}{\partial x}\dfrac{\partial x}{\partial r} + \dfrac{\partial u}{\partial y}\dfrac{\partial y}{\partial r}$，$\dfrac{\partial u}{\partial s} = \dfrac{\partial u}{\partial x}\dfrac{\partial x}{\partial s} + \dfrac{\partial u}{\partial y}\dfrac{\partial y}{\partial s}$。

这些偏导数都是关于新变量 (r,s) 连续可微的函数。于是 u 也关于 (r,s) 连续可微，因而得到：

$du = \partial u/\partial r dr + \partial u/\partial s ds = \partial u/\partial x dx + \partial u/\partial y dy$。

这表明微分形式对于 x,y 为任何连续可微的函数都成立。这称为（一阶）微分的形式不变性。

变量替换规定了一个坐标平面上的动点 $P(x,y)$ 随着另一坐标平面上的动点 $Q(r,s)$ 而变动，因而定义了一个函数 $T: P = T(Q)$。这样得到一个一个矩阵方程：

这里，偏导数所形成的矩阵称为雅可比矩阵。它是微分向量的系数矩阵，相当于一元函数情形的微分系数或导数。

如果动点 P 是在一个三维坐标空间 (r,s,t) 中，则函数应是三元的：$x = \phi(r,s,t), y = \psi(r,s,t)$, 雅可比矩阵则是：

以此类推，一元函数微分的主要定理都能推广到多元微分中。

3.2.7.5 重积分

一元函数的定积分，作为黎曼积分和的极限，推广到二元函数几乎是直接的。只不过把积分区间换成了两个区间 $X(\alpha \le x \le A)$ 和 $Y(b \le y \le B)$，它们的乘积 $R = X \times Y$ 是包含有界闭区域 S 的（各边平行于坐标轴的）最小的矩形。对于 R 上不属于 S 的点，取函数值为 0, 并仿照一元的情形作黎曼和数：

$S_\Delta = \sum f(\zeta_i,\eta_j)\Delta x_i \Delta y_j, (\zeta_i,\eta_j) \in \Delta x_i * \Delta y_j (i = 1,\cdots,n; j = 1,\cdots,m)$

分划 (Δ) 的细密程度由全部 $\Delta x_i, \Delta y_j$ 的最大值 $||\Delta||$ 来度量。于是，可以像一元的情形一样来定义二重积分：

$\iint_S f(x,y)ds = \iint_R f(x,y)dxdy = \lim S_\Delta (||\Delta|| \to 0)$

如果这个极限存在，就说函数 f 在区域 S 上是可积的。

可积的一个充分必要条件仍然是: 函数有界并且几乎处处连续（即不连续点形成一个零测度集合）。不过，这里的零测度集合，作为平面上的点集，是指能用总面积任意小的矩形序列覆盖住。

在可积的前提下，二重积分可以写成：

$\iint_S f(x,y)ds = \int_b^B \int_a^A f(x,y)dx$, 内层积分以 y 为参变量，在不可积（因而相应的 y 值形成一个一维零测度集合）时算作 0。

面积微分 $dR = dxdy$, 作为一个微小矩形的面积，在坐标变换之下成为一个以向量 $\{\frac{\partial x}{\partial r}dr, \frac{\partial y}{\partial r}dr\}$ 和 $\{\frac{\partial x}{\partial s}ds, \frac{\partial y}{\partial s}ds\}$ 为一对邻边的平行四边形的面积。

所以有二重积分的换元公式：

$\iint_S f(x,y)ds = \iint_{r*s} f(x,y,z)(EG - F^2)^{\frac{1}{2}}drds$。

3.2.7.6 三维空间的曲面积分

二重积分可以推广到三维空间中的一块曲面 S 上, 只要这曲面是光滑的，即其上的动点 P(x,y,z) 的坐标能够表示成某一平面矩形 $S = r * s(a \le r \le A, b \le s \le B)$ 上的连续可微的函数，而以 (r,s) 作为 P 的一种新的坐标（曲面坐标）。这里 S 的微小矩形 $(\Delta r) \times (\Delta s)$ 对应着 S 上的微小曲面四边形 ΔS，后者的面积关于前者的面积。$\Delta r \Delta s$ 的线性主要部分便是曲面的面积微分 dS。它等于以切线向量

$\{\frac{\partial x}{\partial r}dr, \frac{\partial y}{\partial r}dr, \frac{\partial z}{\partial r}dr\}$

和

$\{\frac{\partial x}{\partial s}ds, \frac{\partial y}{\partial s}ds, \frac{\partial z}{\partial s}ds\}$

为一对邻边的平行四边形的面积：$ds = (EG - F^2)^{\frac{1}{2}}$,

其中：

$E = (\frac{\partial x}{\partial r})^2 + (\frac{\partial y}{\partial r})^2 + (\frac{\partial z}{\partial r})^2$,

$F = \frac{\partial x}{\partial r}\frac{\partial x}{\partial s} + \frac{\partial y}{\partial r}\frac{\partial y}{\partial s} + \frac{\partial z}{\partial r}\frac{\partial z}{\partial s}$,

$G = (\frac{\partial x}{\partial s})^2 + (\frac{\partial y}{\partial s})^2 + (\frac{\partial z}{\partial s})^2$

从而面积分能够表示成二重积分：

$\iint_s f(x,y,z)ds = \iint_{r*s} f(x,y,z)(EG - F^2)^{\frac{1}{2}}drds$

曲面 S 可以是逐片光滑的, 积分便取为各片上的积分之和。

如果是三维空间的曲线积分，类似地考虑空间中一条光滑的 (或逐段光滑的) 曲线 C 上关于弧长的微分 ds 的积分：$\int_c f(x,y,z)ds$

则有

$\int_c f(x,y,z)ds = \int_a^b f(x,y,z)((dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2)^{\frac{1}{2}}$,

这就与一个直线段 $a \leq s \leq b$ 上的定积分没区别了。

实际上多元定积分在概念上的各种推广，在计算上仍都能回到定积分。

3.2.7.7 牛顿-莱布尼茨公式推广

我们知道，一元微积分之所以成立，就是靠牛顿-莱布尼茨公式。

多元微积分想成立，也得有这种把微分和积分联系起来的公式。

在一元微积分中，根据牛顿-莱布尼茨公式，定积分是微分之逆，在多元微积分中，这个定理仍然是成立的。

二重积分推广：设函数 $f(x,y)$ 在矩形区域

$D = \left|(x,y)\right|(a \leq x \leq b, c \leq y \leq d)$

上连续，如果存在一个二元函数 $F(x,y)$, 使得 $\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$,

则二重积分 $\int_D \int f(x,y)dxdy = F(b,d) - F(b,c)$ 更多重积分也有类似公式。对曲线积分，也有类似公式。设 D 为单连通区域，$P(x,y)$ 和 $Q(x,y)$ 在区域 D 上有连续的一阶偏导数，若存在一个二元函数 $u(x,y)$，使得 $du(x,y) = P(x,y)dx + Q(x,y)dy$ 在区域 D 中任意取两个点 A，B，则对连接 A，B 的任意一条光滑曲线 L，都有:$\int_L P(x,y)dx + Q(x,y)dy = u(B) - u(A)$

另外，必须熟悉的还有斯托克斯公式（格林公式），奥斯特罗格拉茨基公式（高斯公式）等等，只是这些公式没法在豆瓣显示，有兴趣的自己去查书。

多元积分的计算技巧主要是变量替换，教科书中有大量人类积累下来的变量替换的技巧例子，可以通过多做习题，积累下自己的计算技巧，熟能生巧，培养出自己强大的计算能力。

显然，介绍的都是最古典微积分在多元上函数上的推广，现代教科书没有这么复杂，简单明了，例如定义多元函数可微，一般是：设 f 是从欧几里得空间 Ω（或者任意一个内积空间）中的一个开集射到 R^m 的一个函数。对于 Ω 中的一点 x 及其在 Ω 中的邻域 Λ 中的点 $x+h$。如果存在线性映射 A 使得对任意这样的 $x+h$, $\lim \frac{f(x+h) - f(x) - A(h)}{h} = 0(h \to 0)$, 那么称函数 f 在点 x 处可微。

线性映射 A 叫做 f 在点 x 处的微分，记作 df_x。

如果 f 在点 x 处可微，那么它在该点处一定连续，而且在该点的微分只有一个。

当函数在某个区域的每一点 x 都有微分 df_x 时，可以考虑将 x 映射到 df_x 的函数：$df : x \to df_x$，这个函数一般称为微分函数。

而且利用一元微分性质，可以证明：如果 f 是线性映射，那么它在任意一点的微分都等于自身。

在 R^n（或定义了一组标准基的内积空间）里，函数的全微分和偏导数间的关系可以通过雅可比矩阵刻画。

也可以证明如下结论：

可微的必要条件：如果函数 f 在一点 x_0 处可微，那么雅克比矩阵的每一个元素都存在，但反之不真。

可微的充分条件：如果函数 f 在一点 x_0 的雅克比矩阵的每一个元素都在 x0 连续，那么函数在这点处可微，但反之不真。

3.2.7.8 级数

级数主要两个用途，一个是构造新函数，一个是表示、逼近已知函数（主要用于函数的近似计算）。

在微积分中，会涉及一些初等函数之外的函数，一般都是用级数表达的，因为他们的级数形式，便于了解它们的性质。

级数的基本工具是泰勒级数（用有限项的多项式近似表示函数）和三角级数（傅利叶级数，表达周期性函数），级数主要用于连续函数的局部逼近和整体逼近，当然从逻辑上来讲，可以用无穷多项的多项式来准确地表示一个函数，这就是幂级数。利用函数的幂级数展开式，对研究函数的性质和计算都有着非常重要的作用。

当然，能表示成幂级数的函数必须具备任意阶可微的¡件，这对于有些性质较差的函数（如分段函数），我们就不能展开成幂级数，此时付立叶级数却能满足这样的函数的展开。

级数理论的基础是极限，级数是一个无限求和的过程，它与有限求和有着根本的不同，即参与了极限运算，把极限及其运算性质移植到级数中去，就形成了级数的一些独特性质。

所以级数的第一个重要概念是收敛性（也即存在极限）。此外，级数的运算、函数项级数的一致收敛性、一致收敛级数的分析性质、函数的幂级数展开、函数的付立叶级数展开都是级数理论的基本内容。

数列级数

将数列 u_n 的项 $u_1, u_2, \cdots, u_n$ 用加号连接起来的函数就称数项级数简写为 $\sum u_n$，记 $S_n = \sum u_n$, 如果当 $n \to \infty$ 时，S_n 这个数列有极限，则说级数收敛，并以 S 为其和, 否则就说级数发散。

级数收敛的柯西准则：$\sum u_n$ 收敛 <=> 任意给定正数 ϵ，必有自然数 N，当 $n > N$，对一切自然数 p，有 $|u[n+1] + u[n+2] + \cdots + u[n+p]| < \epsilon$。即级数充分靠后的任意一段和的绝对值可任意小。

如果每一 $u_n \geq 0$（或 $u_n \leq 0$），则称 $\sum u_n$ 为正（或负）项级数，正项级数与负项级数统称为同号级数。正项级数收敛的充要条件是其部分和序列 S_m 有上界。

有无穷多项为正，无穷多项为负的级数称为变号级数，其中最简单的是形如 $\sum[(-1)^{n-1}]u_n(u_n > 0)$ 的级数，判别这类级数收敛的基本方法是莱布尼兹判别法：若 $u_n \geq u_{n+1}$，对每一 $n \in N$ 成立，且当 $n \to \infty$ 时 $\lim u_n = 0$，则交错级数收敛。

对于一般的变号级数如果有 $\sum |u_n|$ 收敛，则称变号级数绝对收敛。如果只有 $\sum u_n$ 收敛，但是 $\sum |u_n|$ 发散，则称变号级数条件收敛。（例如 $\sum[(-1)^{n-1}](\frac{1}{n^2})$ 绝对收敛，而 $\sum[(-1)^{n-1}](\frac{1}{n})$ 只是条件收敛）。

对条件收敛的级数有一个重要性质，也即黎曼定理：一个条件收敛的级数，在其项经过适当的排列之后，可以收敛到一个事先任意指定的数；也可以发散到 $+\infty$ 或 $-\infty$；也可以没有任何的和。

函数级数

如果级数的每一项依赖于变量 x,x 在某区间 I 内变化, 即 $u_n = u_n(x), x \in$

I,$\sum u_n(x)$ 称为函数级数。

若 $x = x_0$ 使数项级数 $\sum u_n(x_0)$ 收敛，就称 x_0 为收敛点，由收敛点组成的集合称为收敛域，若对每一 $x \in I$，级数 $\sum u_n(x)$ 都收敛，就称 I 为收敛区间。

显然，函数级数在其收敛域内定义了一个函数，称之为和函数 $S(x)$，即 $S(x) = \sum u_n(x)$ 如果满足更强的条件，$S_m(x)$ 在收敛域内一致收敛于 $S(x)$。

$\sum a_n(x - x)0)^n$ 叫幂级数，收敛域是一个以 x_0 为中心的区间（不一定包括端点），并且在一定范围内具有类似多项式的性质，在收敛区间内能进行逐项微分和逐项积分等运算。例如幂级数 $\sum \dfrac{(2x)^n}{x}$ 的收敛区间是 $[-\frac{1}{2}, \frac{1}{2}]$，幂级数 $\sum[(x-2)^n]/(n^2)$ 的收敛区间是 [1，3]，而幂级数 $\sum(x^n)/(n!)$ 在实数轴上收敛。

不过实际上常用的级数是傅里叶级数（三角函数构成的级数），傅利叶级数的收敛范围一般很复杂，研究它需要对实变函数论、调和分析和泛函分析知识。所以真的理解并掌握傅利叶变换，不熟悉实变函数是没法入门的。

函数级数一致收敛定义：在一个集合 C 上一致地收敛到它的和函数 $s(x)$，是指对任意 $\epsilon > 0$，对于每一个正数级数都存在一个自然数 N（不依赖于 x），使得当 $m > N$ 时 $\left| s(x) = s_m(x) \right| = \left| r_m(x) \right| < \epsilon$，对于一切属于 C 的 x 都成立。

这时级数的和函数 $s(x)$ 是一个无限项的和，便可在整个集合 C 上通过特征性质继承有限项和的一些分析性质：

□ 逐项积分定理：设函数级数级数在有限闭区间 $a \leq x \leq b$ 上一致地收敛，若级数的各项 $s_N(x)$ 都连续，则级数的和也连续并且可以逐项积分。

□ 逐项微分定理：通过微分与积分的互逆关系（微积分基本定理）能够把上述定理转变成逐项微分的形式：设函数级数级数在区间 $a < x < b$ 内收敛，各项都具有连续的导数，若逐项取导数所得的级数在该区间内一致收敛，则原级数的和也具有连续的导数并且可以逐项微分。

函数级数收敛判定

显然下面一个主要问题是函数级数的收敛问题。因为一个函数级数在其收敛范围内代表一个函数，即它的和 $\sum u_n(x)(n = 1, \cdots, \infty) = u(x)$，当和是有限项时（$\sum u_n(x)(n = 1, \cdots, M)$, 这个级数和就是这个 u(x) 函数逐步逼近定义的一种方式。

在函数级数收敛研究过程中，经过约 200 年，才发现一致收敛概念的价值：

这种级数展开在收敛区间内可以逐项微分和积分并且收敛。

级数在逐项取绝对值之后就成为正项级数，显然可以依一致收敛性进行比较，特别是用一个常数级数进行比较，便有 M 判别法。

M 判别法（魏尔斯特拉斯判别法）：假设 u_n 是定义在集合 C 内的一个实数或复数函数的数列，并存在正的常数 M_n，使得 $|u_n(x)| \leq M_n$

对于所有的 $n \geq 1$ 和 C 内所有的 x 成立。进一步假设级数 $\sum M_n(n = 1, \cdots, \infty)$ 收敛。那么级数 $\sum u_n(x)(n = 1, \cdots, \infty)$ 在 C 内一致收敛。（可由常数项级数收敛的柯西准则证明）。

泰勒级数

我们常用的级数函数之一是泰勒级数。

泰勒级数定义：如果 f(x) 在点 x=x0 具有任意阶导数，则幂级数：

$\sum \frac{f^n(x_0)}{n!}(x - x_0)^n(n = 0, \cdots, \infty) = f(x_0) + f'(x_0)(x - x_0) + \frac{f^2(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f(n)(x_0)}{n!}(x - x_0)^n$

称为 $f(x)$ 在点 x_0 处的泰勒级数。

在上述定义中，取 $x_0 = 0$, 得到的级数 $\sum \frac{f^n(0)}{n!}(x)^n(n = 0, \cdots, \infty)$ 称为麦克劳林级数。函数 $f(x)$ 的麦克劳林级数是 x 的幂级数，那么这种展开是唯一的，且必然与 $f(x)$ 的麦克劳林级数一致。

如果 $f(x)$ 的麦克劳林级数在点的某一邻域内收敛，它不一定收敛于 $f(x)$。因此，如果 $f(x)$ 在某处有各阶导数，则 $f(x)$ 的麦克劳林级数虽然能算出来，但这个级数能否在某个区域内收敛，以及是否收敛于 $f(x)$ 还需要进一步验证。

一些函数无法被展开为泰勒级数，因为那里存在一些奇点。

定理一：设函数 f(x) 在 x0 的某个邻域 $N(x_0, \delta_0)$ 内具有任意阶导数，则函数 f(x) 在该邻域内能展开成泰勒级数的充要条件是的泰勒公式中的余项 $Rn(x)$ 满足 $\lim R_n(x) = 0, x \in N(x_0, \delta_0)$

定理二：如果 f(x) 在区间 $(-R + x_0, R + x_0)$ 能展开成泰勒级数

$\sum_{n=0,1,\cdots,\infty} \frac{f^n(a)}{n!}(x - a)^n$,

则右端的幂级数是惟一的。

泰勒级数的重要性质是在研究幂级数收敛过程中得到的：可以严密证明幂

级数在其收敛区间内展开式是唯一的，也即幂级数能够完全代表它的和函数参加分析运算（同时也证明了三角级数展开式不具有唯一性，所以三角函数的收敛集非常样复杂，这就是后来研究三角级数收敛性的学科调和分析能够成为数学主要学科的理由：问题复杂）。

由于幂级数可以逐项微分任意多次，所以幂级数本身就是它的和函数在收敛区间中心处的泰勒级数。所以一个泰勒级数的系数不一定要单纯通过累次微分级数而可以通过某些幂级数的分析运算来求得（因为微分次数越多计算越复杂）。

由于幂级数的求导和积分可以逐项进行，因此求和函数相对比较容易。这是泰勒级数最大的用处：简化计算。

同时，在复变函数中，一个解析函数可被延伸为一个定义在复平面上的一个开区域上的泰勒级数，这样可以简化和拓展解析函数定义方式。

不过在工程中，泰勒级数主要用来近似计算函数的值。

必须强调一点是，对于一些无穷可微函数 f(x)，虽然它们的展开式收敛，但是并不等于 f(x)。例如，分段函数：$f(x) = e^{-1}/x^2$，当 $x \neq 0$ 且 $f(0) = 0$，则当 $x = 0$ 所有的导数都为零，则这个 f(x)x=0 的泰勒级数为零，且其收敛半径为无穷大，虽然这个函数 f 仅在 x = 0 处为零。

下面给出几个常见函数在 x=0 处的泰勒级数，即麦克劳林级数。

指数函数：$e^x = \sum_{n=0,1,\cdots,\infty} \frac{x^n}{n!}$

自然对数：$ln(x+1) = \sum_{n=0,1,\cdots,\infty} (-1)^n + \frac{1}{nx^n}, x \in (-1,1]$

几何级数：$\frac{1}{1-x} = \sum_{n=0,1,\cdots,\infty} x^n, \left| x \right| < 1$

正弦函数：$sinx = \sum_{n=0,1,\cdots,\infty} \frac{(-1)^n}{(2n+1)!x^{2n+1}}$

余弦函数：$cosx = \sum_{n=0,1,\cdots,\infty} \frac{(-1)^n}{(2n)!x^{2n}}$

傅利叶级数

傅利叶级数或者傅利叶变换，是工程师的大杀器，对搞信号分析，模式识别的工程师，基本上就是居家旅游，吃饭睡觉的唯一工具。

由于傅利叶级数涉及很多波形图，豆瓣不支持，只能直观描述，有兴趣的

去查教科书，可能才能清楚我说的是什么。

傅立叶贡献：猜想周期函数都可以展开为常数与一组具有共同周期的正弦函数和余弦函数之和。(但是未能严格证明，拉格朗日就反对他的论文发表，认为不能三角级数表达梯形或箱型周期函数。后来狄利赫里证明了三角级数在一定条件下的收敛唯一性，并用级数连续逼近可以表达梯形或箱型周期函数)。

傅利叶级数展开式中，常数表达的部分称为直流分量，最小正周期等于原函数的周期的部分称为基波或一次谐波，最小正周期的若干倍等于原函数的周期的部分称为高次谐波。因此高次谐波的频率必然也等于基波的频率的若干倍，基波频率 N 倍的波称为 N 次谐波，是 N-1 次泛音。不管几次谐波，他们都是正弦波。正弦波是基本波形。

所以简单说：傅利叶级数就是周期函数展开为一个三角级数，例如：

给定一个周期为 T 的函数

$f(t) = a_0 + \sum_{n=1,\cdots,\infty}[a_n cos(n\omega t) + b_n sin(n\omega t)], \omega_0 = 2\pi/T, \omega t = 2\pi t/T$

根据欧拉公式，也可以等价表示为：

$f(t) = \sum_{k=-\infty,\infty} a_k e^{jk(2\pi/T)t}$,

j 为虚数单位, 其中 $ak = (1/T)\int_T f(t)e^{-jk(2\pi/T)}dt$

其中 $e^{jk(2/T}t$ 是周期为 T 的函数，所以 k 取不同值时的周期信号具有谐波关系（即它们都具有一个共同周期 T)。k=0 时，对应的这一项称为直流分量，k=1 时具有基波频率，称为一次谐波或基波，类似的有二次谐波，三次谐波等等。

欧拉公式：

$e^{jx} = cosx + jsinx$

$cosx = [e^{jx} + e^{-jx}]/2$

$sinx = [e^{jx} - e^{-jx}]/(2j)$

傅里叶最大的贡献是猜想了傅利叶级数的性质，而严格证明了傅利叶级数的收敛性则是狄利赫里。

狄利赫里定理：满足狄利赫里条件的周期函数表示成的傅里叶级数都收敛。狄利赫里条件如下：

在任何周期内，$f(t)$ 须绝对可积；傅里叶级数在任一有限区间中，$f(t)$ 只能取有限个最大值或最小值；在任何有限区间上，$f(t)$ 只能有有限个第一类间断

点。

定理结论是：满足狄利赫里条件的周期函数都可以展开为正弦函数和余弦函数的级数和，并且这个展开是收敛到唯一周期函数的。这是傅利叶变换的基础定理。

既然傅利叶猜想周期函数能够展开成三角函数的级数，那么三角函数的性质就很重要。三角函数最最重要性质是正交性，因为这是证明傅利叶级数收敛的唯一条件。

正交性定义：两个不同向量正交是指它们的内积为 0。（这也就意味着这两个向量之间没有任何相关性，例如，如果两个函数 $\phi_1(r)$ 和 $\phi_2(r)$ 满足条件：$\int \phi_1(r)\phi_2(r)d\tau = 0$, 则称这两个函数相互正交。在三维欧氏空间中，互相垂直的向量之间是正交的。事实上，正交是垂直在数学上的的一种抽象化和一般化）。

内积定义：设向量 $A = [a_1, a_2, \cdots, a_n], B = [b_1, b_2, \cdots, b_n]$，则向量 A 和 B 的内积表示为：

$A \cdot B = a_1 \times b_1 + a_2 \times b_2 + \cdots + a_n \times b_n$

$A = |A| \times |B| \times cos\theta$

$|A| = (a_1^2 + a_2^2 + \cdots + a_n^2)^{\frac{1}{2}}$

$|B| = (b_1^2 + b_2^2 + \cdots + b_n^2)^{\frac{1}{2}}$

其中，$|A|$ 和 $|B|$ 分别是向量 A 和 B 的模，θ 是向量 A 和向量 B 的夹角 ($\theta \in [0, \pi]$)。

若 B 为单位向量，即 $|B| = 1$ 时，$A \cdot B = |A| \times cos\theta$，表示向量 A 在 B 方向的投影长度。向量 A 为单位向量时同理。　当且仅当向量 A 与 B 垂直时，$A \cdot B = 0$。

显然，学过线性代数都知道，一组 n 个互相正交的向量必然是线形无关的，所以必然可以张成一个 n 维空间，也就是说，空间中的任何一个向量可以用它们来线性表出。

函数的正交是向量正交正交的推广, 函数可看成无穷维向量。

所谓三角函数系 $1, cosx, sinx, cos2x, sin2x, \cdots, cosnx, sinnx, \cdots$ 在区间 $[-\pi, \pi]$ 上正交, 就是指在三角函数系中任何不同的两个函数的乘积在区间 $[-\pi, \pi]$ 上的积分等于 0, 即 $\int_{-\pi}^{\pi} cos(nx)dx = 0$

$\int_{-\pi}^{\pi} sin(nx)dx = 0$

$\int_{-\pi}^{\pi} sin(kx) \times cos(nx)dx = 0$ 任意 k 和 n

$\int_{-\pi}^{\pi} cos(kx) \times cos(nx)dx = 0(k, n = 1, 2, 3, \cdots, k \neq n)$

$\int_{-\pi}^{\pi} sin(kx) \times sin(nx)dx = 0(k, n = 1, 2, 3, \cdots, k \neq n)$

$\int_{-\pi}^{\pi} cos(mx) \times cos(nx)dx = 0$（m≠n)

$\int_{-\pi}^{\pi} sin(nx) \times sin(nx)dx = \pi$

$\int_{-\pi}^{\pi} cos(nx) \times cos(nx)dx = \pi$

有了上述性质，傅利叶级数的展开就能大幅简化。

三角函数的另外一个重要性质是有奇偶性。

如果对于函数 $f(x)$ 的定义域内任意一个 x，都有 $f(-x) = -f(x)$，那么函数 $f(x)$ 就叫做奇函数；如果对于函数 $f(x)$ 的定义域内任意的一个 x，都有 $f(x) = f(-x)$, 那么函数 $f(x)$ 就叫做偶函数。奇函数可以表示为正弦级数，偶函数则可以表示成余弦级数。

也即奇函数 $f(x) = \sum b_k sin(kx)$（$k = -\infty, \infty$);

偶函数 $g(x) = \frac{a_0}{2} + \sum a_k cos(kx)$（$k = -\infty, \infty$)。

这些公式用欧拉公式，就可以很容易从上面傅里叶级数的公式中导出。

其实利用函数正交性，可以证明更一般的定理，那就是广义傅利叶技级数的收敛性。广义傅里叶级数是对一切正交函数系定义的，类比三角函数定义的傅利叶级数。

定义：任何正交函数系 $g(x)$，如果定义在 $[a, b]$ 上的函数 $f(x)$ 只具有有限个第一类间断点，那么如果

$f(x)$ 满足封闭性方程：$\int_a^b f^2(x)dx = \sum C_k^2(k = 1, \cdots, \infty)$,

那么级数 $\sum C_k g_k(x)(k = 1, \cdots, \infty)$ 必然收敛于 $f(x)$，其中：

$c_n = \int_a^b f(x)g_n(x)dx$

事实上，无论级数 $\sum C_k g_k(x)(k = 1, \cdots, \infty)$ 是否收敛，总有：

$\int_a^b f^2(x)dx \geq \sum C_k^2(k = 1, \cdots, \infty)$ (贝塞尔 (Bessel）不等式)。

这个性质经常用，可以大幅简化问题。

傅利叶变换和调和分析简介

傅利叶变换在物理学、电子类学科、数论、组合数学、信号处理、概率论、统计学、密码学、声学、光学、海洋学、结构动力学等领域都有着广泛的应用（例如在信号处理中，傅里叶变换的典型用途是将信号分解成频率谱——显示与频率对应的幅值大小）。

由于傅利叶变换的巨大用途（目前尚未有任何数学工具在实际工程和科学应用上可以与之相提并论），下面稍微多说几句。

□ 傅里叶变换定义

简单说，傅立叶变换是一种分析周期函数（例如信号）的方法，它可分析周期函数的频率成分或时变成分，也可用这些成分合成函数（或信号）。（虽然许多波形可作为函数（信号）的成分，比如正弦波、方波、锯齿波等，但是傅立叶变换用正弦波作为信号的成分，因为其容易计算）。

● 连续型傅利叶变换

常用的主要是连续型傅利叶变换。

连续型傅利叶变换的定义：$f(t)$ 是 t 的周期函数，如果 t 满足狄利赫里条件：在一个以 2T 为周期内 $f(X)$ 连续或只有有限个第一类间断点），且 $f(x)$ 单调或可划分成有限个单调区间，则 $F(x)$ 以 2T 为周期的傅里叶级数收敛，和函数 $S(x)$ 也是以 2T 为周期的周期函数，且在这些间断点上，函数是有限值；在一个周期内具有有限个极值点；绝对可积。

则有 $F(\omega)=\int_{-\infty}^{\infty}f(t)e^{-i\omega t}dt$ 称为积分运算 $f(t)$ 的傅立叶变换，即将频率域的函数表示为时间域的函数。

$f(t)=\dfrac{1}{2\pi}\int_{-\infty}^{\infty}F(\omega)e^{-i\omega t}d\omega$ 叫做 $F(\omega)$ 的傅立叶逆变换，即将时间域的函数表示为频率域的函数。

$F(\omega)$ 叫做 $f(t)$ 的像函数，$f(t)$ 叫做 $F(\omega)$ 的像原函数。$F(\omega)$ 是 $f(t)$ 的像。$f(t)$ 是 $F(\omega)$ 原像。

连续形式的傅里叶变换其实是傅里叶级数的推广，因为积分其实是一种极限形式的求和算子而已。对于周期函数，它的傅里叶级数表示被定义为：$f(t)=\sum F_n e^{jn(2/T)t}(n=-\infty,\infty)$,

其中 T 为函数的周期，Fn 为傅里叶展开系数，$F_n=\dfrac{1}{T}\int_{-T/2}^{T/2}f(t)e^{-in(2\pi/T)}dt$,

对于实值函数，函数的傅里叶级数可以写成：

$f(t) = a_0 + \sum[a_n cos(n\omega t) + b_n sin(n\omega t)](n = 1, \cdots, \infty), \omega_0 = 2\pi/T, \omega t = 2\pi t/T$

其中 a_n 和 b_n 是实频率分量的振幅。

当 f(t) 为奇函数（或偶函数）时，其余弦（或正弦）分量为零，而可以称这时的变换为余弦变换（或正弦变换）。

● 离散时间傅里叶变换

针对的是定义域为 Z 的数列。设 $x_n[-\infty, \infty]$ 为某一数列，则其离散时间傅里叶变换被定义为:$X(\omega) = \sum x_n e^{-i\omega n}(n = -\infty, \cdots, \infty)$；相应的逆变换为 $x_n = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{-i\omega t}d\omega$。

离散时间傅里叶变换在时域上离散，在频域上则是周期的，它一般用来对离散时间信号进行频谱分析。

● 离散傅里叶变换

离散函数且满足有限性或周期性条件，序列 $x_n[n = 0, \cdots, N-1]$ 的离散傅里叶变换为:$x[k] = \sum x_n e^{-i2\pi kn/N}(n = 0, \cdots, N-1)$；

其逆变换为：$x_n = \frac{1}{N}\sum X[k]e^{i2\pi kn/N}(k = 0, \cdots, N-1)$

傅里叶变换可以将计算复杂度降低（这个变换在数字电路计算，信号处理等等行业是十分实用且重要的方法）。

□ 傅利叶变换基本性质

● 傅里叶变换具有线性性质：假设函数 $f(x)$ 和 $g(x)$ 的傅里叶变换都存在，a 和 b 为任意常系数，则有 $F[af + bg] = aF[f] + F[g]$

● 尺度变换性质：若函数 $f(x)$ 的傅里叶变换为 $F(\omega)$，则对任意的非零实数 a，函数 $f_a(x\square = f(ax)$ 的傅里叶变换 $F_a(\omega)$ 存在，且等于 $F(a\omega) = \frac{1}{|a|}F(\omega/a)$。也即当 $a > 0$ 时，若将 f(x) 的图像沿横轴方向压缩 a 倍，则其傅里叶变换的图像将沿横轴方向展宽 a 倍，同时高度变为原来的 $1/a$。对于 a<0 时，傅里叶变换的图像关于纵轴做镜像对称。

● 对偶性质：若函数 $f(x)$ 的傅里叶变换为 $F(\omega)$，则存在 $F(x)$ 的傅利叶变换 $= 2\pi f(-\omega)$

● 平移性质：若函数 $f(x)$ 的傅里叶变换为 $F(\omega)$，则对任意实数 a ，函数 $f_a(x\square = f(x)e^{iax}$ 也存在傅里叶变换，且其傅里叶变换 $F_a(\omega) = F(\omega - a)$。也即 $F_a(\omega)$ 可由 $F(\omega)$ 向右平移 a 得到。

● 微分关系：若函数 $f(x)$ 的傅里叶变换为 $F(\omega)$，且其导函数 $f'(x)$ 的傅里叶变换存在，则有 $f'(x)$ 的傅利叶变换 $= i * \omega * F(\omega)$，也即导函数的傅里叶变换等于原函数的傅里叶变换乘以因子 $i * \omega$。更一般地，若 f(x) 的 n 阶导数的傅里叶变换存在，则 $f(x)$ 的 n 阶导数的傅里叶变换 $= (i * \omega)^n * F(\omega)$，即 n 阶导数的傅里叶变换等于原函数的傅里叶变换乘以因子 $(i * \omega)^n$。

□ 傅利叶变换基本定理

● 时域卷积定理

若函数 f(x) 的傅里叶变换为 $F(\omega)$，若函数 g(x) 的傅里叶变换为 $G(\omega)$，若函数 f(x),g(x) 都在 R 上绝对可积，则卷积函数 $f * g(x) = \int_{-\infty}^{\infty} f(x-t)g(t)dt$ 的傅里叶变换存在，且 $f * g(x)$ 的傅利叶变换 $= F(\omega) * G(\omega)$

● 频域卷积定理

若函数 $f(x)$ 的傅里叶变换为 $F(\omega)$，若函数 g(x) 的傅里叶变换为 $G(\omega)$，则有 $f(x) * g(x)$ 的傅利叶变换 $= \frac{1}{2\pi} * [F(\omega) * G(\omega)]$

□ 傅利叶变换用途广泛的原因

主要原因是傅利叶变换有下面这些优点。

● 傅里叶变换是线性算子, 若赋予适当的范数, 它还是酉算子;

● 傅里叶变换属于谐波分析;

● 傅里叶变换的逆变换容易求出, 而且形式与正变换非常类似;

● 正弦基函数是微分运算的本征函数, 从而使得线性微分方程的求解可以转化为常系数的代数方程的求解。在线性时不变杂的卷积运算为简单的乘积运算, 从而提供了计算卷积的一种简单手段;

● 离散形式的傅里叶的物理系统内（线性时不变）, 频率是个不变的性质, 从而系统对于复杂激励的响应可以通过组合其对不同频率正弦信号的响应来获取;

● 卷积定理指出：傅里叶变换可以化复杂的卷积运算为简单的乘积运算，从而提供了计算卷积的一种简单手段;

● 离散形式的傅立叶变换可以利用快速傅里叶变换算法 (FFT)。

□ 从信号分解的几何直观角度（波形图）来简单解释一下傅利叶变换的思想

傅利叶变换核心思想是用正弦曲线来代替原来的曲线而不用方波或三角波来表示。

原因在于分解信号的方法是无穷的，但分解信号的目的是为了更加简单地处理原来的信号。用正余弦来表示原信号会更加简单，因为正余弦拥有原信号所不具有的性质：正弦曲线保真度。一个正弦曲线信号输入后，输出的仍是正弦曲线，只有幅度和相位可能发生变化，但是频率和波的形状仍是一样的。且只有正弦曲线才拥有这样的性质，正因如此我们才不用方波或三角波来表示。

为什么选择三角函数而不用其他函数进行分解？因为很多现象可以抽象成一个线性时不变系统（也即输入输出信号满足线性关系，而且系统参数不随时间变换，无论用微分方程还是传递函数或者状态空间描述都可以），而且一个正弦曲线信号输入后，输出的仍是正弦曲线，只有幅度和相位可能发生变化，但是频率和波的形状仍是一样的。也就是说正弦信号是系统的特征向量，同时指数信号也是系统的特征向量（表示能量的衰减或积聚，衰减或者扩散现象大多是指数形式的，或者既有波动又有指数衰减，也即 e^{a+ib} 形式），所以除了指数信号和正弦信号以外的其他波形都不是线性系统的特征信号。

由于正弦信号是很多线性时不变系统的特征向量，于是傅里叶变换就有了用武之地。对于更一般的线性时不变系统，复指数信号 (表示耗散或衰减) 是系统的特征向量，于是拉普拉斯变换就有了用武之地。

显然，傅里叶级数和傅里叶变换就能处理特征值与特征向量的问题，这样用正余弦来表示原信号会更加简单，因为正余弦拥有原信号所不具有的性质：正弦曲线保真度。且只有正弦曲线才拥有这样的性质。

这也解释了傅利叶变换的强大用途的原因：因为正弦量 (或复指数) 是特征向量。

□ 傅里叶变换的推广

从数学的角度理解积分变换就是通过积分运算，把一个函数变成另一个函数。也可以理解成是算内积，然后就变成一个函数向另一个函数的投影：$F(s) =$

$\int_a^b f(t) * K(s,t)dt$，$K(s,t)$ 是积分变换的核 (Kernel)。

当选取不同的积分域和变换核时，就得到不同名称的积分变换（也即向核空间投影，将原问题转化到核空间。所谓核空间，就是这个空间里面装的是核函数）。

当然，选取什么样的核主要看面对的问题有什么特征。不同问题的特征不同，就会对应特定的核函数。把核函数作为基函数。将现在的坐标投影到核空间里面去，问题就会得到简化。之所以叫核，是因为这是最核心的地方。至于常用傅里叶变换和拉普拉斯变换是因为复指数信号才是描述这个现实世界的特征函数。

□ 傅利叶变换使用的一个例子

图像的频率是表征图像中灰度变化剧烈程度的指标，是灰度在平面空间上的梯度。如：大面积的沙漠在图像中是一片灰度变化缓慢的区域，对应的频率值很低；而对于地表属性变换剧烈的边缘区域在图像中是一片灰度变化剧烈的区域，对应的频率值较高。设 f 是一个能量有限的模拟信号，则其傅里叶变换就表示 f 的谱。

从物理效果看，傅里叶变换是将图像从空间域转换到频率域，其逆变换是将图像从频率域转换到空间域。换句话说，傅里叶变换的物理意义是将图像的灰度分布函数变换为图像的频率分布函数，傅里叶逆变换是将图像的频率分布函数变换为灰度分布函数。

傅里叶变换以前，图像（未压缩的位图）是由对在连续空间（现实空间）上的采样得到一系列点的集合，用一个二维矩阵表示空间上各点，则图像可由 $z = f(x,y)$ 来表示。由于空间是三维的，图像是二维的，因此空间中物体在另一个维度上的关系就由梯度来表示。

对图像进行二维傅里叶变换得到频谱图，就是图像梯度的分布图，当然频谱图上的各点与图像上各点并不存在一一对应的关系，即使在不移频的情况下也是没有。

从傅里叶频谱图上看到的明暗不一的亮点，实际上图像上某一点与邻域点差异的强弱，即梯度的大小，也即该点的频率的大小（图像中的低频部分指低梯度的点，高频部分相反）。梯度大则该点的亮度强，否则该点亮度弱。

傅里叶变换后的频谱图，也叫功率图。

可以看出，图像的能量分布，如果频谱图中暗的点数更多，那么实际图像是比较柔和的（因为各点与邻域差异都不大，梯度相对较小），反之，如果频谱图中亮的点数多，那么实际图像一定是尖锐的，边界分明且边界两边像素差异较大的。

对频谱移频到原点以后，可以看出图像的频率分布是以原点为圆心，对称分布的。将频谱移频到圆心可以分离出有周期性规律的干扰信号（带有正弦干扰，移频到原点的频谱图上可以看出除了中心以外还存在以某一点为中心，对称分布的亮点集合，这个集合就是干扰噪音产生的），这时可以很直观的通过在该位置放置带阻滤波器消除干扰。

图像经过二维傅里叶变换后，其变换系数矩阵表明若变换矩阵 Fn 原点设在中心，其频谱能量集中分布在变换系数短阵的中心附近，若所用的二维傅里叶变换矩阵 Fn 的原点设在左上角，那么图像信号能量将集中在系数矩阵的四个角上。这是由二维傅里叶变换本身性质决定的。同时也表明一股图像能量集中低频区域。

变换之后的图像在原点平移之前四角是低频，最亮，平移之后中间部分是低频，最亮，亮度大说明低频的能量大（幅角比较大）。

这样通过傅利叶变换，就能简化计算，识别图像特征。

□ 调和分析简介

傅里叶分析从诞生之日起，就围绕着“傅里叶级数究竟是否收敛于自身”这样一个中心问题进行研究，这也是调和分析的中心问题。

可以说调和分析就是傅里叶分析，调和分析是研究作为基本波形的叠加的函数或者信号的表示的数学分支。它研究并推广傅立叶级数和傅立叶变换的概念。基本波形称为调和函数，和分析因此得名。主要用途是信号处理、量子力学、模式识别、人工智能、神经科学等等。

当初傅里叶只是提出周期函数可用三角级数表示的猜想，并未证明。是狄利克雷给出了周期函数的傅里叶级数收敛于它自身的充分条件：一个周期上分段单调的周期函数的傅里叶级数，在它的连续点上必收敛于 $f(x)$；如果在 x 点不连续，则级数的和是 $(f(x+0)+f(x-0))/2$。

狄利克雷的定理表明：函数在一个周期内的分段单调性，可能导致该函数在不同区间上的不同解析表示，这自然应当把它们看做同一个函数的不同组成部分，而不是像当时人们所理解的那样，认为一个解析表达式就是一个函数。这是对函数概念的一大突破和重大贡献。

黎曼对傅里叶级数的研究也作出了贡献。黎曼在 1854 年《用三角级数来表示函数》论文中，引进了现在称为黎曼积分的概念及其性质，证明了如果周期函数 (x) 在 $[0, 2\pi]$ 上有界且可积，则当 n 趋于无穷时的傅里叶系数趋于 0；有界可积函数的傅里叶级数在一点处的收敛性，仅仅依赖于 (x）在该点近旁的性质。这是一个本质定理，现在称之为局部性原理。

海涅在 1870 年证明：有界函数 $f(x)$ 可以唯一地表示为三角级数，但是由于证明不完备（因为傅里叶级数未必一致收敛，从而无法确保逐项积分的合理性，逻辑上就可能存在不一致收敛的三角级数。但这个级数确实表示一个函数），导致了康托研究函数用三角级数表示是否唯一的问题的由来：为此康托引进了点集的极限点以及导集等概念，这导致了实变函数的诞生。

魏尔斯特拉斯在 1861 年首次利用三角级数构造了处处不可求导的连续函数。他的这一发现震动了当时的数学界，因为长期的直观感觉使人们误认为，连续函数只有在少数一些点上才不可求导。

这个发现，直接导致了勒贝格积分和点集测度理论诞生。（勒贝格积分与勒贝格测度，现在已成为数学各分支中不可缺少的重要概念和工具）。勒贝格利用勒贝格积分和点集测度把黎曼的工作又推进了一步，得到如下重要结果：任何勒贝格可积函数的傅里叶级数，不论收敛与否，都可以逐项积分；对于 $[0, 2\pi]$ 上勒贝格平方可积的函数，帕舍伐尔等式成立（$||x||^2 = \sum |(x, ek)|^2$）。

连续函数的傅里叶级数，是否必处处收敛？1876 年杜布瓦－雷蒙发现，存在连续函数，它的傅里叶级数在某些点上发散；后又证明，连续函数的傅里叶级数可以在一个无穷点集上处处发散。这反面结果的发现提醒人们对傅里叶级数的收敛性应持审慎态度。这些重要发现，导致了对傅利叶级数收敛性质的进一步探讨，结果成果越来越多，最后形成现代数学一个主要分支：调和分析。

另外一个研究傅里叶级数收敛的方向是复变函数论方法。因为傅里叶级数的指数函数表达式可以看成单位圆内的解析函数（可积函数的傅里叶级数它是

复变量 z 的幂级数的实部。所以复变函数论是研究傅里叶级数的一个重要工具。

利用复变函数，哈代和里斯兄弟建立单位圆上 H 空间的理论。

50 年代以前，傅里叶分析的研究领域基本上限于一维的具体空间，50 年代以后的研究，逐渐向多维和抽象空间推广，例如傅立叶级数在希尔伯特空间的意义研究，以此建立了与泛函分析的一个联系；再例如基于拓扑群上的函数或测度以及由它们构成的空间或代数来研究傅利叶变换；以及考尔德伦－赞格蒙奇异积分理论，伯克霍尔德的一般 H 空间理论，以及群上的傅里叶分析。

从群论的观点看，无论是周期函数还是非周期函数，它们的定义域都是拓扑群 G（要在群上运用傅里叶分析方法，先就要能在群上定义傅里叶变换，外与彼得合作对一般紧李群建立了外尔-彼得定理, 奠定了紧群上调和分析的基础，哈尔对满足某些条件的局部紧群证明了特殊测度（哈尔测度）的存在性），就是说，G 有一个代数运算，称为群运算，以及与之相协调的极限运算，称为 G 的拓扑。傅里叶级数或傅里叶积分的任务，正是研究 G 上定义的函数 $f(x)$ 分解为群上许多“特殊”函数（例如 e 或 e）之和的可能性，以及通过傅里叶系数或傅里叶变换来研究自身的性质。例如，以乘法为群运算的全体正实数构成一拓扑群 R，它的拓扑就是欧氏空间的拓扑，那么测度 $d\mu = xdx$ 就是 R 上的哈尔测度。可以证明对于群 R 上的可积函数 $f(x)$ 的傅里叶变换收敛性。群代数、测度代数、傅里叶代数、傅里叶－斯蒂尔杰斯代数这些是群上调和分析最主要的研究对象。

群论观点的引入，使得隐藏在周期性函数现象背后的内在联系，被揭示得更清楚更深刻了，使得调和分析内部各分支之间以及调和分析与其他学科例如泛函分析、代数学、群表示论、模形式等的联系变得更为密切。因此，群上调和分析可以说是一门既具应用价值（正如它对概率论、数论与微分方程等所起的作用所说明的）又具理论意义的综合性学科。

有兴趣的人可以去查调和分析教科书。

3.2.7.9 现代微积分

上面介绍的微积分是最古典的微积分，也即是 18 世纪 19 世纪的微积分，也是现在绝大多数理工科院校教授的微积分，他们特点是涉及的微分，积分和级数讨论的函数的自变量定义的区域基本都是一维的直线，二维的曲线和三维的曲面。但是现在好的大学数学系的本科学生学的微积分是现代微积分，例如哈佛。普林斯顿、耶鲁、斯坦福、MIT、加州理工等等的数学系，当然中国科大数学系本科的微积分也是现代微积分。

现代微积分的标志之一就是从流形上的微积分开始。古典微积分讨论对象–函数的定义一般都是实数数域上的一个映射。而现代微分讨论的对象是流形上的映射。

流形（Manifold）的概念最早是在 1854 年由 Riemann 提出，简单说就是局部具有欧氏空间性质的空间，一般可以直观理解流形是把许多平直的片折弯并粘连而成。它是数域概念的推广，例如可以象数域一样定义距离（一般用测度表示，包括面积体积等等概念），定义方向（包括向量场），定义运算（包括各种算子，变换，内积等等），以及定义流形上的微分和积分。

流形是一个几何概念，流形是任意维度的抽象空间，简单说流形包括各种维数的曲线曲面。(线段是一维的，曲面是二维的，三维空间中的所有旋转是三维的)。微积分研究其可微性。流形概念来自于物理，例如经典力学的相空间和构造广义相对论的时空模型的四维伪黎曼流形都是流形的实例。

欧几里得空间就是流形最简单的实例，像地球表面这样的球面也是一个流形。

欧几里德空间是一个特别的度量空间，定义如下：设 V 是实数域 R 上的线性空间（向量空间），若 g 是 V 上的二元实值函数，满足如下关系：

- $g(x,y)=g(y,x)$
- $g(x+y,z)=g(x,z)+g(y,z)$
- $g(kx,y)=kg(x,y)$
- $g(x,x)\geq 0$,

而且 $g(x,x)=0$ 当且仅当 $x=0$ 时成立。

这里 x,y,z 是 V 中任意向量，k 是任意实数。

内积空间是对欧氏空间的一般化（一个线性空间定义了内积运算之后就是欧几里德空间，向量空间又称线性空间，是线性代数的中心内容和基本概念之一）。内积空间和度量空间都是泛函分析的基本研究对象。

举几个经典欧几里德空间例子：

E^n：在 n 维实向量空间 R^n 中定义内积 $(x,y)=x_1y_1+\cdots+x_ny_n$，则 R^n 为欧几里德空间。（任意一个 n 维欧几里德空间 V 等距同构于 E^n）

设 V 是 $[0,1]$ 区间上连续实函数全体，则 V 是 R 上线性空间，对于如下内积是欧几里德空间：(f,g) 定义为 fg 在 $[0,1]$ 区间上的积分值。

简单来讲，流形上的微积分课程首先得介绍欧几里得空间性质，包括范数、线性变换和连续映射，然后介绍可微映射及其导数和逆映射定理，然后要介绍欧几里得空间上的可积函数的特征，然后介绍微分流形特征，及其微分形式和外微分，流形上的积分，斯托克斯公式等等。由于具体内容有太多数学符号，在豆瓣无法表达，例如流行上的微分就涉及流形映射的雅可比矩阵，所以只能简单提一提，具体内容，有兴趣的可以找书来看，现在物理学和一些应用科学，例如模式识别，不懂流形微积分，基本没法玩。

就我的经历来看，在大学本科讲流形上的微积分，对大多数学生来讲毫无难度。淘汰数学系学生的两道门槛，一道是 $\epsilon-\delta$ 语言体系（也包括充分必要条件，逻辑完备性，反例等等概念）的充分理解和把握，一道是代数结构抽象语言（也包括同构，同态，同胚等等概念）的充分理解和把握。一般抽象能力和逻辑能力跟不上的学生，都会在这两道门槛前被淘汰。

显然根据前面介绍，我们知道微积分学的基础概念其实是无限，传统微积分把无限当成一个过程（也即维斯特拉斯用 $\epsilon-\delta$ 语言公理体系定义的极限的概念），但是在不断发现不连续函数，不可积函数后（其实微分方程很多解的函数都不是初等函数，很多都只能用级数表达或间断函数表达，这里面就有许多不可积或不可微，甚至连续不可微的函数，例如魏尔斯特拉斯函数就是一类处处连续而处处不可导的实值函数。魏尔斯特拉斯函数是一种无法用笔画出任何一部分的函数，因为每一点的导数都不存在，画的人无法知道每一点该朝哪个方向画。魏尔斯特拉斯函数的每一点的斜率也是不存在的。在魏尔斯特拉斯这

个反例出现之前，数学家们认为除了少数一些特殊的点以外，连续的函数曲线在每一点上总会有斜率），所以传统微积分逻辑基础需要的公理体系就存在了漏洞，因为这个公理体系假设了无限是一个具体的东西，一种真实的存在，但是很多反例在质疑这种假设。为了弥补这个漏洞，就自然出现了进一步研究实数无限性质的必要，这就是下面我们要介绍的实变函数的内容。

3.2.7.10 证明解存在性的逻辑价值的一个例子

对工程师来讲，能够用来计算的数学工具才是有用的，由于微积分强大的计算能力，就成为了必备工具（另外工程师常用的还有线性代数和数理统计），但是从逻辑角度来讲，证明解的存在性重要性比计算解要有价值得多，因为去计算不存在的解是无用功。举例来讲，微观经济学里面的阿罗–德布鲁边际均衡模型证明的结论：一定假设下，供需曲线一定相交于一点（也即均衡价格是存在的），这个结论是整个微观经济学的逻辑基础，没有这个存在性证明，微观经济学其实不是科学，只是假设和幻想。这个基础定理不但可以引申证明微观经济学的一些核心概念的正确性，例如边际效益递减（也即增长是有极限的），最优增长路径存在（也即著名的萨缪尔森大道定理）和经济体系（不管开放或封闭），只要资源约束条件是凸集，就有多目标非劣解等等。这些定理就是存在性定理的杰作。

当然我们非经济经济专业人士一般知道阿罗不是因为这个均衡定理，而是另外一个存在性定理：阿罗不可能性定理。这个定理摧毁了福利经济学的基础，也摧毁了绝对公平信奉者的逻辑基础。

下面我们先简单介绍一下阿罗和德布鲁的成就，然后介绍阿罗德布鲁一般均衡，最后介绍阿罗不可能定理。

简单介绍一下阿罗和德布鲁，这是两个学经济学无法绕开的高峰，张五常就是阿罗的铁粉，因为阿罗可以迅速把一切经济问题变成数学问题。

阿罗 (Kenneth Arrow)（1972 年诺贝尔经济学奖）是新古典经济学的开创者之一。除了在一般均衡领域的成就之外，阿罗还在风险决策、组织经济学、信息经济学、福利经济学和政治民主理论方面进行了创造性的工作。

德布鲁（法国数学家和数理经济学家，因为均衡定理证明 1983 年获诺贝尔经济学奖），他的工作改写了现代数理经济学，他最重要的贡献是与阿罗合作，联名发表了一篇具有划时代意义的文章《竞争性经济中均衡的存在》（1954）。在这篇文章中，运用拓扑学方法，对一般均衡的存在提供了数学证明。他获得诺贝经济学奖的成果一共只有 102 页：《价值理论：对经济均衡的公理分析》，他开创了一种研究解决问题的先河：德布鲁用集合论和凸性分析来研究均衡问题，彻底摆脱了一般均衡理论主要运用代数和方程的传统，从而彻底解决了亚当·斯密、瓦尔拉斯以来的一般均衡理论只是假设或直觉的逻辑基础（瓦尔拉斯利用代数和方程企图证明一般均衡存在，但是证明被验证是逻辑错误的，因为这种方法本身存在循环假设，无法内在地解决均衡的存在性这一基本问题）。

德布鲁在这 102 页的证明中，开创了以下概念：资源未被充分利用的度量；帕累托的最优的数学表达（用了数学中的超平面分离定理，在一般意义上建立了价格系统效用最大化配置和帕累托最优配置这两个概念之间的等价性）；相关商品的均衡存在性（一般竞争均衡理论）；用效用函数表示偏好次序关系；总量超额需求函数（效用的需求理论）；经济核算的收敛定理等。

德布鲁的的每一篇文章都给出与经济学核心相关的公理证明，轻松地证明了一个又一个均衡定理。德布鲁 114 页的《价格理论》奠定了新古典经济学的框架，书中用一般均衡理论讨论了商品、价格、消费等概念的实质意义，还把阿罗刚拓展的不确定环境中的资源配置问题纳入书中。德布鲁还是纳什均衡（因此获得 1994 年诺贝经济学奖）的先驱，因为通过讨价还价来决定资源配置，最终也会有一个均衡解，就是德布鲁开始研究的（纳什的成果直观讲就是：在一个复杂经济系统中，每个人根据市场统一的价格进行交易和各人各自讨价还价形成价格的机制是完全不同的。如果价格不同，资源配置的效率自然也就不同。纳什证明：大规模讨价还价最终形成的价格之“核”，是一个不动点，也即存在一般均衡，也即讨价还价是能够实现资源配置的）。

现在德布鲁的成果，已成为微观经济理论的统一构架。他使用的公理化分析方法已成为经济分析的标准形式。

现在介绍阿罗德布鲁模型。

一般均衡理论是经济学理论的核心。一般均衡概念来自于亚当·斯密，也

即“看不见的手”：市场会通过价格调整，自动找到供需平衡（直观讲，就是供需曲线一定会相交于一点）。

可是，经济学家们一直没法证明这个均衡点是存在的。

十九世纪末，瓦尔拉斯企图用线性代数来证明这一均衡点存在。但是这个企图失败了。直到 1954 年，阿罗–德布鲁在《计量经济学》杂志上发表了一篇题为“竞争经济的存在性均衡”论文，提出了阿罗–德布鲁一般均衡模型，用集合论作为基本工具，对经济体制进行了结构抽象，通过一些假设条件，证明了一般均衡的存在，从逻辑上验证了亚当 • 斯密“看不见的手”的猜想，这是整套新古典经济学的根基。从此，理论经济学被视为科学。阿罗和德布鲁的证明方法简单来说就是：引入一个虚构的市场主体来选择价格体系，从而将给定的经济体系问题转化为一个一般化博弈的均衡存在性问题。

阿罗–德布鲁的证明要用到布劳渥 (Brouwer) 不动点定理：如果 f 是 n + 1 维实心球 $B^{n+1} = \{x \in R^{n+1}, |x| \leq 1\}$ 到自身的连续映射（$n = 1, 2, 3, \cdots$），则 f 存在一个不动点 $x \in B^{n+1}$（即满足 $f(x_0) = x_0$）。（简单直观说就是任何封闭单位点的连续函数在 n 维欧几里德空间本身必须有一个不动点）。

这个定理广泛应用于代数方程、微分方程、积分方程等的求解问题，在数学上非常重要，也是微观经济学的基础定理。

这个定理与哥德尔不完全定理，是一切复杂问题解决的大杀器，建议熟悉。例如霍金再《哥德尔与 M 理论》中就认为，从哥德尔不完全定理出发，建立一个单一的描述宇宙的大统一理论是太可能的。现在人工智能也把哥德尔不完全性定理当成基础定理：根据哥德尔不完全性定理，机器不可能具有人的心智。（哥德尔定理的简单表述是：任意一个包含一阶谓词逻辑与初等数论的形式系统，都存在一个命题，它在这个系统中既不能被证明也不能被否定。或者我们可以这样直观理解：我们永远不能发现一个万能的公理系统能够证明一切数学真理，而不能证明任何谬误。哥德尔不完全性定理的价值是：真与可证是两个概念，可证的一定是真的，但真的不一定可证）。

阿罗–德布鲁构造的模型包括若干假设（完全是描述性的，不具有严格意义）：

假设 1：假设存在 L 种商品，商品的数量为实数，且这里的商品是指最佳划

分的商品类，即进一步增加商品类别的划分，由此产生的消费分配并不能增加消费者的满足（保证产出集合凸性假设）。

假设 2：假定存在 H 个消费者，每一个消费者的偏好是一个完备的、连续的、传递的次序，且消费者偏好是非充分满足性和凸性，每一个当事人被赋予拥有每一个厂商的股权（保证消费集合的性假设）。

□ 集合凸性定义：集合 S，对于任意 x，y□S，存在 a□[0，1] 使得 ax+(1-a)y□S。

假设 3：假设存在 J 个厂商，厂商的生产计划是可行的且能自由决策（这个假设排除了产品的不可分性、规模收益递增和从专业化中获得收益等，且从厂商生产和消费者来说，商品不被加以区分等）。

在上面假设下，阿罗–德布鲁证明：有一组确定的解能够同时满足一般均衡方程组，并且在总量水平上，供给与需求同时均等地决定价格（也即一般均衡状态在完全竞争经济中是可以达到的，并且使之达到均衡状态的价格和产量不是唯一的，只有相对价格的变化才影响消费者、厂商和要素拥有者的决策。如果所有市场在一组价格下处于均衡状态，那么所有这些价格都以同样比例上升或下降后，这些市场仍然处于均衡状态）。

这个定理证明有兴趣的可以去查书，因为涉及太多数学符号，就不介绍了。

阿罗-德布鲁证明一般均衡最主要的假设是：消费与生产集合都是凸集，每个经济主体都拥有一些由其它经济主体计值的资源。

阿罗–德布鲁模型中不需要有固定的生产系数，也不必有一致的利润率，没有股票市场，因为股票不是阿罗-德布鲁商品，模型也不存在企业破产的问题。因为所有的经济主体的生产和消费行为都必须符合预算约束，一旦超过预算，就对其实施无限破产处罚，模型中货币对资源配置没有实质的影响（但是现实中存在货币的理由：交易需求、预防需求、价值贮藏、记价单位等，在阿罗-德布鲁模型中都已经顾及）。

下面介绍阿罗不可能定理。

阿罗不可能性定理的直观表述是：如果众多的社会成员具有不同的偏好，而社会又有多种备选方案，那么在民主的制度下不可能得到令所有的人都满意的结果。

阿罗的课题是：投票选举方式能否保证产生出合乎大多数人意愿的领导者

或者说将问题简化为：“将每个个体表达的先后次序综合成整个群体的偏好次序”。

阿罗采用数学的公理化方法进行分析，结论是：绝大多数情况下是不可能的。更准确的表达：当至少有三名候选人和两位选民时，不存在满足阿罗公理的选举规则，也即随着候选人和选民的增加，“程序民主”必将越来越远离“实质民主”。

所以阿罗给出了被西方经济学界称为不可思议的定理：如果众多的社会成员具有不同的偏好，而社会又有多种备选方案，那么在民主的制度下不可能得到令所有的人都满意的结果。也即少数服从多数民主原则并不能将个人的偏好汇集成社会的偏好。

简单说明：假如有一个非常民主的群体，或者说是一个希望在民主基础上作出自己的所有决策的社会，对它来说，群体中每一个成员的要求都是同等重要的。一般地，对于最应该做的事情，群体的每一个成员都有自己的偏好。为了决策，就要建立一个公正而一致的程序，能把个体的偏好结合起来，达成某种共识。这就要进一步假设群体中的每一个成员都能够按自己的偏好对所需要的各种选择进行排序，对所有这些排序的汇聚就是群体的排序了。

定理简单说明：假设甲乙丙三人，面对 ABC 三个备选方案，偏好排序如下：甲 (a>b>c)；乙 (b>c>a)；丙 (c>a>b)（甲 (a>b>c) 代表–甲偏好 a 胜于 b，又偏好 b 胜于 c 等等）。

1、若取 a，b 对决，那么按照偏好次序排列如下：甲 (a>b)；乙 (b>a)；丙 (a>b)；社会次序偏好为 (a>b)；2、若取 b，c 对决，那么按照偏好次序排列如下：甲 (b>c)；乙 (b>c)；丙 (c>b)；社会次序偏好为 (b>c)；3、若取 a，c 对决，那么按照偏好次序排列如下：甲 (a>c)；乙 (c>a)；丙 (c>a)；社会次序偏好为 (c>a)

于是得到三个社会偏好次序–(a>b)、(b>c)、(c>a)，其投票结果显示社会偏好有如下事实: 社会偏好 a 胜于 b；偏好 b 胜于 c；偏好 c 胜于 a。显而易见，这种所谓的社会偏好次序包含有内在的矛盾，即社会偏好 a 胜于 c，而又认为 a 不如 c。所以按照投票的大多数规则，不能得出合理的社会偏好次序。

所以依靠简单多数的投票原则，要在各种个人偏好中选择出一个共同一致的顺序，是不可能的。这样，一个合理的公共产品决定只能来自于一个可以胜

任的公共权利机关，要想借助于投票过程来达到协调一致的集体选择结果，一般是不可能的。

证明阿罗不可能性定理需要利用 May 关于完美投票的定理。

May 完美投票定义：

□ 没有弃权票（也即选票定义无限制）；

□ 两个投票者相互换票, 获胜者不变（选票对称且匿名）；

□ 投票系统对候选人平等（也即当所有投票者换投另一人, 相应的胜出候选人也会改变。例如简单多数胜出就是一个平等的投票系统）；

□ 单调性（也即胜出者不会因为得到更多的票而失去胜利，失败者不会因为失去票而得到胜利）。

May 定理：如果有奇数个投票人，那么多数胜出投票是唯一符合上述 4 条规定的投票系统。

May 定理的数学形式：设 N 是一个有限集 (全体投票者, 假设有 2n+1 个人), N 的全体子集记为 P(N),0,1 为两点集, 表示候选人, 投票为函数 v: N →0,1, 记 V 为 v 的 1 原像集, 即投候选人 1 的投票人全体, 则 V 是 N 的一个子集合，反过来也一样, 即 N 的任何子集合都可能作为投候选人 1 的全部投票人。以 |V| 记 V 中的人数。投票系统为集合函数 f:P(N) → 0,1 (集合函数即集合的子集合类上的函数)

多数胜出投票系统指: 对任意 V□P(N), |V|>n 时, f(V)=1, |V|≤n 时, f(V)=0

将 N 排序, 以 π 记 N 的一个位置置换 (重新排序), 排好顺序的 N 的一个 A, 以 πA 记 A 在 π 的换位下得到的新的 N 的子集合，定义域无限制相当于 v 的定义域是 N。假设满足下面三个条件：

□Neutrality：A□P(N), Ac 记 A 在 N 中的余集 (补集), 则 f(Ac) = 1-f(A)。

□Monotone：任意 A, B □P(N), A 包含于 B 中, 则有 f(A)≤f(B)。

□anonymous：f(πA)=f(A)

则有下属结论（May 定理）: 在有奇数个投票者即 |N|=2n+1 时, 多数优胜系统 f, 是唯一满足上述三条性质的集合函数。

根据上述定义，有两个特殊的投票系统或者集合函数，一个是独裁系统，即除了一个投票人之外，所有的票都被无视，相当于集合函数只在包含这一个人

的子集合上非 0。另外一个是强加系统，也即存在一个候选者无视所有的投票，计票输出函数强行赋值给 1，也即投票系统函数是平凡的集合函数。

显然根据上述 May 定理，只需要简单改造，就能证明阿罗不可能定理，例如：

□ 投票系统仍然为集合函数，但函数的取值为候选人的一种名次排列 (全序)，要求每个投票人给出一种候选人的全序，投票系统必须给出候选人的全序，同时候选人平等（也即候选人置换后，投票系统函数输出的全序不变）；

□ 传导性 (把 May 假设从 Neutrlity 换成 Consensu) 如果对两个候选人，每个投票人都给出同样的序关系 A>B，则投票系统对此二候选人给出同样的序关系 A>B；

□ 第三方独立性（即投票系统输出两个候选人的序关系，独立于其他候选人）；

□ 非独裁性（即如果有一个投票者 (独裁者)，同意 A>B，其余投票者都相反，则投票系统输出中，A<B）。

这样假设改进后，极易证明阿罗不可能定理：不存在这样的投票系统，满足上述四个条件，还能输出一个全序出来。

阿罗证明的思路：假设群体 S 上有 m 个个体成员，群体中出现的各群体种事件构成一个集合 X，每个个体对每一事件都有自己的态度，即每个人都对集合 X 有一个偏好关系 >i=1，2，…，m。即可以按自己的偏好为事件排序。

定义群体的偏好为：其中 P 是一种由每个个体偏好得出群体偏好的规则。按这个规则从个体排序 (偏好) 得到群体排序 (偏好)，而且这个排序符合民主社会的民主决策的各种要求。注意这个排序是自反的，即如果 A>B，那么，B<A；是传递的，即如果 A>B，B>C，则有 A>C；并且还是完全的，即要么 A>B，要么 B>A，二者只有其一而且必有其一。这首先要考察一下民主社会的民主决策的各种要求是什么，阿罗用 4 个公理表述出这些要求。

公理 1：个体可以有任何偏好；而且是民主选择–每个社会成员都可以自由地按自己的偏好进行选择；

公理 2：不相干的选择是互相独立的；

公理 3：社会价值与个体价值之间有正向关联；

公理 4：没有独裁者–不存在能把个体偏好强加给社会的可能。

阿罗证明，满足这 4 条公理表述的要求的民主决策的规则是不存在的：如果 X 中的事件个数不小于 3，那么就不存在任何遵循原则 U，P，I，D 的规则 (称为社会福利函数)。这表明满足所有一般条件的民主选择要么是强加的，要么就是独裁的结果。

换句话说，阿罗不可能性定理指出，多数规则的一个根本缺陷就是在实际决策中往往导致循环投票。

在得多数票获胜的规则下，每个人均按照他的偏好来投票。不难看出，大多数人是偏好 X 胜于 Y，同样大多数人也是偏好 Y 胜于 Z。按照逻辑上的一致性，这种偏好应当是可以传递的 (transitivity)，即大多数人偏好 X 胜于 Z。但实际上，大多数人偏好 Z 胜于 X。因此，以投票的多数规则来确定社会或集体的选择会产生循环的结果。结果，在这些选择方案中，没有一个能够获得多数票而通过，这就是" 投票悖论"，它对所有的公共选择问题都是一种固有的难题，所有的公共选择规则都难以避开这两难境地。

那么，能不能设计出一个消除循环投票，做出合理决策的投票方案呢? 阿罗的结论是: 根本不存在一种能保证效率、尊重个人偏好、并且不依赖程序 (agenda) 的多数规则的投票方案。简单地说，阿罗的不可能定理意味着，在通常情况下，当社会所有成员的偏好为已知时，不可能通过一定的方法从个人偏好次序得出社会偏好次序，不可能通过一定的程序准确地表达社会全体成员的个人偏好或者达到合意的公共决策。

证明过程：

证明 2 阿罗的不可能定理证明过程 第一步：假设可以找到一个决定群体, 或者独断群体, 这个群体如果一致投 $A>B$, 投票系统将输出 $A>B$（这个群体类似独裁系统，他把这个群体叫做 *quorum*）；

第二步：由于候选人的平等性条件，*quorum* 与具体的候选人 A, B 无关；

第三步：如果投票人子集合 S，T 都是 *quorum*, 则 $S\cap T$ 也是 *quorum*；

第四步：从全体投票人去掉随便一个人，由于非独裁性条件, 这个集合必然是一个 *quorum*；

结论：所有这些 *(*全体投票者中去掉一个人*)* 的 *quorum,* 相交也应该为 *quorum,* 但是为空集，这与与传导性条件 *2* 矛盾。证毕。

这个定理曾经震惊西方，因为证明一个社会不可能有完全的每个个人的自由–否则将导致独裁；一个社会也不可能实现完全的自由经济–否则将导致垄断。也即证明了社会福利曲线不可能存在。

阿罗的不可能定理对福利经济学是毁灭性的。因为福利经济学假设公平是存在的。

所以当年一大批经济学家痛恨阿罗，因为敲掉人家饭碗了，李特尔，萨缪尔森等等领衔的上百篇文章对他的定理进行驳斥。但是阿罗的不可能性定理经受住了所有技术上的批评，其基本理论从来没有受到重大挑战，可以说无懈可击。所以后来这种攻击也就不了了之。毕竟萨缪尔森也是诺贝经济学奖获得者（1970 年诺贝经济学奖获得者，第一个获得经济学奖的美国人），而且是数理经济学的先驱人物，是第一个把数学分析应用于经济学研究中的大家，数学也是不错的。不然也证明不了大道定理这种完全超出人直觉的东西。

大道定理简单说就是：在多部门经济体系中存在着多条均衡增长路径, 其中增长速度最快的一条称为“大道增长路径”或称为“冯 • 诺依曼路径”。形象说法就是：当我们要尽快从 A 点走到 B 点时，最好的办法就是是先从 A 点走到高速公路上，然后高速向前，最后才离开高速公路而到达 B 点。这条高速公路就是最有增长路径。萨缪尔森把制度变量作为约束条件，建立多部门多变量的经济增长模型，利用变分法的的欧拉方程和最优控制的庞特里亚金最大值原理，求出其一阶条件，并解出最优路径，并证明最优路径是收敛的，这就证明了在多条均衡增长路径中，存在一条最优增长路经或最快增长路径。当然假设不同条件下就会有不同的大道定理。

另外萨缪尔森还证明了著名的斯托尔珀—萨缪尔森定理：简单说就是：长期来看，开展国际贸易后，出口产品生产中密集使用的生产要素（也就是本国充裕的生产要素）的报酬会提高，而进口产品生产中密集使用的生产要素（也就是本国稀缺的生产要素）的报酬会下降，而且无论这些生产要素在哪个行业中使用都是如此。

萨缪尔森另外一个定理是：赫克歇尔—俄林—萨缪尔森定理：只要存在产品价格的差异，两国就会继续开展贸易，但最终的结果将是两国两种产品的价格完全相等，而生产要素的价格也完全相等，此时如果其他条件不变，贸易也就停止。也即两国间开展贸易的结果会使两国的生产要素价格最终相等。这个定理也被称为要素价格均等化定理。

当然萨缪尔森还有其他成就，例如福利经济学里面的社会福利函数。

这里顺便瞎扯一下福利经济学。西方经济学大概分为三类，一类是纯粹的理论经济学，大概就是用经济学名词的数学，例如现在基于微分几何，代数拓扑的微观经济学，宏观经济学和计量经济学，不是数学系毕业的，基本看不懂。第二类是应用经济学，主要包括三大领域，例如研究一个国家如何获得经济高速增长的发展经济学，发展的成果如何合理分配以保持持续增长动力的福利经济学，以及保证经济增长和分配公平的制度经济学，这是西方现在应用经济学三大支柱，不过目前也基本是数学；第三类就是一切沾点经济学边的所谓技术经济学，不过大多数理论经济学家不认为他们是经济学，例如产业经济学，区域经济学，环境经济学，金融经济学，信息经济学，劳动经济学，法律经济学，管理经济学，公共经济学，国际经济学，社会经济学，货币经济学，行为经济学，国防经济学等等（目前在中国挂某某经济学的学科有 400 多个，不过大多数是名不副实，挂羊头卖狗肉，中国社科院有的自封所谓经济学家有的甚至就是民科，有的所谓经济学大家，其实连弹性（供需曲线的一次导数）和边际（供需曲线的二次导数）这种基础概念都弄不清楚。80 年代末期斯坦福，哈佛等等大学的一些经济学教授受 286 邀请来中国讲课，在中国转了一大圈，最后结论就是：只有中科院系统所里才有真正的符合国际定义的经济学家，其他地方都是自己认为的）。

福利经济学研究的主要内容有：社会经济运行的目标，或称检验社会经济行为好坏的标准；实现社会经济运行目标所需的生产、交换、分配的一般最适度的条件及其政策建议等。

福利经济学从提出至今，大致走过了三个阶段。首先是庇古的福利经济学，有两个基本命题：第一是国民生产纯产值越大，社会经济福利就越大；第二是国民收入分配越是平等，社会经济福利就越大。这个理论就是说：国家经济实力

上升必然导致公民福利增加。显然中国现在的事实证明这个理论纯属瞎扯，这个理论早就破产了，甚至被成为脑残理论。

第二个阶段是以帕累托最优状态命名的福利经济学，强调整体经济效率增加对福利的好处：在经济运行时，如对现状进行的改变，使得至少有一个人的福利增进了，这种改变就有利；如果使得至少一个人的福利减少了，这种改变就不利；但是，如果使得一个人的福利增进的同时，而使得另一个人的福利减少，就不能说这种改变一定有利或者不利。然而就是在此时，却达到了帕累托最优状态。可以证明：在完全竞争的条件下，如果消费者追求最大效用满足，生产者追求最多利润获得，生产要素所有者追求最大收入，加上没有经济外部效应，就一定能够达到社会福利最大的帕累托最优状态（按照这个理论，容易证明如果一个人独霸全世界的山河土地，仍会存在着相应的一个帕累托最优状态）。这个理论是为资本家剪羊毛提供心理安慰：我剥削有理，我不剥削你，你连饭都没得吃，我吃肉，你喝汤。这个理论被称为鸡脚杆上刮油有理理论，在明智开发的时代，显然也无法生存了。

第三阶段是以伯格森、萨缪尔森为代表的社会福利函数福利经济学，他们认为：帕累托的最优状态只解决了经济效率问题，没有解决合理分配问题，不同的收入分配会对消费和生产发生不同的影响，而经济效率仅是社会福利最大的必要条件，而合理分配产品收入是社会福利最大的充分条件，只有同时解决效率和公平的问题，才能达到社会福利的唯一最优状态。他们以政治投票与货币投票具有相似特征出发，提出用政治投票的方式构建社会福利函数。

但是阿罗摧毁了这些人的公平梦想：无法以投票的方式产生人人都能接受的唯一社会福利函数。

阿罗不可能定理最本质的理解是：个人私自利益与社会整体利益无论如何必然存在矛盾，不能在满足所有个人私有利益的前提下，逻辑地导出社会整体利益同时也被满足的结论。

阿罗不可能定理其实也摧毁了整个西方经济学关于市场竞争好处的理论基础：西方认为市场经济有好处是假设市场经济必然导致资源配置最优化，因为他们假设个人利益被满足的同时，会自动地导向全社会资源配置的总体优化（福利经济学第一定理：不管初始资源配置怎样，分散化的竞争市场可以通过个人

自利的交易行为达到瓦尔拉斯均衡，而这个均衡一定是帕雷托有效的配置；福利经济学第二定理：每一种具有帕累托效率的资源配置都可以通过市场机制实现。也即可以使效率问题与分配问题相分离。有时收入分配的结果并不尽如人意，但这并不能否认市场的作用）。

所以，下面要进一步介绍的实变函数就是以证明存在性为主，而不是计算解为主的学科。一般说来，现在非数学家是无法获得诺贝尔经济学奖的，因为现在经济学研究的问题已经超出直觉和常识太远了，非数学家，完全无法理解问题本质。现在非数学家研究经济学，基本就是民科路线：望文生义，胡搅蛮缠。目前国内一些老年的所谓经济学大家，基本都是望文生义，胡搅蛮缠的模范人物。

顺便说一句，豆瓣对数学公式不支持，尤其是积分，矩阵，复杂求和等等符号无法表达，所以这篇帖子惨不忍睹，写得实在艰难。错误在所难免，发现请指出来。

没有仔细检查，先发出来，大家一起找错误，符号太困难，错误可能有点多。