数论大观园

尽我所能地走遍数论的每一个角落

作者：逯晓零
组织：万物联盟数学分部
时间：2023/06/01
版本：1.0

我们必须知道，我们必将知道。–David Hilbert(万物联盟数学部常驻大使)

目录

第一章 导言 **1**

第二章 数的学科乃数学也 **3**
2.1 从自然数到有理数 3
2.2 实数 4
2.3 复数 5
2.4 p 进数 6
2.5 数论在哪里 7

第三章 初等数论 **8**
3.1 同余 8
3.2 积性函数 9
3.3 杂七杂八 11

第四章 代数数论 **12**
4.1 基本定理 12
4.2 费马大定理 18
4.3 岩泽理论 19

第五章 解析数论 **25**
5.1 类数公式 25
5.2 素数分布 30
5.3 零散内容 40

第六章 实数与逼近 **47**
6.1 超越数 47
6.2 逼近原理 54
6.3 非初等函数 59

第七章 方程与曲线 **68**
7.1 不定方程 68
7.2 代数曲线 78
7.3 零散内容 90

第八章 计算与验证 **91**
8.1 计算 91
8.2 验证 94
8.3 散列函数 95

参考文献 **97**

第一章 导言

十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。十分无关紧要的一章，请任意地跳过。

大家好，我是**逯晓零**，一个隶属于万物联盟数学分部的底层人员，唯一的职责就是给各位顶级的大佬打黑工，不过我也乐在其中，因为我实在太喜欢数学了，有关我们组织的信息以后再继续披露，但可以介绍一下我们数学部的部长，就是下面的这位。

我写的前两篇文章朗兰兹纲领简介和代数分析几何简介都是在她的指导下完成的。部长最强大的地方并不是数学能力，而是出众的学习理解能力，据说只有她不想学的，没有她学不会的，其实她的主业是历史分部的部长，但她也十分关心我们数学分部，比如这次她指派我来完成数论现状的总结工作，而这也是这篇文章的由来

了。

这是一次十分重要的工作，我可不能有丝毫懈怠，所以回到正题，有关部长的事可以入部了解。这次的文章也会遵循我一贯的风格，在严谨给出定义定理前后关系的基础上，忽略定理的证明，你可能想知道为什么我会去忽视证明呢？以我的个人观点来看，如果补上证明的话，我们的文章就不能称为简介了，而应该被称为教材了，现代数学的教材本身就是比较无聊单调的，无非就是“定义-定理-证明”的循环往复，这是数学本身的特点，稍微好些的教程会给例子，优秀的教程还能配合图形讲得绘声绘色，但是以我的个人能力可做不到这么多。稍微想想就能明白，我在数学分部快 7 年了，却还是底层部员，这就足以说明一切了。也不知道以前我有没有讲过，就是该如何看待和学习数学，我就把部长的教诲来稍微传达一下吧。

- 首先，严谨地读透定义和定理，并试着以自己的语言表达出来。
- 其次，你认为定理理所当然且不想浪费时间时，可以跳过证明的阅读。
- 再者，你认为定理不可思议时，无论如何都要去阅读证明。
- 最后，无聊的时候，看看曾经被跳过的证明。

部长曾说，“通义明理，数之始也”。也就是说，把定义读通透，把定理搞明白，数学的学习才能开始，而我的主要工作就是把这部分给完成，有人可能会觉得这太过于形式主义了，对的，这就是形式主义，我们部长所推崇的是 Bourbaki 学派的作风，将严谨作为数学的第一要务，我可以稍微透露一下部长目前在数学方面的课题。主要是一个与计算机部共同研究的课题，在Coq和Lean之类工具的基础上研究能被广泛应用的数学证明辅助工具，并尽可能地将已有的数学内容全部形式化为相应的证明代码。这种事情前途未卜，我也只能默默在后面为她加油了，部长曾说我的工作对她的研究十分重要，所以我也不能辜负她的期待，该开始努力工作，进入正文了。

注

万物联盟 [–认识与认识外的组织–] 是求知大陆最大最权威的学术组织，主管核心学术期刊的索引目录，并持有着求知大陆的最高学术期刊《点》(Dot)[月刊] 和最大的实验室“昙”。组织本身为非盈利性组织，由各国基金会赞助运行，每年举行一次“觅之会”，除了学术报告、学术交流等基本内容以外，备受瞩目的则是由此评选出的“描思涂奖”，其主要面向以下几个领域：数学、物理、计算机、化工、生命、文学。万物联盟以自然求知为主，社会求知为辅，其八大核心主部为：数学部、自然部、计算机部、工业部、生命部、历史部、哲学部、语言部。在人员选择上的基本理念为“宁缺勿滥”，数学部的代理部长为历史部的部长：梅缇，其它各分部的部长为，自然部：希迩睿、计算机部：橘梅忒、工业部：橘艾泊、生命部：忒珀洛菊、哲学部：空娜莱丝、语言部：梵珂萱。组织的基本标语为“最难理解的东西是最贴近理解者的存在”，除主部以外，还有各种形式多样的依附于某一主部的附属部，组织本身不具有实体而由各主部的根据地自发地联合而成。

《**点**》是万物联盟的核心主刊，每月发行一次，分八个板块，分别来自组织的八大主部，自然计算机工业生命部各 4 篇，数学部 2 篇，历史哲学语言部各 1 篇。另外，在每年期刊的尾部，会附上当前年份内的核心学术期刊的索引目录，也被称为认证目录。

“昙”是万物联盟的实验室集群总称，它分别由自然部的“场”、计算机部的“脑”、工业部的“流”、生命部的“胞”、历史部的“仓”共同构成，另外算力最强的计算机由数学部接管，以进行数学验证。在所有的实验室中，“仓”最为特殊，其主要作用不是实验，而是留存，用于保存由其它实验室产出目前不需要却有收藏价值的东西，有时会拿来做觅之会的礼品。

觅之会是万物联盟一年一度的大集会，由各主部轮流主持，通常举行于每年的夏季，持续 8 天。除《点》以外的在核心学术期刊索引目录中的作者将有机会在此次会议中做学术报告，此会议不设任何分会，因此在觅之会上你将可以体会来自各种领域的学术成果，参加觅之会对于扩展眼界具有极大的帮助。

描思涂奖是求知大陆的最高学术奖项，属于荣誉奖项，不会带来实际的资金奖励，而是一块奖牌配上主部的限量精良工业品。基本每年的工业礼品都包含了当前最前沿的技术，而且复刻难度极大，基本只能在万物联盟的核心实验室中限量制作。[1]

[1]我在这里偷偷地告诉你，这章的目的只是为了凑够 100 页这个好看的数字。

第二章 数的学科乃数学也

顾名思义，所谓**数学**就是研究“数”的科学，虽然这并不精确，但“数”在数学中具有无可撼动的地位，所以在数论的开始，我们有必要介绍一下“数”都有哪些？而哪些又是我们所研究的“数”。

本章的前置知识为：抽象代数、数学分析、线性代数。

2.1 从自然数到有理数

克罗内克曾经说过“上帝创造了自然数，其余的数都是人造的。”其告诉了我们一个简单的事实，即只需要完成自然数 $\mathbb{N}$ 的公理，其余的数都可以定义出来。完全严格的自然数导出有两种方式，一是哥德尔在证明不完备定理时，其条件所给的数论公理，二是从公理集合论的角度，通过空集嵌套来导出可数的有理数集，而它们的核心思想其实都来自于众所周知的 Peano 公理。

公理 2.1 (Peano 公理)

1. 0 是自然数。
2. 每个自然数 n 的后继 n^+ 是自然数。
3. 0 不是任何自然数的后继。
4. 不同自然数的后继不同，即 $n^+ = m^+ \Rightarrow n = m$。
5. 设 $P(n)$ 是一个关于自然数的命题。如果 $P(0)$ 为真且 $P(n)$ 为真可以推出 $P(n^+)$ 为真，则对任意的自然数 n 有 $P(n)$ 为真。

♡

有关它的讨论数不胜数，比如此时可以定义出具体的阿拉伯数字 $1 = 0^+, 2 = 1^+, \ldots$，可以定义出加法 $n + 0 = n, n + m^+ = (n + m)^+$，可以定义出乘法 $n0 = 0, nm^+ = nm + n$，我强推由陶哲轩写的这本书 [44]，其有不少观点都值得我们好好说道一番。

比如自然数符号的无关紧要性，通常使用十进制的阿拉伯数字，即 $\{0, 1, 2, 3, \ldots\}$，还能使用罗马数字 {O,I,II,III,IV,...}，还能使用二进制 $\{0, 1, 10, 11, 100, \ldots\}$ 等等，但无论如何 $1 + 1 = 2$ 是永远成立的，其关键在于你如何看待这个式子，也就是我们之前提及的学习数学的第一步，搞清定义 (1 是什么？2 是什么？+ 是什么？) 和定理 (1+1=2) 的含义，如果这两点都不明晰的话，证明就无从谈起了，1+1=2 的标准含义是“0 的后继加上 0 的后继等于 0 的后继的后继”，而在常规的进位制下，就是上面所写的形式。虽然舆论和大众都喜欢议论一些模糊不清的东西，但作为专业的数学学习者就不要去瞎掺和了，一笑而过就是最好的态度。

有了上述基础，要证明交换律、结合律、消去律、分配律等都是水到渠成的事，而且相关的书籍也非常多，我想告诉你的是另一种与 Peano 公理等价的表述，在 Peano 公理中，我们给出了 0、自然数和后继三个性质概念，然后引入了加法乘法并完成了性质的证明。在实际应用中，我们其实更关注的是自然数的运算，后继之流并非什么重要概念，所以如果从运算性质来完成自然数的定义，自然会更好一些，也就是我们下面的公理。

公理 2.2 (自然数公理)

自然数 $\mathbb{N} = \{0, 1, 2, 3\ldots\}$ 是具有下述性质的对象构成的整体

1. 有加法和乘法，并且满足结合律 $((a+b)+c = a+(b+c), (ab)c = a(bc))$、交换律 $(a+b = b+a, ab = ba)$、分配律 $(a(b + c) = ab + ac)$。
2. 存在 1 是乘法的单位元，即满足 $1n = n$。
3. 乘法满足消去律且存在 0 不满足，即 $a \neq 0, ma = na \Rightarrow m = n$。
4. 对任意的 m, n 有且只有下面的三种情况：$m = n$、$m + x = n$ 有唯一 x、$n + y = m$ 有唯一 y。
5. 设 $P(n)$ 是一个关于自然数的命题。如果 $P(0)$ 为真且 $P(n)$ 为真可以推出 $P(n^+)$ 为真，则对任意的自然数 n 有 $P(n)$ 为真。

♡

由 Peano 公理推自然数公理没啥好说的，主要看怎么反过来，由第二和第三条给出了自然数中的两个特殊存在，0 和 1，通过第四条可以说明它们在自然数中唯一存在，因此可以定义 Peano 公理中的后继为 $n^+ = n+1$，剩下的就简单了，也没啥好说的。此时通过这个定义我们可以很容易地发现，自然数 $\mathbb{N}$ 构成加法交换幺半群，$\mathbb{N}-\{0\}$ 构成乘法交换幺半群，那么我们自然就引出了扩充自然数的想法，也就是我们接下来的内容。

我们知道一个简单的事实，交换幺半群可以唯一完备化为交换群，这其实自由群构造中用到的事实，放到这里我们应该考虑比较完整的加法交换幺半群 $\mathbb{N}$，而不是乘法，其完备化的结果就是整数 $\mathbb{Z}$ 中的加法交换群成分。当然构造的过程并不是简单暴力地让 n 对应一个 $-n$，而是下面这样。

定义 2.1 (整数)

$\mathbb{Z} = \mathbb{N}\times\mathbb{N}/\sim, (a,b)\sim(c,d) \Leftrightarrow a+d=b+c$ ♣

如果我们把上面的有序对 (a,b) 写成 $a-b$ 的话，你可能就能清晰地理解上面地概念了，比如各种负数 $-n$ 都可以视为 $(0,n)$ 或者 $0-n$。为什么非要这么定义呢？又是笛卡儿积，又是等价关系，这时可以回忆一下我在上一篇文章 [45] 的范畴论中讲的一件事，公理可以凭空造物，定义和构造则必需从已有进行选择，换句话说，我们不能仅凭 n 就凭空造出 $-n$，除非我们不是构造整数，而是直接在整数上构造公理，而这与我们最开始的理念是相悖的，因此上面的这种定义才是普遍采用的方式。此时将乘法兼容过来，$\mathbb{Z}$ 就已经是一个性质较为丰富的交换整环了。

我们还知道一个简单的事实，交换整环可以唯一完备化为域，而做法与自然数扩充为整数是如出一辙的，区别在于一个用的是加法，一个用的是乘法，具体写出来也是很简单的。

定义 2.2 (有理数)

$\mathbb{Q} = \{(a,b)\in\mathbb{Z}\times\mathbb{Z} : b\neq 0\}/\sim, (a,b)\sim(c,d) \Leftrightarrow ad=bc$ ♣

实际将数对 (a,b) 写出来就是 $\frac{a}{b}$，有人可能会好奇像 $a-b$ 和 $\frac{a}{b}$ 算不算凭空造物呢？在有前后文的基础上不算。因为它们本质上一个数对，我们只是换了一种适合我们书写习惯的形式罢了，就和写“罗马数字 II、二进制的 10、十进制的 2”都是一件事是同样的道理。

到有理数的构造是比较简单的，有时我们甚至可以直接拿“特征为零的素域”的定义式的内容来作为我们的公理，一步到位地来到线性计算的顶点。我们注意到，到目前为止的两步扩充都有抽象代数提供的唯一性做担保，这意味着在 Peano 公理下，我们只能得到唯一有理数理论，而有理数代表了我们加法和乘法玩到极致的结果，也就是说完全清晰计算已经到头来，而分析多样的近似计算才刚刚开始。

$$\text{扩张链：}\mathbb{N}\overset{\text{交换幺半群的唯一扩充}}{\Rightarrow}\mathbb{Z}\overset{\text{交换整环的唯一扩充}}{\Rightarrow}\mathbb{Q}$$

2.2 实数

实数的构造就比较玄乎了，但我们还是要试着去阐明它。理解实数的关键是连续，如何在 $\mathbb{Q}$ 上定义序，我们就不重复了，一个显而易见的事实是，任意两个有理数 $a<b$ 之间一定存在一个有理数 $a<\frac{a+b}{2}<b$，另一个事实是，有理数域具有封闭性，只靠所定义的常规运算，我们无论如何也是无法达到实数的领域内的。所以我们必需以其它的方法计算出实数，其中的一种就是极限逼近了，这个思想来自我们常规我们对实数计算的习惯，例如 $\sqrt{2}$，我们的基本做法就是利用方程 $x^2=2$，然后不断一位一位地试出结果来，就像下面的过程。

$$1.4, 1.41, 1.414, 1.4142, 1.41421, \ldots$$

认识实数时，我们需要时刻理解，它就是一个收敛序列，当然对于同一个实数，可能存在不同的收敛序列，而其中的有理数序列是我们最为关注的。

定义 2.3 (有理数序列)

(1) 我们将映射 $f:\mathbb{N}-\{0\}\to\mathbb{Q}, n\mapsto r_n$ 称为有理数序列，并简记为 $\{r_n\}$。

(2) 设 $\{r_n\}$ 是有理数序列。如果存在有理数 r 使得对任意的 $\varepsilon>0$ 都存在 $N=N(\varepsilon)$ 满足当 $n>N$ 时都有 $|r_n-r|<\varepsilon$，则称 r 是 $\{r_n\}$ 的极限，记为 $\lim\limits_{n\to\infty} r_n=r$。

(3) 设 $\{r_n\}$ 是有理数序列。如果对任意的 $\varepsilon>0$ 都存在 $N=N(\varepsilon)$ 使得当 $n,m>N$ 时都有 $|r_n-r_m|<\varepsilon$，则称 $\{r_n\}$ 是有理数基本序列。

♣

上面的第二个定义给出了有理数极限的概念，而第三个定义给出了收敛的概念，将两者结合起来就可以得到实数的定义了。我们令 $\mathscr{M}$ 表示所有有理数基本序列构成的集合。

定义 2.4 (实数)

$$\mathbb{R}=\mathscr{M}/\sim,\{a_n\}\sim\{b_n\}\Leftrightarrow\lim_{n\to\infty}(a_n-b_n)=0$$

♣

简单来讲，实数就是满足柯西收敛原理的序列构成的等价类，收敛说明了实数可以被逼近的性质，而等价类说明了能逼近实数的等价类是不唯一的，而这正是实数的两个核心实质，读者可以多去品味和理解。通过 $\varepsilon-\delta$ 语言来推导实数，最大的优点是得到实数定量的运算性质，如果只需了解实数定性的逻辑性质的话，我们可以使用另一种实数的构造法。

定义 2.5 (戴德金分割)

任意子集 $X\subset\mathbb{Q}$，如果满足 $(1)\varnothing\neq X\neq\mathbb{Q}(2)r\in X,r'\in\mathbb{Q},r'<r\Rightarrow r'\in X(3)r\in X\Rightarrow\exists r'\in X,r<r'$，则称 X 是一个实数。

♣

简单来讲就是比这个实数小的所有有理数构成的整体，这里给出仅作了解，因为实数的核心作用是分析学，只有能运算的实数才是有用的，最后我们给出实数完备的几个定理来结束本节，以下所有的序列均指实数序列。

定理 2.1 (实数完备性)

(1) 序列 $\{a_n\}$ 收敛当且仅当，对任意的 $\varepsilon>0$ 都存在 $N=N(\varepsilon)$ 使得当 $n,m>N$ 时都有 $|a_n-a_m|<\varepsilon$。

(2) 单调有界序列必有极限。

(3) 对任意一个区间套 $\{[a_n,b_n]\}$(即 $[a_{n-1},b_{n-1}]\subset[a_n,b_n]$, $\lim\limits_{n\to\infty}|b_n-a_n|=0$) 均有 $\cap_{n=1}^{\infty}[a_n,b_n]$ 是单点集。

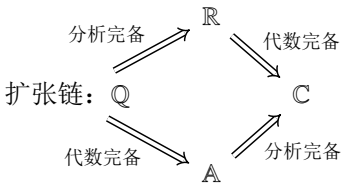
(4) 非空有上 (下) 界数集必有上 (下) 确界。

(5) 无限点集上至少有一个聚点。

(6) 闭区间上的任意一个开覆盖必存在一个有限子覆盖。

♡

2.3 复数



上面的 $\mathbb{A}$ 代表代数数，其在复数 $\mathbb{C}$ 中的补集称为超越数，e,π 都是典型的超越数，当然我们更关注复数，复数似乎是一个很好理解的概念，无非就是形如 $z=a+bi,a,b\in\mathbb{R}$ 的数，它的核心作用是得到代数基本定理。

定理 2.2 (代数基本定理)

$\mathbb{C}$ 是代数闭域，或者任何复系数一元 n 次多项式方程在复数域上至少有一根，或者任意 $f(x) \in \mathbb{C}[x]$ 的分裂域为 $\mathbb{C}$。

♡

其实我是知道的，大部分人对复数的代数性质并不感兴趣，因为如果能充分理解实数，我们也只会感叹“就这”，问题在于我们并没有充分理解实数，因此复数的分析性质才是我们对复数关注的核心，从复分析来谈论实分析，就相当于从高维审视低维，使得很多东西变得简单了起来，这些只有通过大量实例才能体会，由于我们的主题是数论，所以我们就一笔带过了。

定理 2.3 (欧拉公式)

$$e^{i\theta} = \cos(\theta) + i\sin(\theta)$$

♡

2.4 p 进数

数域中最奇妙的莫过于 p 进数了，我认为的比较好用的教材是 GTM198[27]。既然要讨论 p 进数，那么我们的第一步自然是稍微回忆一下 p 进数该如何定义和引入，主要有三种方法，其分别代表了我们看待 p 进数的几种基本观点。第一种是和上面实数的引入方式类似，区别在于改变绝对值的定义，只有几步就能搞定。

定义 2.6 (p 进数的序列定义)

(1) 对于有理数 $r \in \mathbb{Q}$，我们定义 $r = p^m \frac{u}{v}, (uv, p) = 1$ 中的 $m = v_p(r)$ 为 r 的 p 进赋值，并定义 p 进绝对值为 $|r|_p = p^{-v_p(r)}, |0|_p = 0$。

(2) 设 $\{r_n\}$ 是有理数序列。如果存在有理数 r 使得对任意的 $\varepsilon > 0$ 都存在 $N = N(\varepsilon)$ 满足当 $n > N$ 时都有 $|r_n - r|_p < \varepsilon$，则称 r 是 $\{r_n\}$ 的 p 进极限，记为 $\lim\limits_{n\to\infty} r_n = r$。

(3) 设 $\{r_n\}$ 是有理数序列。如果对任意的 $\varepsilon > 0$ 都存在 $N = N(\varepsilon)$ 使得当 $n, m > N$ 时都有 $|r_n - r_m|_p < \varepsilon$，则称 $\{r_n\}$ 是有理数 p 进基本序列。

(4)$\mathbb{Q}_p = \mathscr{M}_p/\sim, \{a_n\} \sim \{b_n\} \Leftrightarrow \lim\limits_{n\to\infty}(a_n - b_n) = 0, \mathscr{M}_p$表示所有有理数 p 进基本序列构成的集合

♣

第二种是使用进位制级数方式，它其实就相当于小数的意思，所有的 10 进制实数都可以表示为如下的形式

$$r = \sum_{n=-\infty}^{m} c_n 10^n \in \mathbb{R}, c_n = 0, 1, ..., 9$$

当然 10 可以替换为任意一个正整数 b 以得到 b 进制，而我们只需要把方向转过来就是 p 进数了

$$\mathbb{Q}_p = \{\sum_{n=m}^{\infty} c_n p^n : m \in \mathbb{Z}, c_n \in \{0, 1, ..., p-1\}\}$$

对这种情况，比较疑惑的是 p 是否可以像实数一样选取非素数？其实我也不是很清楚，而且在这种表示下，p 进数之间似乎没有区别，比如 2 进数 $\mathbb{Q}_2$ 和 3 进数 $\mathbb{Q}_3$ 似乎只是和实数一样换了个进位制。笔者认为，这主要是因为我们把传统的算术观给带入了进来，p 进数的算术其实大有不同，witt 向量是我认为目前最好的理解方式，这篇文章 [26] 有一个系统的总结，从中也可以解答我们的疑惑，素数的要求主要还是来自素域 $\mathbb{F}_p$ 的要求，单纯使用 $\mathbb{Z}_n$ 环是达不到要求的。第三种则是逆向极限的方式，就两步，先得到 p 进整数环

$$\mathbb{Z}_p = \varprojlim_n \mathbb{Z}/p^n\mathbb{Z}$$

然后再由整环扩张为其分式域 $\mathbb{Q}_p = \mathrm{Frac}(\mathbb{Z}_p)$ 即可。有人可能不是很理解逆向极限，它其实相当于无限笛卡儿积的子集

$$\varprojlim_n \mathbb{Z}/p^n\mathbb{Z} \subset \prod_{n=1}^{\infty} \mathbb{Z}/p^n\mathbb{Z} = \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p^2\mathbb{Z} \times \mathbb{Z}/p^3\mathbb{Z}...$$

但逆向极限还多了一个给定的映射链，在我们的 p 进整数中就是下面的情况

$$...\mathbb{Z}/p^{n+1}\mathbb{Z} \xrightarrow{f_n} \mathbb{Z}/p^n\mathbb{Z} \xrightarrow{f_{n-1}} ...\mathbb{Z}/p\mathbb{Z} \xrightarrow{f_1} \mathbb{Z}/p\mathbb{Z}$$

其中的映射是环的自然投影 $f_n : \mathbb{Z}/p^{n+1}\mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}, a \mapsto b, a \equiv b(\mathrm{mod} p^n)$，想必看到这里的都熟悉抽象代数，这种集合配映射的结构应该再熟悉不过了。有关 p 进数，我们还需要了解它的扩域结构，其并不像实数那么简单，从赋值就能窥见一二了，$v_p(\mathbb{Z}_p) = \mathbb{Z}_{\geq 0}, v_p(\mathbb{Q}_p) = \mathbb{Z}$，有限代数扩张可以让其赋值稍微接近有理数，但还是不能完全填满有理数，但只需通过代数闭包就能达到了 $v_p(\overline{\mathbb{Q}}_p) = \mathbb{Q}$，但此时我们又会失去赋值的完备化，故进一步就得到了 $\mathbb{C}_p$，虽然我们在代数和分析上都达到了完备，但其不满足一种叫**球面完备** (Spherically Complete) 的性质，它就是实分析中闭球套定理，重要性就不言而喻了，为了到达这种完备就有了**泛 p 进域** (universal p-adic field)Ω_p，它可以使得赋值到达实数域 $v_p(\Omega_p) = \mathbb{R}$。

$$\text{扩张链：} \mathbb{Q} \overset{\text{p 进分析完备}}{\Rightarrow} \mathbb{Q}_p \overset{\text{代数完备}}{\Rightarrow} \overline{\mathbb{Q}}_p \overset{\text{p 进分析完备}}{\Rightarrow} \mathbb{C}_p \overset{\text{球面完备}}{\Rightarrow} \Omega_p$$

当然，我们在这里也只是稍微科普一下有哪些数，我们真正研究数论时，到 p 进数域的有限扩张就已经足够了，而且如果有些东西连给数学带来大影响都做不到的话，基本就只能沦为课后习题，然后被遗忘到角落里。

2.5 数论在哪里

至此我们已经了解了数学中的各种数，那么我们的数论又在研究哪些数呢？我们应该毫不犹豫地回答，自然数 $\mathbb{N}$，为啥那？当然是“上帝只创造了自然数”。开玩笑啦，数论的起点是自然数，但它的研究方法涉及了各种各样的数，整环 $\mathbb{Z}$、有理数域 $\mathbb{Q}$ 的有限扩张、实分析 $\mathbb{C}$、复变函数 $\mathbb{C}$、p 进分析 $\mathbb{Q}_p$，基本每一个领域都或多或少的为数论提供了大量帮助，而最为纯粹的初等数论的发展巅峰其实是高中的数学竞赛。这怪不得数论，而是有一座大山压在数论身上，它就是哥德尔不完全性定理。

定理 2.4

(1) 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为真，也不能被证明为否。

(2) 如果系统 S 含有初等数论，当 S 无矛盾时，它的无矛盾性不可能在 S 内证明。

有关这玩意的讨论太多了，比如这本书 [37]，实际上基本任何一本数理逻辑的书都会介绍，我们也不想说太多。我们只需了解到这样一个事实，基本大部分数论的结论都没有初等数论的解答，它不一定是命题，但确实有着大量的资料支撑着这个结论。

第三章 初等数论

初等数论的基本内容并不多，而且很多东西是符合我们基本认知的，我在 [46] 的数论形式化中已经讲过，只需要将欧几里得环的相关内容套到 $\mathbb{Z}$ 上，就能得到大部分初等数论中的定义和性质了，包括**整除**、**素性**、**唯一分解性**、**带余除法**、**公因子**、**公倍数**等。但这并不代表初等数论就点到为止了，还有很多值得探究的东西尚未涉及。

3.1 同余

同余就是一个可以说道的东西之一，其基本定义十分简单。

$$a \equiv b (\mathrm{mod}\ n), n \mid (a-b)$$

它的大部分基本性质都比较简单，有时对于 $n = p$ 为素数的情况，甚至可以用有限素域 $\mathbb{F}_p$ 的观点进行看待。稍有难点的是同余方程，讨论最多的是一元多项式方程 $f(x) \equiv 0 (\mathrm{mod}\ n), f(x) \in \mathbb{Z}[x]$，或者方程组，我们可以以次数为基准一点点来讨论。首先是一次同余方程

$$ax + b \equiv 0 (\mathrm{mod}\ n)$$

只能说毫无难度，所以对于一次方程，我们更喜欢讨论方程组，而剩余定理给出了我们想要的结果。

定理 3.1 (中国剩余定理)

设 $m_1, ..., m_k$ 是两两互素的正整数，记 $M = m_1 ... m_k, M_i = M/m_i$，则下面的同余方程组

$$x \equiv a_1 (\mathrm{mod}\ m_1), x \equiv a_2 (\mathrm{mod}\ m_2), ..., x \equiv a_k (\mathrm{mod}\ m_k)$$

存在唯一的解为

$$x \equiv a_1 M_1' M_1 + a_2 M_2' M_2 + ... + a_k M_k' M_k (\mathrm{mod}\ M)$$

其中的 M_i' 是满足 $M_i M_i' \equiv 1 (\mathrm{mod}\ m_i)$ 的正整数解。

♡

其基本想法是对于有解的一次同余方程，其解的形式一定为 $x \equiv c (\mathrm{mod}\ n)$，这样连立的一次方程组就变成了方程组解的联立，而这就是定理所要求的条件，而且模非素数的方程由可以分解为模互素的方程组，所以说，剩余定理可谓是解同余方程的基础了。当我们开始考虑二次同余方程时，就来到了二次剩余这个出美妙定理的地方，考虑 $(a, m) = 1$，如果同余方程 $x^2 \equiv a (\mathrm{mod}\ m)$ 有解，则称 a 为 m 的二次剩余，而无解时称 a 为 m 的非二次剩余。当 $m = p$ 为奇素数且 $a \mid p$ 时，上述方程只会出现无解或两个不同解的情况，于是就可以定义著名的勒让德符号。

$$(\frac{a}{p}) = \begin{cases} 0 & p \mid a \\ 1 & \text{a 为 p 的二次剩余} \\ -1 & \text{a 为 p 的非二次剩余} \end{cases}$$

它的主要作用就是通过计算的手段来判断同余方程是否有解，而计算的公式中就有我们最爱的二次互反律了。

定理 3.2

(1) 分解公式 $(\frac{ab}{p}) = (\frac{a}{p})(\frac{b}{p})$

(2) 欧拉判别法 $(\frac{a}{p}) \equiv a^{\frac{p-1}{2}} (\mathrm{mod}\ p)$

(3) 平方消灭法 $(\frac{a^2}{p}) = 1$

(4) 二次互反律 $(\frac{p}{q})(\frac{q}{p}) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}$

(5) 特殊值计算 $(\frac{-1}{p}) = (-1)^{\frac{p-1}{2}}, (\frac{2}{p}) = (-1)^{\frac{p^2-1}{8}}$

♡

实际上还有将勒让德符号下面的素数替换为一般整数的雅可比符号，但笔者觉得意义不大就是了。在二次剩余上再前进一步就是原根的概念了，设 $(a,n) = 1$，我们把满足 $a^r \equiv 1(\mathrm{mod}\ n)$ 的最小整数 r 称为 **a 模 m 的阶**，并记 $r = \mathrm{ord}_n a$，这实际上就是有限域中乘法群结构的阶的概念，那么我们自然就能意识到生成元的地位了，如果有 $\mathrm{ord}_n a = \varphi(n)(\varphi(n)$ 为欧拉函数)，我们就称 a 为**模 m 的原根**。原根和指数都比较无聊，主要拿来给竞赛生练练手，更多的方程实例差不多也属于习题级的存在，就不讨论了。至于同余的定理倒没什么，像威尔逊定理 p为素数 $\Leftrightarrow (p-1)! \equiv -1(\mathrm{mod}\ p)$、费马小定理 $a^p \equiv a(\mathrm{mod}\ p)$、欧拉定理 $a^{\varphi(n)} \equiv 1(\mathrm{mod}\ n)$ 之类的实际上都可以在有限域的理论中找到。

3.2 积性函数

所谓**算术函数** (或数论函数) 指映射 $f : \mathbb{N} - \{0\} \to \mathbb{C}$，我们也可以把它看为复数列，显然一般的复数列应该在复分析中进行讨论。在数论中，我们主要讨论积性函数，即满足下面性质的算术函数

$$(m,n) = 1 \Rightarrow f(mn) = f(n)f(m)$$

如果去除互素条件 $(m,n) = 1$ 后依旧有 $f(mn) = f(n)f(m)$，则称其为完全积性函数。对于积性函数 f，我们只需要知道素因数的分解 $n = p_1^{a_1} \ldots p_s^{a_s}$ 就能计算出相应的函数值

$$f(n) = f(p_1^{a_1}) \ldots f(p_s^{a_s})$$

即只需要知道素数幂的函数值就能得到所有的函数值。如果想只知道素数的函数值就能得到所有的函数值则需要完全积性函数，从上面的性质可以知道，积性函数只需加上 $f(p^n) = (f(p))^n$ 的条件就能升级为完全积性函数了。单有一个概念是不够的，我们需要认识一些常见的积性函数和相关性质，首先是见过无数次的欧拉函数。

定理 3.3 (欧拉函数)

设 $\varphi(n)$ 表示不超过 n 且与 n 互素的正整数的个数，即满足 $0 < m \le n, (m,n) = 1$ 的整数 m 的个数。

(1)$\varphi(n)$ 是积性函数

(2)$\varphi(p^a) = p^a - p^{a-1}$

(3)$\sum_{d|n} \varphi(d) = n$

♡

积性函数的讨论无非就两步，其一是否完全积性，其二相应地给出素数或素数幂的计算值。上面的最后一个性质是和函数性质，我们把 $F(n) = \sum_{d|n} f(d)$ 称为 f 的**和函数**，一个简单的性质是积性函数的和函数是积性函数，欧拉函数的和函数就是恒等函数。接着我们可以直接枚举各种积性函数，$\sigma_k(n)$ 表示 n 的所有正因子的 k 次幂之和、$\tau(n) = \sigma_0(n)$ 表示 n 的所有正因子个数、$\sigma(n) = \sigma_1(n)$ 表示 n 的所有正因子之和，它们都是积性函数，并且有下面的计算性质

$$\sigma_k(p^a)=1+p^k+p^{2k}+\ldots+p^{ak},\sigma(p^a)=\frac{p^{a+1}-1}{p-1},\tau(p^a)=a+1$$

实际上，$\tau(n)$ 可以视为不变函数 $\mathbf{1}(n)=1$ 的和函数，$\sigma(n)$ 可以视为单位 (或恒等) 函数 $\mathbf{Id}(n)=n$ 的和函数，注意到 **1** 和 **Id** 是完全积性函数，故完全积性函数的和函数不一定是完全积性函数。我们把

$$\mu(n)=\begin{cases}1 & n=1\\ (-1)^r & n\text{无平方因子且}n=p_1\ldots p_r\\ 0 & \text{其它情形}\end{cases}$$

称为**莫比乌斯函数**，它是一个积性函数。由于它的和函数十分有特点

$$\varepsilon(n)=\sum_{d|n}\mu(d)=\begin{cases}1 & n=1\\ 0 & n>1\end{cases}$$

因此我们可以得到著名的反演公式。

定理 3.4 (莫比乌斯反演公式)

对任意算术函数 $f(n)$ 和它的和函数 $F(n)=\sum_{d|n}f(d)=\sum_{d|n}f(\frac{n}{d})$ 有

$$f(n)=\sum_{d|n}\mu(d)F(\frac{n}{d})=\sum_{d|n}\mu(\frac{n}{d})F(d)$$

♡

借助反演公式我们可以容易的证明，如果和函数 $F(n)$ 是积性函数，则原函数 $f(n)$ 是积性函数。上面我们见到的都是值域为整数的积性函数，实际上还有一类值域为复数的完全积性函数十分常见，就是狄利克雷特征。

定义 3.1

设 $\chi(n)$ 是不恒为零的完全积性函数[a]，如果存在正整数 q 使得 $\chi(n+q)=\chi(n)$ 并且 $(n,q)>1\Rightarrow\chi(n)=0$，则称 $\chi(n)$ 是模 q 的狄利克雷特征，并把最小的 q 称为它的周期。

[a]恒零函数 $f(n)=0$ 也是一个完全积性函数，不恒为零加积性隐含了 $\chi(1)=1$ 的性质

♣

周期 $q=1$ 时可以得到恒等函数，所以狄利克雷特征是一类满足某些性质的算术函数，并不是唯一的，例如固定素数 p 的勒让德符号 $\chi(n)=(\frac{n}{p})$，又比如主特征 “$\chi^o(n)=1,(n,q)=1;\chi^o(n)=0,(n,q)>1$”。容易发现狄利克雷特征的下述计算性质

$$\chi(1)=1,\chi(-1)=\pm1,(n,q)=1\Rightarrow(\chi(n))^{\varphi(q)}=1$$

因此狄利克雷特征的值域只有 0 和单位根 $e^{\frac{2\pi ki}{n}}$，由于狄利克雷特征之积还是狄利克雷特征，容易发现所有模 q 的狄利克雷特征构成一个有限交换群，单位元是主特征 χ^o，χ 的逆元是它的共轭特征 $\overline{\chi}$，实际上可以写出所有的狄利克雷特征，就把它留给读者自己去探索了。剩下的积性函数都不怎么常见，也不知道有什么用处，但我们可以考虑所有算术函数构成的整体所具有的性质，如果我们对所有不恒为零的算术函数 $f_1(n),f_2(n)$ 定义狄利克雷卷积

$$f(n)=f_1(n)*f_2(n)=\sum_{d|n}f_1(d)f_2(\frac{n}{d})$$

显然它是一个算术函数，由此定义的乘法可以使得所有非零算术函数构成了一个交换群，它的单位元是莫比乌斯函数的和函数 $\varepsilon(n)$，$f(n)$ 的逆元 $f^{-1}(n)$ 采用递归的方式进行定义，即有

$$f^{-1}(n)=\begin{cases}\frac{1}{f(1)} & n=1\\ -\frac{1}{f(1)}\sum\limits_{d|n,d>1} f(d)f^{-1}(\frac{n}{d}) & n>1\end{cases}$$

如果直接使用算术函数相加作为加法，则所有算术函数构成一个整环，加法单位元为恒零函数 $\mathbf{0}(n)=0$。在这些理论的基础上，我们有 $\mu*\mathbf{1}=\varepsilon$，即莫比乌斯函数的逆元为恒等函数。最后，我顺便写个连分数测试一下 latex 的渲染能力

$$a_0+\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\cfrac{1}{a_4+\cdots}}}}$$

有关连分数其实没什么好说的，它就是数的另一种写法而已，比较有用的是两个事实，有限连分数是有理数，无限循环连分数是二次无理数并且二次无理数是循环连分数，这样我们可以通过截取根式连分数的一部分来近似计算根式的值，而且逼近效果尚可。

3.3 杂七杂八

还有一个丢番图方程 (或称不定方程) 的重要课题，由于其涉及面过于广泛，我们在后面再进一步研究。实际上，初等数论就只是一个问题的产出地，比如十分著名的孪生素数猜想、哥德巴赫猜想、3x+1 猜想、回文数猜想等都出自于初等数论，当然远不止如此，大量的不定方程就不说了，我们还能介绍几个著名的猜想。我们把满足 $\sigma(n)=2n$ 的正整数 n 称为**完全数**，一个著名的结论是

定理 3.5

偶数 n 是完全数当且仅当 $n=2^{m-1}(2^m-1)$，其中 $m\geq 2$ 是使得 2^m-1 为素数的整数。

于是马上就能萌生两个猜想，是否有奇完全数？偶完全数是否有无限个？对于后一个问题，我们可以引出另一个猜想，首先有一个结论

定理 3.6

对于正整数 m，如果 2^m-1 是素数，则 m 是素数。

于是寻找偶完全数实际就是找形如 2^m-1 的素数，我们把 $M_n=2^n-1$ 称为第 n 个**梅森数**，如果 M_p 还是素数就把它称为**梅森素数**。于是研究偶完全数就相当于研究梅森素数了，上述问题就变成了，梅森素数是否有无限个？另一个则是由正 n 边形尺规作图引出的费马素数

定理 3.7

我们把 $F_n=2^{2^n}+1$ 称为费马数。

(1)$m\neq n\Rightarrow (F_m,F_n)=1$

(2) 正 n 边形可以尺规作图当且仅当 $n=2^s p_1\ldots p_r$ 且 $p_1,\ldots,p_r$ 为两两不同的费马素数。

目前已知 $F_0=3,F_1=5,F_2=17,F_3=257,F_4=65537$ 是素数外尚未发现其它的费马素数，于是自然有猜想，费马素数是否只有这 5 个？只有这样才能完全解决正 n 边形尺规作图的问题。数论的问题还有很多，想要了解的小伙伴可以参考这本书 [14]。我想读者的建议是，不要过分去研究这些初等数论的问题，它只能做饭后甜点，而不能作为正餐，一方面初等数论的大部分问题相当于数字游戏，基本“无用又无聊”，另一方面，大部分未解决的初等数论问题都需要超越初等数论本身的工具，数学是一步一个脚印的，去研究这些工具的相关理论更有意义，有时说不定一不小心就解决了某个初等数论的猜想。

第四章 代数数论

有关代数数论，我在上上篇文章 [46] 中已经介绍得十分详细了，在里面基本可以找到所有代数数论的相关概念，包括**代数整数环**、**理想类群**、**赋值**、**分歧理论**等，甚至我们还在后面进一步介绍了类域论，所以我们直接跳过这一章也并不过分。但既然要写我们就讨论一些比较实际的问题，大部分内容抄自 [39]。

4.1 基本定理

定理 4.1 (代数数论基本定理)

(1) 理想类群的有限性：任意数域 K 的类数 $h_K = |C_K|$ 是有限的。

(2) 狄利克雷单位定理：对任意数域 K，设 $[K,\mathbb{Q}] = n = r_1 + 2r_2$，$r_1$ 为实素点个数，r_2 为复素点个数，记 $r = r_1 + r_2 - 1$，则有 $O_K^\times = \mathbb{Z}^{\oplus r} \oplus W_K$。

Minkowski 定理

在这一节中我们主要来讲述这两个定理的证明思路，基本所有书籍采用的都是传统的证明方法，也就是需要依赖几何数论中的 Minkowski 定理。我们先引入一些比较简单的概念，对于子集 $H \subset \mathbb{R}^n$，如果存在 $\mathbb{R}^n$ 的一组基 $\{a_1,...,a_n\}$，使得在加法群结构中有分解 $H = \mathbb{Z}a_1 \oplus ... \oplus \mathbb{Z}a_n$，我们就把 H 称为 $\mathbb{R}^n$ 的一个**格**。这里的 H 实际就是我们常规所见到的格点，只不过不一定是方格，只要是平行格都是可以的。在 $\mathbb{R}^n$ 上的常规勒贝格测度表示为 μ，它就是我们经常所说的体积的扩展，我们所认识的体积就是勒贝格测度，但勒贝格测度却比我们所认识的体积更广。此时我们把格 H 的一个单元的体积记为 $V(H)$，就是下面的表达式

$$V(H) = \mu(P(a_1,...,a_n)) = \mu(\mathbb{R}^n/H) = \mu(\{\sum_{i=1}^n c_i a_i : 0 \le c_i < 1\})$$

上面构造商集的方法，我们在通过复数域构造复环面中见过一次，等价的定义就是差值是否落在格 H 上。如果选取标准正交基 $\{e_i\}$，并假设坐标变换为 $a_i = \sum_{j=1}^n r_{ij} e_j$，则我们可以用平行体的行列式体积公式得到 $V(H) = |\det(r_{ij})|$，这些内容基本可以全凭几何直观，所以我们不做过多赘述。$S \subset \mathbb{R}^n$ 为**凸集**指 $\forall x, y \in S, \frac{x+y}{2} \in S$，$S \subset \mathbb{R}^n$ 为**关于原点对称**指 $\forall x \in S, -x \in S$，于是我们有下面的定理。

定理 4.2

设 H 是 $\mathbb{R}^n$ 中的一个格，S 是 $\mathbb{R}^n$ 中的勒贝格可测集。

(1) 如果 S 是关于原点对称的凸集且 $\mu(S) > 2^n V(H)$，则 $S \cap H$ 存在非零元。

(2) 如果 S 是关于原点对称的紧凸集且 $\mu(S) \ge 2^n V(H)$，则 $S \cap H$ 存在非零元。

注 这个定理虽然带了个名字，但证明起来其实并不困难，由于凸性的存在，这里的勒贝格可测集一定是“实”的，所以我们无需担心出现一些奇怪的存在，直接把它认定为体积即可，而 $\mathbb{R}^n$ 中的紧集实际就是闭集的意思，所以这个定理稍加改造就可以给任何一个中学数竞生做了。由于对称且“实心”(由凸性给出)，一定有 $\mathbf{0} \in S \cap H$，所以我们的目的是再来一个格点，定理的意思就是体积达到一定程度的时候就一定能包含新的一个格点。我们考虑特殊情况，并稍加思考就能明白这个系数从哪里来，最简单的情况是 $\mathbb{Z}^2 \subset \mathbb{R}^2$，我们只需在原点周围的格点围一圈就得到了上界集

$$S = \{(x,y) \in \mathbb{R}^2 : -1 < x, y < 1\}$$

这是无法加上新点的极限了，只要稍微凸出一点，就有了新的格点，所以在 (1) 中使用的是严格大于号，如果加上紧集的条件，我们则必需往回退一点才能保证不触碰新点，所以在 (2) 中使用的是大等号。

证明 先来证明一个简单的性质，$\mu(S) > V(H) \Rightarrow \exists s, s' \in S, s - s' \in H$。记 $P = \mathbb{R}^n/H$，对任意的 $h \in H$ 集合 $h + P = \{h + p \in \mathbb{R}^n : p \in P\}$ 互不相交，且有 $\mathbb{R}^n = \cup_{h\in H} h + P$，我们用它将 S 进行切割

$$\mu(S) = \sum_{h\in H} \mu(S \cap (h + P)) = \sum_{h\in H} \mu((-h + S) \cap P)$$

后一半是反向处理，即将 $S \cap (h + P)$ 部分平移回 P 内得到 $(S - h) \cap P$ 部分。由于 $\mu(S) > V(H)$，所以一定存在两个碎片 $h \neq h'$ 使得它们有交点 $x \in ((-h + S) \cap P) \cap ((-h' + S) \cap P)$，此时只需要令

$$x = -h + s = -h' + s', s, s' \in S$$

就有 $0 \neq s - s' \in H$。

(1) 对于通常情形，我们构造 $S' = \frac{1}{2}S = \{\frac{s}{2} \in \mathbb{R}^n : s \in S\}$，根据条件可得

$$\mu(S') = \frac{\mu(S)}{2^n} > V(H)$$

由上述性质，存在 $x, y \in S'$ 使得 $x \neq y, x - y \in H$。此时由对称凸性可得

$$2x, -2y \in S \Rightarrow x - y = \frac{1}{2}((2x) + (-2y)) \in S$$

即有 $0 \neq x - y \in S \cap H$。

(2) 对于紧集情况，我们只需采用类似的方法使用逼近边界即可，构造 $S_m = (1+\frac{1}{m})S = \{(1+\frac{1}{m})s \in \mathbb{R}^n : s \in S\}$，此时

$$\mu(S_m) = (1 + \frac{1}{m})^n \mu(S) > 2^n V(H)$$

利用 (1) 的结论，存在 $0 \neq h_m \in S_m \cap H$，由于 $\lim\limits_{m\to\infty} S_m = S$ 且 S 是紧集，故 $h = \lim\limits_{m\to\infty} h_m \in H$，又由于 H 在 $\mathbb{R}^n$ 中离散且 h_m 非零，故 $h \neq 0$。即有 $0 \neq h \in S \cap H$。

类数公式

我们来稍微回忆一下数域 K 的理想类群 C_K 的定义。首先，$O_K = \overline{\mathbb{Z}} \cap K$ 表示代数整数环，其表示数域 K 中代数整数构造的整体；接着，这个环的所有非零理想构成的集合 I_K^o 配上理想的乘法 $IJ = \{\sum a_i b_i \in O_K : a_i \in I, b_i \in J\}$ 构成一个交换幺半群；然后，借助交换幺半群的唯一扩充性可以得到分式理想群 I_K，并把 I_K^o 的元素称为整理想；最后，借助其主分式理想子群 $P_K \leq I_K$，即可得到理想类群 $C_K = I_K/P_K$。一个简单的事实是，任意一个分式理想 I 都存在整理想 $A, B \in I_K^o$ 满足 $IB = A$，或写为 $I = \frac{A}{B} = AB^{-1}$，我们定义整理想的范为 $N_K(A) = |O_K/A|, A \in I_K^o$，并相应地定义分式理想的范为 $N_K(I) = \frac{N_K(A)}{N_K(B)}$，理想范的性质主要集中在整理想上，我们单单就列举一下，其非常自然而且证明也不困难，读者可以自行探索。

定理 4.3 (理想范的性质)

设 $A, B \in I_K^o$，且有唯一素理想分解 $A = \mathfrak{p}_1^{e_1} \ldots \mathfrak{p}_r^{e_r}, e_i \geq 1$。

(1) $N_K(A) = N_K(\mathfrak{p}_1)^{e_1} \ldots N_K(\mathfrak{p}_r)^{e_r}$

(2) $N_K(AB) = N_K(A)N_K(B)$

(3) 对 A 的任意一组基 $\{a_1, \ldots, a_n\}$ 有 $d_K(a_1, \ldots, a_n) = N_K(A)^2 d(K)$

(4) 若 $A = (\alpha), \alpha \in O_K$ 为主理想，则有 $N_K(A) = N_{K/\mathbb{Q}}(\alpha)$

♡

想要证明类数的有限性，实际就是证明理想类 $I + P_K \in C_K$ 的个数是有限的。接下来的思路比较奇怪，假设我们证明了存在一个上界 C 使得每个理想类 $I + P_K$ 均存在一个满足 $N_K(A) \leq C$ 的整理想 A。注意到 $N_K(A) \in \mathbb{Z}_{>0}$，而且如果有 $N_K(A) = |O_K/A| = q$，则有 $A \cap O_K = (q) \Leftrightarrow A \mid (q)O_K$，由 $(q)O_K$ 的素理想分解可知，其最多只能组合出有限个整理想因子。因此满足 $N_K(A) \leq C$ 的整理想 A 只有有限多个，也就是说类数 $h_K = |C_K|$ 是有限的。而这个上界 C 和相应的假设都是可以证明的，即下面的定理。

定理 4.4

数域 K 的每个理想类 $I+P_K\in C_K$ 均存在一个整理想 A 使得

$$N_K(A)\leq(\frac{4}{\pi})^{r_2}\frac{n!}{n^n}\sqrt{|d(K)|}$$

♡

注 定理中的大多数，都是确定数域 K 后就能确定的，比如指数 $n=[K,\mathbb{Q}]$、复素点个数 r_2、域的行列式 $d(K)=d_{K/\mathbb{Q}}(1,\alpha,...,\alpha^{n-1}),K=\mathbb{Q}(\alpha)$，所以得到的就是一个常数 C，有些书喜欢称其为 Minkowski 常数，不过对它叫什么我无所谓，知道是这么一回事就行了。而证明这个定理，我们主要分三步，首先，构造非零整理想 $\mathfrak{a}$ 到格 $\sigma(\mathfrak{a})$ 的对应，并得到相应的体积值 $V(\sigma(\mathfrak{a}))=2^{-r_2}N_K(\mathfrak{a})\sqrt{|d(K)|}$；其次，得到其一个非零元素 $0\neq x\in\mathfrak{a}$ 的范数值估计 $|N_{K/\mathbb{Q}}(x)|$；最后，得到我们的结论。

证明 (1) 根据 $n=r_1+2r_2$，我们把实嵌入记为 $\sigma_i:K\hookrightarrow\mathbb{R}(1\leq i\leq r_1)$，复嵌入记为 $\sigma_{r_1+j}=\overline{\sigma}_{r_1+r_2+j}:K\hookrightarrow\mathbb{C}(1\leq i\leq r_2)$。考虑一个非零整理想 $\mathfrak{a}$，我们可以给出

$$\sigma:\mathfrak{a}\subset K\to\mathbb{R}^n,a\mapsto(\sigma_1(a),...,\sigma_{r_1}(a),\mathrm{Re}(\sigma_{r_1+1}(a)),...,\mathrm{Re}(\sigma_{r_1+r_2}(a)),\mathrm{Im}(\sigma_{r_1+1}(a)),...,\mathrm{Im}(\sigma_{r_1+r_2}(a)))$$

由于整理想是自由阿贝尔群，故有直和分解 $\mathfrak{a}=\mathbb{Z}a_1\oplus...\oplus\mathbb{Z}a_n$。设 $\mathbb{R}^n$ 的标准正交基为 $\{e_i\}$，并记变换为 $\sigma(a_i)=\sum_{j=1}^{n}x_{ij}e_j$，即有

$$x_{ij}=\begin{cases}\sigma_j(a_i) & 1\leq j\leq r_1\\ \mathrm{Re}(\sigma_j(a_i)) & r_1+1\leq j\leq r_1+r_2\\ \mathrm{Im}(\sigma_j(a_i)) & r_1+r_2+1\leq j\leq n\end{cases}$$

这样我们环的结构就已经出来了，$\sigma(\mathfrak{a})=\mathbb{Z}\sigma(a_1)\oplus...\oplus\mathbb{Z}\sigma(a_n)$，而对于基的线性无关性可以通过算体积得到

$$V(\sigma(\mathfrak{a}))=|\det(x_{ij})|=2^{-r_2}|\det(\sigma_j(a_i))|=2^{-r_2}\sqrt{|d_K(a_1,...,a_n)|}=2^{-r_2}N_K(\mathfrak{a})\sqrt{|d(K)|}$$

它给出了 $\det(x_{ij})\neq0$，从而 $\sigma(\mathfrak{a})$ 是我们想要的格。

(2) 为了使用 Minkowski 定理，我们还需要构造一个关于原点对称的紧凸集，即下面这个

$$B_t=\{y=(y_1,...,y_n)\in\mathbb{R}^n:\sum_{i=1}^{r_1}|y_i|+2\sum_{j=1}^{r_2}\sqrt{y_{r_1+j}^2+y_{r_1+r_2+j}^2}\leq t\}$$

我们令 $t^n=(\frac{4}{\pi})^{r_2}N_K(\mathfrak{a})\sqrt{|d(K)|}n!$，可以计算出它的体积为

$$\mu(B_t)=\int_{B_t}\cdots\int\mathrm{d}y_1...\mathrm{d}y_n=2^{r_1}(\frac{\pi}{2})^{r_2}\frac{t^n}{n!}=2^nV(\sigma(\mathfrak{a}))$$

故存在 $0\neq x\in\mathfrak{a}$ 使得 $\sigma(\mathfrak{a})\in B_t$，即有

$$|N_{K/\mathbb{Q}}(x)|=\prod_{i=1}^{n}|\sigma_i(x)|\leq(\frac{1}{n}\sum_{i=1}^{n}|\sigma_i(x)|)^n\leq\frac{t^n}{n^n}=(\frac{4}{\pi})^{r_2}\frac{n!}{n^n}\sqrt{|d(K)|}N_K(\mathfrak{a})$$

(3) 对任意一个理想类 $C=I+P_K\in C_K$，可知 $\forall n\in\mathbb{Z},\frac{1}{n}I\in C$，故一定能找到一个整理想 $\mathfrak{a}$ 使得 $\mathfrak{a}'=\mathfrak{a}^{-1}\in C$。由 (2) 可知存在 $0\neq x\in\mathfrak{a}$ 满足一个不等条件，此时我们引入

$$A=x\mathfrak{a}^{-1}=x\mathfrak{a}'\in C$$

接着可以计算出我们的上界

$$N_K(A)=N_K(x)N_K(\mathfrak{a}')\leq(\frac{4}{\pi})^{r_2}\frac{n!}{n^n}\sqrt{|d(K)|}N_K(\mathfrak{a})N_K(\mathfrak{a}')=(\frac{4}{\pi})^{r_2}\frac{n!}{n^n}\sqrt{|d(K)|}N_K(O_K)=(\frac{4}{\pi})^{r_2}\frac{n!}{n^n}\sqrt{|d(K)|}$$

比起类数的有限性，上面的不等式估计反而更有意义，所以我们才把它作为核心证明的定理。借助这个不等式我们可以估计数域 K 的行列式值，假设 $n=[K,\mathbb{Q}]\geq2$，由于非零整理想 A 一定满足 $N_K(A)\geq1$，故反过来可估计得到

$$|d(K)|\geq((\frac{\pi}{4})^{r_2}\frac{n^n}{n!}N_K(A))^2\geq(\frac{\pi}{4})^n(\frac{n^n}{n!})^2\geq\frac{\pi}{3}(\frac{3\pi}{4})^{n-1}>1$$

这表明至少有一个素数在 K 中分歧。借助上一个不等式链的倒二个值可知，存在一个绝对常数 M，使得每个数域 K 均有

$$n \leq M \ln(|d(K)|), M = (\ln(\frac{4}{9}) \ln(\frac{3\pi}{4}))^{-1}$$

此时如果固定 $d(K) = d$，则 n 的取值有限，且更进一步可知 r_1, r_2 取值有限，利用 Minkowski 定理和类似的方法可以证明均存在单代数扩张 $K = \mathbb{Q}(x)$，且 x 被限定在由 n, d, r_1, r_2 确定的上界中。于是我们可以得到满足 $d(K) = d \in \mathbb{Z}$ 的数域个数是有限的。实际上，我们有下面的类数公式，但它们的证明涉及解析数论的方法，所以对于其证明我们下一章再说。

$$\lim_{s \to 1}(s-1)\zeta_K(s) = \frac{2^{r_1}(2\pi)^{r_2} R_K h_K}{\omega_K \sqrt{|d_K|}}$$

$$\lim_{s \to 0} s^{-r_1-r_2+1}\zeta_K(s) = -\frac{R_K h_K}{\omega_K}$$

单位定理

单位群 $O_K^\times$ 表示代数整数环 O_K 中所有乘法可逆元 (也称为单位) 构成的全体，而在它的分解 $O_K^\times = \mathbb{Z}^{\oplus r} \oplus W_K$ 中，$\mathbb{Z}^{\oplus r}$ 表示 r 个整数环的直和，即秩为 r 的自由阿贝尔群，W_K 称为 K 的**单位根群**，由 K 中的所有单位根[1]构成，是一个有限阶循环群，以下是些简单的例子。

例题 4.1 (1)$K = \mathbb{R}, W_K = \{-1, 1\}$

(2)$K = \mathbb{Q}(i), W_K = \{\pm 1, \pm i\}$

(3)$K = \mathbb{Q}(\sqrt{-3}), W_K = \{\pm 1, \pm\omega, \pm\omega^2\}, \omega = \frac{1}{2}(-1 + \sqrt{-3})$

我们需要注意到，O_K 在加法上是自由阿贝尔群，即有 $O_K = \mathbb{Z}\omega_1 \oplus ... \oplus \mathbb{Z}\omega_n$，其中 $n = [K, \mathbb{Q}]$ 是指数，$\omega_1, ..., \omega_n \in O_K$ 是一组存在性的基且并不唯一。对于乘法，我们则考虑单位群 $O_K^\times \subset O_K$，则它一定是有限生成阿贝尔群，根据结构定理其一定存在自由子群和挠子群的分解 $O_K^\times = F_K \oplus T_K$，我们想要证明的即是将这两部分给精确化为

$$\text{自由部分:} F_K \cong \mathbb{Z}^{\oplus r} \quad \text{挠部分:} T_K \cong W_K$$

方便起见，我们先简记 $U_K = O_K^\times$，首先我们需要一个计算判定归属的方法，即用某些计算值来判断元素是否为单位，是否为单位根，即有下面定理。

定理 4.5

(1)$u \in U_K \Leftrightarrow N_{K/\mathbb{Q}} = \pm 1$

(2)$u \in W_K \Leftrightarrow \forall 1 \leq i \leq n, |\sigma_i(u)| = 1$

♡

上述性质的理解和证明都不困难，甚至十分地自然，比如 $u \in U_K$ 可逆，则意味着存在 u^{-1} 且有 $N(u)N(u^{-1}) = 1$，结合范的取值为整数，那么可能的结果也就只有 $\mathbb{Z}$ 的可逆元 $\{-1, 1\}$ 了。接着我们可以来分步完成单位定理的证明了。

(1) 构造格 $l(U_K)$，并分离出挠子群 W_K

设 $[K, \mathbb{Q}] = n = r_1 + 2r_2$，$\sigma_i : K \hookrightarrow \mathbb{R}(1 \leq i \leq r_1)$ 为实嵌入，$\sigma_{r_1+j} = \overline{\sigma}_{r_1+r_2+j} : K \hookrightarrow \mathbb{C}(1 \leq i \leq r_2)$ 为复嵌入，U_K 为单位群，W_K 为单位根群。映入映射

$$l : U_K \to \mathbb{R}^{r_1+r_2}, u \mapsto (\ln|\sigma_1(u)|, ..., \ln|\sigma_{r_1}(u)|, 2\ln|\sigma_{r_1+1}(u)|, ..., 2\ln|\sigma_{r_1+r_2}(u)|)$$

[1]所谓 K 的单位根 ω，即存在正整数 n 使得 $\omega^n = 1$

由一些列简单的推导

$$u \in \ker l \Leftrightarrow \forall 0 \le i \le r_1 + r_2, \ln|\sigma_i(u)| = 0 \Leftrightarrow \forall 0 \le i \le n, |\sigma_i(u)| = 1 \Leftrightarrow u \in W_K$$

故有 $\ker l = W_K$，由 $U_K = \ker l \oplus \mathrm{im} l$ 可得 $U_K = W_K \oplus l(U_K)$。注意到对 $u \in U_K, l(u)$ 的分量直和为

$$\ln|\sigma_1(u)| + ... + \ln|\sigma_{r_1}(u)| + 2\ln|\sigma_{r_1+1}(u)| + ... + 2\ln|\sigma_{r_1+r_2}(u)| = \sum_{i=1}^{n} \ln|sigma_i(u)| = \ln|N_{K/\mathbb{Q}}(u)| = 0$$

因此 $l(U_K)$ 不仅是 $\mathbb{R}^{r_1+r_2}$ 的子群，还是下面超平面

$$H = \{(x_1, ..., x_{r_1+r_2}) \in \mathbb{R}^{r_1+r_2} : x_1 + ... + x_{r_1+r_2} = 0\}$$

的子群。此时易证，对任意有界集 $B \subset \mathbb{R}^{r_1+r_2}$，只存在有限多个 $u \in U_K$ 使得 $l(u) \in B$，即 $l(U_K) \cap B$ 是有限集。换句话说，$l(u)$ 是 $\mathbb{R}^{r_1+r_2}$ 的离散子群，从而是一个格，是 U_K 的自由子群，并且 $l(U_K) \subset H$ 给出了秩估计 $\mathrm{rank} l(u) \le r = r_1 + r_2 - 1$。

(2) 元素存在性定理

简便起见我们记 $l(a) = (a_1, ..., a_{r_1+r_2})$，我们将借助 Minkowski 定理来证明一个找元素的定理。即对 $0 \ne a \in O_K$ 和 $1 \le k \le r_1 + r_2$，我们可以找到一个 $0 \ne b \in O_K$ 满足 $i \ne k \Rightarrow b_i < a_i$。我们同样先构造一个关于原点对称的紧凸集

$$B = \{(x_1, ..., x_n) \in \mathbb{R}^n : |x_i| \le c_i (1 \le i \le r_1), x_j^2 + x_{j+r_2}^2 \le c_j (r_1 + 1 \le j \le r_1 + r_2)\}$$

$$c_k = (\frac{\prod_{i=1}^{n} c_i}{c_k})^{-1} (\frac{2}{\pi})^{r_2} \sqrt{|d(K)|}, 0 \le c_i \le e^{a_i} (i \ne k, 1 \le i \le r_1 + r_2)$$

接着我们计算它的体积

$$\mu(B) = 2^{r_1} c_1 ... c_r \pi^{r_2} c_{r_1+1} ... c_{r_1+r_2} = 2^n 2^{-r_2} \sqrt{|d(K)|} = 2^n V(l(O_K))$$

故存在 $0 \ne b \in O_K$ 使得 $l(b) \in B$ 且当 $i \ne k$ 时有

$$b_i \le c_i < a_i$$

而且其还满足一个范的约束不等式

$$|N(b)| \le c_1 ... c_{r_1+r_2} = (\frac{2}{\pi})^{r_2} \sqrt{d(K)}$$

(3) 构造 r 个线性无关元

接着我们固定一个 $1 \le k \le r_1 + r_2$，只需先任取 $a_1 \in O_K - \{0\}$ 开始，反复利用 (2) 的结论即可以得到一系列的 $a_2, a_3, ... \in O_K - \{0\}$，其均满足 $i \ne k \Rightarrow (a_{j+1})_i < (a_j)_i$，且范被一个上界控制 $|N(a_j)| \le M$。

由于范被一个上界控制，故主理想集 $\{a_j O_K : j = 1, 2, ...\}$ 是有限的，即存在 $j > h$ 使得 $a_j O_K = a_h O_K$，我们令 $a_j = u_k a_h$ 就有 $u_k \in U_K$，我们只需变换 k 就能得到一系列的元素 $\{u_1, ..., u_{r_1+r_2}\} \subset U_K$，我们记 $l(u_k) = (y_{k,1}, ..., y_{k,r_1+r_2})$，总共有 $r + 1 = r_1 + r_2$ 个 $r_1 + r_2$ 维向量。我们需要说明，里面有 $r = r_1 + r_2 - 1$ 个向量是线性无关的，即证明矩阵 $Y = (y_{i,j})$ 的秩为 r。

我们只需注意到两个事实，第一，由于 $l(U_K)$ 落在超平面 H 上，所以任意 $l(u_k)$ 的分量之和为零 $\sum_{i=1}^{r_1+r_2} y_{k,i} = 0$，即矩阵 Y 的每行元素之和为零，第二，由 (2) 的估计和 (3) 的构造可知，取定 k 后对任意 $i \neq k$ 均有 $y_{k,i} = (l(a_j))_i - (l(a_h))_i < 0$，由于和为零又可得 $y_{k,k} > 0$，即矩阵 Y 的除了对角元为正以外均为负。

由之前的结论可知 $\mathrm{rank}Y \leq r$，故一定有 $\det Y = 0$，即无需考虑有 $r+1$ 个线性无关元的情况。我们记 $m = r_1 + r_2 = r + 1$，并设

$$Y = (v_1, ..., v_m), v_i = \begin{pmatrix} y_{1,i} \\ \vdots \\ y_{m,i} \end{pmatrix} \text{是列向量}$$

假设 $\{v_i\}$ 中有 $m-1$ 个列向量是线性相关的，不失一般性设它们为前 $m-1$ 个列向量 $\{v_1, ..., v_{m-1}\}$，此时存在不全为零的 $t_i \in \mathbb{R}$ 使得

$$\sum_{i=1}^{m-1} t_i v_i = 0$$

设 $\max_{1 \leq i \leq m-1} |t_i| = |t_k|$，此时通过适当的乘法，我们可以调整 $t_k = 1$ 且 $|t_j| \leq 1 (1 \leq j \leq m-1, j \neq k)$，于是在 Y 的第 k 行将会有

$$0 = \sum_{j=1}^{m} y_{k,j} < \sum_{j=1}^{m-1} y_{k,j} \leq \sum_{j=1}^{m-1} t_j y_{k,j} = 0$$

故 $\{v_i\}$ 中任意 $m-1$ 个列向量是线性无关的，即 $\mathrm{rank}Y = m - 1 = r$。

应用

单位定理实际就是在说任意的 $u \in U_K$ 可以唯一地表示为

$$u = \omega u_1^{a_1} ... u_r^{a_r}, \omega \in W_K, a_i \in \mathbb{Z}$$

我们把 $\{u_1, ..., u_r\}$ 称为 K 的一个**基本单位组**。其具有强大的力量，我们可以用它得到不少的数论命题。考虑实二次域 $K = \mathbb{Q}(\sqrt{d}), d > 0$ 且无平方因子，容易算得 $n = 2, r_1 = 2, r_2 = 0, r = 1, W_K = \{-1, 1\}, U_K = \mathbb{Z} \oplus W_K$，即 U_K 只有一个基本单位 ε 并且可以表示为 $U_K = \{\pm\varepsilon^n : n \in \mathbb{Z}\}$。由单位的性质可知，只有 $\pm\varepsilon, \pm\varepsilon^{-1}$ 可以作为基本单位，故满足 $\varepsilon > 1$ 的基本单位是唯一的。

我们进一步限制 $d \equiv 2, 3 (\mathrm{mod} 4)$，则有 $O_K = \mathbb{Z}[d] = \{a + b\sqrt{d} : a, b \in \mathbb{Z}\}$，于是有等价链

$$a + b\sqrt{d} \in U_K \Leftrightarrow N(a + b\sqrt{d}) = a^2 - db^2 = \pm 1 \Leftrightarrow (a, b) \text{是方程} x^2 - dy^2 = \pm 1 \text{的整数解}$$

定理 4.6

设 $K = \mathbb{Q}(\sqrt{d}), d \equiv 2, 3 (\mathrm{mod} 4), d > 0$ 且无平方因子，$\varepsilon = a + b\sqrt{d} \in U_K$ 且 $\varepsilon > 1$，继续设 $\varepsilon^n = (a + b\sqrt{d})^n = a_n + b_n\sqrt{d}, a_n, b_n \in \mathbb{Z}$

(1) 当 $N(\varepsilon) = 1$ 时，方程 $x^2 - dy^2 = -1$ 无整数解，方程 $x^2 - dy^2 = 1$ 的全部整数解为 $\{(\pm a_n, \pm b_n) : n \in \mathbb{Z}\}$

(2) 当 $N(\varepsilon) = -1$ 时，方程 $x^2 - dy^2 = -1$ 的全部整数解为 $\{(\pm a_{2n+1}, \pm b_{2n+1}) : n \in \mathbb{Z}\}$，方程 $x^2 - dy^2 = 1$ 的全部整数解为 $\{(\pm a_{2n}, \pm b_{2n}) : n \in \mathbb{Z}\}$。

这就是 Pell 方程解的结构定理，证明比较简单就不写了，有关像 Pell 方程这样的不定方程，后面还会有更多的探讨，这里我们就点到为止了。另一种情况是考虑虚二次 $K = \mathbb{Q}(i)$，它可以获得高斯整数环 $\mathbb{Z}[i] = O_K$，其满足 $n = 2, r_1 = 0, r_2 = 1, r = 0$，故 $U_K = W_K = \{\pm 1, \pm i\}$ 为有限群，实际我们有下面的更一般性结论

$$U_K\text{是有限群} \Leftrightarrow K=\mathbb{Q}\text{或}K\text{为虚二次域}$$

由于后面的需要，我们再来讨论一下分圆域 $K=\mathbb{Q}(\zeta_{p^t})$ 的单位结构，其中 $\zeta_{p^t}=e^{\frac{2\pi i}{p^t}}, t\geq 1, p$ 为奇素数，容易知道 W_K 是 $2p^t$ 阶循环群且有

$$n=[K,\mathbb{Q}]=\varphi(p^t)=p^t-p^{t-1}, r_1=0, r_2=\frac{1}{2}\varphi(p^t), r=\frac{1}{2}\varphi(p^t)-1$$

此时我们可以找到它的一个极大实子域 $K_+=\mathbb{Q}(\zeta_{p^t}+\zeta_{p^t}^{-1})$，并且有 $[K_+:\mathbb{Q}]=\frac{1}{2}\varphi(p^t), r=\frac{1}{2}\varphi(p^t)-1$，于是我们可以找到一组它们共同的实基本单位组

$$\{u_1,...,u_r\}, r=\frac{1}{2}\varphi(p^t)-1$$

4.2 费马大定理

或许代数数论最辉煌的成就之一就是证明了 Kummer 定理，它证明了费马大定理几乎无限多种情形，所谓的**费马大定理**就是：当 $n\geq 3$ 时，方程 $x^n+y^n=z^n$ 没有满足 $xyz\neq 0$ 的整数解。我在以前的视频中说过完整的证明是，先约化为 $n=p\geq 5$ 为素数的情形，再构造椭圆曲线 $E_{a^p+b^p=c^p}: y^2=x(x-a^p)(x-c^p)$，最后通过模性定理说明其不存在即可。而我们将利用代数数论完成下面定理的证明。

定理 4.7 (Kummer 定理其一)

设 p 为奇素数，h_p 为分圆域 $K=\mathbb{Q}(\zeta)(\zeta=\zeta_p)$ 的类数，如有 $p\nmid h_p$，则费马方程 $x^p+y^p=z^p$ 没有满足 $p\nmid xyz\neq 0$ 的整数解。

♡

注 开始证明之前，我们先进行一些约简，从而给自己更多的条件。在上述加了 $p\nmid xyz$ 的条件下，如果 $p=3$ 我们可以很容易地推出 $x^3,y^3,z^3\equiv\pm 1(\mathrm{mod}9)$，此时可得 $x^3+y^3\equiv 0,\pm 2\neq\pm 1\equiv z^3(\mathrm{mod}9)$，故我们可以进一步假设 $p\geq 5$。接着，如果 $x\equiv y\equiv -z(\mathrm{mod}p)$，则可以推出 $p\nmid z, 3z^p\equiv 0(\mathrm{mod}p)$ 矛盾，即我们可以假设 $x\not\equiv y(\mathrm{mod}p)$。最后，如果 $d=(x,y)$ 则有 $d\mid z$，$(x/d,y/d,z/d)$ 也是原方程的解，所以还能假设 x,y,z 两两互素。带着假设，在整数环 $\mathbb{Z}[\zeta]$ 中我们把 $x^p+y^p=z^p$ 重写为

$$(x+y)(x+\zeta y)...(x+\zeta^{p-1}y)=z^p$$

证明 (1) 我们需要先了解 $\mathbb{Z}[\zeta]$ 的一些性质

(i)$\mathbb{Z}[\zeta]$ 中的 p 个主理想 $(x+\zeta^i y)$ 两两互素。反之，存在 $i<j$ 和公共素理想因子 $\mathfrak{p}$ 使得 $x+\zeta^i y\equiv x+\zeta^j y\equiv 0(\mathrm{mod}\mathfrak{p})$，从而

$$0\equiv(x+\zeta^i y)-(x+\zeta^j y)\equiv\zeta^i y(1-\zeta^{j-i})\equiv\zeta^i\frac{1-\zeta^{j-i}}{1-\zeta}y(1-\zeta)(\mathrm{mod}\mathfrak{p})$$

由于 $\zeta, \frac{1-\zeta^{j-i}}{1-\zeta}\in U_K$，故 $y(1-\zeta)\equiv 0(\mathrm{mod}\mathfrak{p})$，类似地由

$$0\equiv\zeta^j(x+\zeta^i y)-\zeta^i(x+\zeta^j y)=\zeta^i x(1-\zeta^{j-i})(\mathrm{mod}\mathfrak{p})$$

给出 $x(1-\zeta)\equiv 0(\mathrm{mod}\mathfrak{p})$。接着由于 $(x,y)=1$ 互素，故有 $1-\zeta\equiv 0(\mathrm{mod}\mathfrak{p})$，即 $\mathfrak{p}|(1-\zeta)$。由于 $(1-\zeta)$ 是 $\mathbb{Z}[\zeta]$ 的素理想且有分歧 $pO_K=(1-\zeta)^{p-1}$ 从而有 $\mathfrak{p}=(1-\zeta)=(1-\zeta^i)$。进一步推导可得

$$x+y\equiv x+\zeta^i y\equiv 0(\mathrm{mod}\mathfrak{p}), x+y\in\mathfrak{p}\cap\mathbb{Z}=p\mathbb{Z}, x+y\equiv 0(\mathrm{mod}p)$$

由此可得 $z^p\equiv x^p+y^p\equiv x+y\equiv 0(\mathrm{mod}p), p\mid z$，与 $p\nmid xyz$ 矛盾。

(ii) 对每个 $a\in\mathbb{Z}[\zeta]$ 均存在 $b\in\mathbb{Z}$ 使得 $a^p\equiv b(\mathrm{mod}p)$，我们只需要将 $a=a_0+a_1\zeta+...+a_{p-1}\zeta^{p-1}$ 的形式表达出来，再直接计算

$$a^p\equiv a_0^p+a_1^p\zeta^p+...+a_{p-1}^p(\zeta^{p-1})^p=a_0^p+a_1^p+...+a_{p-1}^p(\mathrm{mod}p)$$

即可构造出 $b=a_0^p+a_1^p+\ldots+a_{p-1}^p\in\mathbb{Z}$。

(iii) 如果 $n\mid a=a_0+a_1\zeta+\ldots+a_{p-1}\zeta^{p-1}$ 满足至少有一个 $a_i=0$，则 $\forall i,n\mid a_i$。这是整基的显然性质，即 $\{1,\zeta,\ldots,\zeta^{p-1}\}$ 中任意 $p-1$ 个元素都是整基。虽然平常我们都写完全，但实际做基时可以少一个元素，比如 $p=3$ 时，$a_0+a_1\omega+a_2\omega^2=a_0+a_1\omega+a_2(-1-\omega)=(a_0-a_2)+(a_1-a_2)\omega$。

(2) 我们将原来的方程分解式看成理想的分解式

$$\prod_{i=0}^{p-1}(x+\zeta^i y)=(z)^p$$

由于左边的理想两两互素，故存在理想 $\mathfrak{a}_i$ 使得 $(x+\zeta^i y)=\mathfrak{a}_i^p\in P_K$，即 $\mathfrak{a}_i$ 的等价类满足 $(\mathfrak{a}_i+P_K)^p=1$ 为理想类群 C_K 的单位元。由于 $p\nmid h_K$，即 C_K 中没有 p 阶元，故只能让 $\mathfrak{a}_i+P_K=1$ 为单位元，即 $\mathfrak{a}_i\in P_K$ 为主理想。我们设 $\mathfrak{a}_i=(a),a\in\mathbb{Z}[\zeta]$，则有 $x+\zeta^i y=\varepsilon a^p,\varepsilon\in U_K,\varepsilon=\zeta^r\varepsilon_1,\varepsilon_1\in\mathbb{R}$，我们考虑 $i=1$ 的情况，则有

$$x+\zeta y=\zeta^r\varepsilon_1 a^p\equiv\zeta^r\varepsilon_1 b(\mathrm{mod}p)$$

类似地有 $\overline{x+\zeta y}\equiv\zeta^{-r}\varepsilon_1 b(\mathrm{mod}p)$ 故有 $\zeta^{-r}(x+\zeta y)\equiv\zeta^r\overline{(x+\zeta y)}(\mathrm{mod}p)$，化简可得

$$x+\zeta y-\zeta^{2r}x-\zeta^{2r-1}y\equiv 0(\mathrm{mod}p)$$

如果 $\{1,\zeta,\zeta^{2r},\zeta^{2r-1}\}$ 互不相同则有 $p\mid x,p\mid y$，与 $p\nmid xyz$ 矛盾。故其中至少有两个相等，又显然有 $1\neq\zeta,\zeta^{2r}\neq\zeta^{2r-1},1=\zeta^{2r-1}\Leftrightarrow\zeta=\zeta^{2r}$，故只有 $\mathrm{C}_4^2-3=3$ 种情况需要再做讨论。

(3)(i) 当 $1=\zeta^{2r}$ 时，上式可化为 $\zeta y-\zeta^{-1}y\equiv 0(\mathrm{mod}p)$，即有 $p\mid y$，与 $p\nmid xyz$ 矛盾。

(ii) 当 $1=\zeta^{2r-1}\Leftrightarrow\zeta=\zeta^{2r}$ 时，上式可化为 $(x-y)-(x-y)\zeta\equiv 0(\mathrm{mod}p)$，即有 $p\mid x-y$，与 $x\not\equiv y(\mathrm{mod}p)$ 矛盾。

(iii) 当 $\zeta=\zeta^{2r-1}$ 时，上式可化为 $x-\zeta^2 x\equiv 0(\mathrm{mod}p)$，即有 $p\mid x$，与 $p\nmid xyz$ 矛盾。

定理 4.8 (Kummer 定理其二)

设 p 为奇素数，h_p 为分圆域 $K=\mathbb{Q}(\zeta)(\zeta=\zeta_p)$ 的类数，如有 $p\nmid h_p$，则费马方程 $x^p+y^p=z^p$ 没有满足 $p\mid xyz\neq 0$ 的整数解。♡

注 我们可以试着把定理加强，有时反而可能会更好证明。我们把无整数解加强为无 $x,y,z\in\mathbb{Z}[\zeta]$ 的解，由于 $(1-\zeta)|p$(可以由主理想 (p) 在 $\mathbb{Z}[\zeta_p]$ 上分歧得到)，所以我们把满足 $p\mid xyz$ 的条件加强为 $(1-\zeta)|xyz$，读者需要注意我们所加的条件都是在扩大可能的无解范围。然后加上 x,y,z 两两互素是毫无悬念的简化，由互素性和上一个条件，我们可以进一步假设 $(1-\zeta)\mid z$，即令 $z=(1-\zeta)^m z_0,z_0\in\mathbb{Z}[\zeta]$，此时条件为 $(1-\zeta)\nmid xyz_0$，设任一单位 $\varepsilon\in U_K$，方程进一步强化为 (取 $\varepsilon=1$ 就是原来的方程了)

$$x^p+y^p=\varepsilon(1-\zeta)^{pm}z_0^p$$

至于怎么证明，其与我们代数数论不是主要关系，所以不做深入探究，想了解的读者可以参考这本书 [30]。

4.3 岩泽理论

与代数数论有着紧密关系，又有着与解析数论的不浅联系，这就是我们将要介绍的岩泽理论 (Iwasawa Theory)，下面的大部分内容摘自这本书 [47] 的第十章。所谓岩泽理论实际就是指由日本数学家岩泽健吉所研究的理论，其本身应该具备的内涵是丰富的，而通常情况下指的就是两个关于理想类群和分圆域的定理，当然了各种等价和变形都看成一个，我想在这里指出，如果真的想要理解和研究岩泽理论的话，就不应该去看像“岩泽理论简介”之类的文章，比如我将要写的内容，而应该去看分圆域的相关书籍，比如 GTM 的这几本书 [5, 21, 33]。在数学领域中一般很少给理论冠上人名，冠上人名的理论一般都是在一个理论进行了很深入的研究、获得了一些艰辛的结论、并且具有深挖的可能，不拿分圆域之于岩泽理论来说事，在以前就有不少的例子，抽象代数之于伽罗瓦理论、微分流形之于黎曼几何，其中的最后一点也是很重要的，如果这些结论没有继续探索的方向，或者说没有留下一些猜想，或者说没有人去研究它，那么它的最后就是 XX 定理，而不是 XX 理论，不过岩泽理论也

算是吃上了时代红利。

岩泽公式

首先，我们要将岩泽理论所研究的对象给构造出来，并引出作为岩泽理论起点的岩泽公式。对于数域 K，如果它的扩域 K' 满足 $Gal(K'/K) \cong \mathbb{Z}_p$，就把 K' 称为 K 的 $\mathbb{Z}_p$ **扩域**，这里使用的是 p 进整数 $\mathbb{Z}_p$ 的加法群结构，虽然定义看起来挺奇怪的，但每个数域都可以构造出这样的一个扩域来。我们记 n 次单位根为 $\zeta_n = e^{\frac{2\pi i}{n}}$，考虑伽罗瓦群的同构

$$Gal(\mathbb{Q}(\zeta_{p^{n+1}})/\mathbb{Q}) \cong (\mathbb{Z}/p^{n+1}\mathbb{Z})^\times \cong (\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/p^n\mathbb{Z})$$

前半部分是分圆域的经典性质了，我们只需把 $\sigma(\zeta_n) = \zeta_n^m, \sigma \in Gal(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ 对应到 $m \in (\mathbb{Z}/n\mathbb{Z})^\times$ 即可，后半部分来自于 p 进单位的性质 $\mathbb{Z}_p^\times = (\mathbb{Z}/p\mathbb{Z})^\times \times (1+p\mathbb{Z}_p)$，然后使用指对数变换 $\exp : \mathbb{Z}_p \to (1+p\mathbb{Z}_p), \log : (1+p\mathbb{Z}_p) \to \mathbb{Z}_p$ 进一步转化为 $\mathbb{Z}_p^\times = (\mathbb{Z}/p\mathbb{Z})^\times \times \mathbb{Z}_p$，接着使用逆向极限来看待 p 进整数就有 $(\varprojlim_n \mathbb{Z}/p^{n+1}\mathbb{Z})^\times = (\mathbb{Z}/p\mathbb{Z})^\times \times \varprojlim_n \mathbb{Z}/p^n\mathbb{Z}$，分离可得我们想要的结果。交换群是显然的，所以 $(\mathbb{Z}/p\mathbb{Z})^\times$ 可以视为正规子群 $(\mathbb{Z}/p\mathbb{Z})^\times \times \{1\}$ 作用在 $\mathbb{Q}(\zeta_{p^{n+1}})$ 上，根据伽罗瓦理论，我们可以得到相应的不变子域，我们设其为 $\mathbb{Q}_n = \mathbb{Q}(\zeta_{p^{n+1}})^{(\mathbb{Z}/p\mathbb{Z})^\times} \subset \mathbb{Q}(\zeta_{p^{n+1}})$ 且有同构 $Gal(\mathbb{Q}_n/\mathbb{Q}) \cong \mathbb{Z}/p^n\mathbb{Z}$，我们令 $\mathbb{Q}_\infty = \cup_n \mathbb{Q}_n$，则有

$$Gal(\mathbb{Q}_\infty/\mathbb{Q}) \cong \varprojlim_n \mathbb{Z}/p^n\mathbb{Z} = \mathbb{Z}_p$$

接着，扩张到一般数域 K 上，只需简单的相乘 $K_\infty = K\mathbb{Q}_\infty = \{kq \mid k \in K, q \in \mathbb{Q}_\infty\}$ 即有

$$Gal(K_\infty/K) \cong Gal(\mathbb{Q}_\infty/K \cap \mathbb{Q}_\infty) \cong p^n\mathbb{Z}_p \cong \mathbb{Z}_p$$

第一部分是群的同构定理，第二部分则是因为 $Gal(\mathbb{Q}_\infty/\mathbb{Q}) \cong \mathbb{Z}_p$ 的闭子群只能是 $p^n\mathbb{Z}_p$ 的形式，最后一部分是简单的平移同构。接着我们来换个视角，即假设已知 $\mathbb{Z}_p$ 扩张 $K_\infty/K, Gal(K_\infty/K) \cong \mathbb{Z}_p$ 看看能得到什么，一个简单的想法是考虑伽罗瓦群的正规闭子群 $\mathbb{Z}/p^n\mathbb{Z}$，然后考虑伽罗瓦对应得到的同构 $Gal(K_n/K) \cong \mathbb{Z}/p^n\mathbb{Z}$，即可以获得一个域扩张塔

$$K = K_0 \subset K_1 \subset \ldots \subset K_\infty = \cup_n K_n$$

最后对于得到的这些对象 $K = K_0, K_1, K_2, \ldots, K_\infty = \cup_n K_n$，我们有下面的著名定理。

定理 4.9 (Iwasawa’s Theorem)

对于数域 K 的一个 $\mathbb{Z}_p$ 扩张 K_∞/K，设得到的扩张塔为 $K = K_0 \subset K_1 \subset \ldots \subset K_\infty$，并记 $h_n = |C_{K_n}|$ 为 K_n 的理想类群 C_{K_n} 的元素个数，$p^{e_n} \mid h_n, p^{e_n+1} \nmid h_n$(或简写为 $p^{e_n} \parallel h_n$)。则存在整数 $\lambda > 0, \mu > 0, \nu, n_0$ 使得任意 $n \geq n_0$ 都有

$$e_n = \lambda n + \mu p^n + \nu$$

♡

对于此定理我们需要知道，其条件应该是数域 K 和 $\mathbb{Z}_p$ 扩域 $K_\infty \subset K(\zeta_{p^\infty}) = \cup_n K(\zeta_{p^n})$，我们所给的例子也叫做**分圆** $\mathbb{Z}_p$ **扩域**，由于伽罗瓦群同构并不能得到扩域同构，所以 $\mathbb{Z}_p$ 扩域本身可能不唯一，而后面的扩张塔则是由伽罗瓦理论导出是唯一存在确定的，所以四个常数 λ, μ, ν, n_0 依赖于 K 和 K_∞。实际上还能得到对于我们所给的分圆 $\mathbb{Z}_p$ 扩域有 $\mu = 0$，我们来给些更具体的例子，条件 $p^{e_n} \parallel h_n$ 实际在告诉你我们在考虑理想类群的 p-Sylow 子群，我们干脆就简单地记它为 $A_{K_n} \leq C_{K_n}$，于是题意为 $h_n = |A_{K_n}|$。考虑 $K = \mathbb{Q}$，并使用我们构造的分圆 $\mathbb{Z}_{37}$ 扩域，则有 $K_n = \mathbb{Q}(\zeta_{37^n})$，此时比较简单地有

$$A_{K_n} \cong \mathbb{Z}/37^n\mathbb{Z}, h_n = 37^n, e_n = n, \lambda = 1, \mu = 0, \nu = 0, n_0 = 1$$

如果使用分圆 $\mathbb{Z}_{691}$ 扩域，则有 $K_n = \mathbb{Q}(\zeta_{691^n})$，并且可得

$$A_{K_n} \cong \mathbb{Z}/691^n\mathbb{Z} \oplus \mathbb{Z}/691^n\mathbb{Z}, h_n = 691^n 691^n, e_n = 2n, \lambda = 2, \mu = 0, v = 0, n_0 = 1$$

稍微复杂点，我们令 $K = \mathbb{Q}(\sqrt{-31})$ 考虑分圆 $\mathbb{Z}_3$ 扩域，则有 $K_n = K(\zeta_{3^n})$，并且可得

$$A_{K_n} \cong \mathbb{Z}/3^n\mathbb{Z}, h_n = 3^n, e_n = n, \lambda = 1, \mu = 0, v = 0, n_0 = 1$$

我们令 $K = \mathbb{Q}(\sqrt{-1399})$ 考虑分圆 $\mathbb{Z}_3$ 扩域，则有 $K_n = K(\zeta_{3^n})$，并且可得

$$A_{K_n} \cong \mathbb{Z}/3^{n+2}\mathbb{Z} \oplus \mathbb{Z}/3^{n-1}\mathbb{Z}, e_n = 2n+1 (n \geq 2), \lambda = 2, \mu = 0, v = 1, n_0 = 2$$

我们令 $K = \mathbb{Q}(\sqrt{-762})$ 考虑分圆 $\mathbb{Z}_3$ 扩域，则有 $K_n = K(\zeta_{3^n})$，并且可得

$$A_{K_n} \cong \mathbb{Z}/3^n\mathbb{Z} \oplus \mathbb{Z}/3^5\mathbb{Z} (n \geq 5), e_n = n+5, \lambda = 1, \mu = 0, v = 5, n_0 = 5$$

有了上面这么多的例子，理解这个定理应该是小菜一碟的事了。至于怎么构造一个非分圆 $\mathbb{Z}_p$ 扩域，典型例子就是逆分圆 $\mathbb{Z}_p$ 扩域 (anticyclotomic $\mathbb{Z}_p$-extension)，它只能定义在虚二次域上，在此之前我们先来得到所有 $\mathbb{Z}_p$ 扩域的个数，回忆单位群 $U_K = O_K^\times$，$r = \mathrm{rank}_{\mathbb{Z}}(U_K)$ 表示其自由子群的秩，$\mathfrak{p}$ 为 O_K 的一个素理想则相应对应着 K 的一个素点，其完备化得到的局部域记为 $K_\mathfrak{p}$，此时我们定义一个对角映射

$$\Delta : U_K \to \prod_{\mathfrak{p}|(p)} U_{K_\mathfrak{p}}, u \mapsto (u, u, ..., u)$$

主理想 $(p) \subset \mathbb{Z}$ 在 K 上的素理想分解个数有限，所以映射右边笛卡尔积个数也是有限的，所谓的 **Leopoldt 误差** (Leopoldt defect) 指 $\delta_K = \mathrm{rank}_{\mathbb{Z}}(\ker \Delta)$，而 **Leopoldt 猜想**为：对任意数域 K 有 $\delta_K = 0$，其等价于上述对角映射为单射。设 $S_\infty(K)$ 表示数域 K 的所有阿基米德素点，则我们可以得到 Leopoldt 误差的一个估计式为

$$0 \leq \delta_K \leq \frac{|S_\infty(K)| - 1}{2}$$

设 $\tilde{K}$ 表示数域 K 所有 $\mathbb{Z}_p$ 扩域的复合[2]，则可以证明

$$Gal(\tilde{K}/K) \cong \mathbb{Z}_p^{r_2+1+\delta_K}$$

这里的指数 $r_2+1+\delta_K$ 就是 $\mathbb{Z}_p$ 扩域的个数。考虑虚二次域 $K = \mathbb{Q}(\sqrt{-d})$，我们有 $n = 2, r_1 = 0, r_2 = 1, |S_\infty(K)| = r_1 + r_2 = 1$，由估计式可得 $\delta_K \leq \frac{1-1}{2} = 0$，即满足 Leopoldt 猜想，进一步计算有 $r_2 + 1 + \delta_K = 2$，即

$$Gal(\tilde{K}/K) \cong \mathbb{Z}_p^2$$

换言之，K 有两个 $\mathbb{Z}_p$ 扩域，一个是我们的分圆 $\mathbb{Z}_p$ 扩域 $K^{cyc} = K_\infty$，我们把另一个记为 K^{acti} 并称为**逆分圆 $\mathbb{Z}_p$ 扩域**。这个性质 $Gal(K_n^{cyc}/K) \cong Gal(K_n^{acti}/K) \cong \mathbb{Z}/p^n\mathbb{Z}$ 是定义给出的显然结果，它们的主要区别在于，$Gal(K^{cyc}/\mathbb{Q})$ 为交换群，而 $Gal(K^{acti}/\mathbb{Q}) \cong \mathbb{Z}_p \rtimes \mathbb{Z}/2\mathbb{Z} \subset \mathbb{D}_\infty$，这里使用的是半直和，其结果不一定是交换群而是无限二面体群，设 $\mathbb{D}_{2n} = \{a^s b^t \mid a^n = 1, b^2 = 1, bab = a^{-1}\}$ 表示二面体群，则有 $\mathbb{D}_\infty = \cup \mathbb{D}_{2n}$。通常我们把 λ, μ, v 称为 $\mathbb{Z}_p$ 扩张 K_∞/K 的**岩泽不变量** (Iwasawa invariant)，至于这里的 n_0 只是足够大 n 的一个下界，学者并不怎么关心它，在这篇论文 [17] 中，我们可以找到更多非分圆 $\mathbb{Z}_p$ 扩域的计算实例，其借助了逆分圆 $\mathbb{Z}_p$ 扩域的工具，有兴趣的读者可以自行探索。

[2]即把所有 $\mathbb{Z}_p$ 扩域使用一般乘法乘起来，由于它们都是 K^{ab}(阿贝尔闭包) 的子域，所以运算是可以互相兼容的

岩泽代数

岩泽公式显然不可能是终点，而是起点，我们很自然地会去思考如何去计算特定数域的岩泽不变量 λ,μ,ν。在数论里最爱做的事就是对应了，会不会还有一个什么类似的对象也有着同样的不变量呢？确实有，我们设 R 为完备赋值环[3]，其唯一极大理想的生成元为 $\pi\in(\pi)\subset R$，$R[[X]]=\{\sum_{i=0}^{\infty}a_iX^i\mid a_i\in R\}$ 表示形式幂级数环，此时我们有定理。

定理 4.10 (Weierstrass preparation theorem)

任意非零形式幂级数 $0\neq f(X)\in R[[X]]$，可以唯一地写为下面的形式

$$f(X)=\pi^{\mu}(X^{\lambda}+a_1X^{\lambda-1}+\ldots+a_{\lambda})u(X)$$

其中 $\mu,\lambda\in\mathbb{Z}_{\geq 0},a_1,\ldots,a_{\lambda}\in(\pi)=\pi R,u(X)\in R[[X]]^{\times}$

♡

定理的证明到处都有没啥好说的，关键在于其给出了非零形式幂级数 f 的两个不变量 $\mu(f),\lambda(f)$。讨论另一个例子前，我们插播一个抽象代数中模的定理。

定理 4.11 (有限生成模的结构定理)

(1) 设 R 是 PID(主理想整环)，M 是有限生成 R-模，$\mathrm{Tor}M=\{x\in M\mid\exists a\neq 0,ax=0\}$ 为 M 的挠子模，则存在自由子模 N 使得有唯一分解

$$M=\mathrm{Tor}M\oplus N$$

(2) 设 M 为有限生成挠模 (即 $\mathrm{Tor}M=M$)，设其零化子 $\mathrm{Ann}M=\{a\in R\mid\forall x\in M,ax=0\}$ 有 $\mathrm{Ann}M=(a),a\in R$，生成元的唯一分解为 $a=up_1^{n_1}\ldots p_r^{n_r}$，$p_i$ 为互不相伴的素元，$u\in R^{\times}$ 为可逆元，则有唯一分解

$$M=M/(p_1^{n_1})\oplus\ldots\oplus M/(p_r^{n_r})$$

其中 $M/(p_i^{n_i})=\{x\in M\mid p_i^{n_i}x=0\}$ 是 M 的子模。

♡

上述定理的一个典型应用就是将有限生成交换群看成一个 $\mathbb{Z}$-模，从而得到有限生成交换群的基本定理。同样地考虑交换环 $A=R[[X]]$ 上的有限生成挠模 M，但由于 R 并不是主理想整环，所以我们要弱化模同构的条件，对于模同构 $f:M\to N$ 有：$|\ker f|=1\Leftrightarrow f$为单射、$|\mathrm{coker}f|=1\Leftrightarrow f$为满射。于是对于模同态 $f:M\to N$，如果满足 $|\ker f|,|\mathrm{coker}f|$ 均有限，就把 f 称为**伪同构** (pseudo-isomorphism)，在这种意义下，对任意的有限生成挠 A-模 M 都有唯一伪同构分解

$$M\sim A/(f_1^{n_1})\oplus\ldots\oplus A/(f_r^{n_r})$$

我们把 $f=f_1^{n_1}\ldots f_r^{n_r}$ 生成的 A 的理想 $(f)=fA$ 称为 M 的**特征理想**，并记为 $\mathrm{Char}(M)$，再进一步定义一些不变量为 $\lambda(M)=\lambda(f),\mu(M)=\mu(f)$。接着我们来把这些内容与上一节联系起来，考虑 $\mathbb{Z}_p$ 扩张 K_{∞}/K，我们记 $\Gamma=Gal(K_{\infty}/K)\cong\mathbb{Z}_p,\Lambda=\mathbb{Z}_p[[\Gamma]]$，$X=\varprojlim_n A_{K_n}$，我们先要说明 X 是一个 Λ-模。去除逆向极限，即相当于说明 A_{K_n} 是一个 Λ_n-模 ($\Lambda_n=\mathbb{Z}_p[[Gal(K_n/K)]]$)，设 K_n 的极大非分歧阿贝尔 p 扩域为 L_n，并记 $L_{\infty}=\cup_nL_n$，由整体域的互反律可得同构，并可以进一步得到作用

$$A_{K_n}\cong Gal(L_n/K_n),Gal(K_n/K)\times Gal(L_n/K_n)\to Gal(L_n/K_n),(\sigma,s)\mapsto\tilde{\sigma}s\tilde{\sigma}^{-1}$$

其中 $\tilde{\sigma}$ 是 σ 在 $Gal(L_n/K_n)$ 上的延拓 (即满足 $\tilde{\sigma}|_K=\sigma$)，因为 $Gal(L_n/K_n)$ 是交换群，故 $\tilde{\sigma}$ 可以任意选取。验证没什么好说的，总结可得下面的核心定理。

[3]即存在局部域 K 使得 $R=O_K$，典型例子为 p 进整数环 $\mathbb{Z}_p=O_{\mathbb{Q}_p}$

定理 4.12

对任意数域 K，记 $\Lambda = \mathbb{Z}_p[[Gal(K_\infty/K)]]$，则 $X = \varprojlim_n A_{K_n}$ 是一个有限生成挠 Λ-模。 ♡

读者需要知道的是，这部分的内容实际上是为了证明前一部分的岩泽公式而准备的，正是因为这个定理，我们可以将 A_{K_n} 的阶数计算转移到一个有特殊构造的有限生成挠模上 (模的不变量就是岩泽公式中的不变量)，而且在这部分更核心的内容中，我们可以看到岩泽不变量 λ, μ, ν 实际是来自一个无穷形式幂级数的，那么由此来迈向分析那一边就是一个自然的想法了。

岩泽主猜想

黎曼函数 $\zeta(s)$，想必基本所有学数学的人都很熟悉，它的局部表达式为 $\mathrm{Re}(z) > 1, \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$，$z = 1$ 是它的唯一极点，由复变解析函数的唯一延拓定理来保证全纯点的取值。更进一步，对于一个狄利克雷特征 $\chi : \mathbb{N} - \{0\} \to (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$，我们还能定义**狄利克雷 L 函数**为

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}, \mathrm{Re}(z) > 1$$

对于平凡特征 $\chi = \mathbf{1} : x \mapsto x$ 有 $L(s, \mathbf{1}) = \zeta(s)$，学数论的人应该也是十分熟悉的。接着我们需要构造与 L 函数类似的在 p 进域中的函数，直接用狄利克雷 L 函数像下面这样定义

$$L_p(1 - r, \chi) = (1 - \chi\omega^{-r}(p)p^{r-1})L(1 - r, \chi\omega^{-r})$$

并不是明智的选择，我们知道伯努利数可以拿来计算整点处的函数值，所以我们决定采用伯努利数来定义会更本质一些，**广义伯努利数**由下面函数生成

$$\sum_{j=1}^{N} \frac{\chi(j)te^{jt}}{e^{Nt} - 1} = \sum_{m=0}^{\infty} B_{m,\chi} \frac{t^m}{m!}$$

选取平凡特征时退化为常规的伯努利数，更细致的探究就不讲了，可以参考这本书 [20] 或这篇文章 [4]，根据我们的核心书籍 [33] 可以得到对任意整数 $m \geq 1$ 有

$$L(1 - m, \chi) = -\frac{B_{m,\chi}}{m}$$

我们的目的是将 L 函数从复数域 $\mathbb{C}$ 扩展到 p 进数域，由于分析完备的要求，有两种选择 $\mathbb{Q}_p \to \mathbb{Q}_p$ 或 $\mathbb{C}_p \to \mathbb{C}_p$，我们倾向于选择后者，实际上就和我们从实数域 $\mathbb{R}$ 和复数域 $\mathbb{C}$ 中选择 $\mathbb{C}$ 一样，分析和代数的双重完备才有将代数与分析联系起来的可能。只不过它也是存在性的，并且在可数个点上满足我们的需求，而没有 $\mathrm{Re}(z) > 1$ 这么大的范围。

定理 4.13 (Kubota,Leopoldtp)

设狄利克雷特征为 $\chi : (\mathbb{Z}/N\mathbb{Z}) \to \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}_p}^\times$，Teichmüller 特征为 $\omega(x)^p = \omega(x), x \in \mathbb{Z}$ 在 $\mathbb{Z}_p^\times$ 中的唯一解

$$\omega(x) = \lim_{n \to \infty} x^{p^n}$$

(1) 对于非平凡的 $\chi \neq \mathbf{1}$，存在唯一的定义在 $\{s \in \mathbb{C}_p \mid |s|_p < p^{-\frac{p}{p-1}}\}$ 上的 p 进亚纯函数 $L_p(s, \chi)$ 满足对任意正整数 $n > 0$ 有

$$L_p(1 - n, \chi) = -(1 - \chi\omega^{-n}(p)p^{n-1})\frac{B_{n,\chi\omega^{-n}}}{n}$$

(2) 对于平凡的 $\chi = \mathbf{1}$，存在唯一除极点 $s = 1$ 外处处解析的 p 进亚纯函数 $L_p(s, \mathbf{1})$ 满足和上面一样的等式，

且 $s=1$ 是留数为 $1-\frac{1}{p}$ 的一阶极点。

不要因为这里的狄利克雷特征换了个值域就不认得了，值得注意的是特征 $\chi\omega^{-n}$ 通常在互素情况 $(a,p)=1$ 下有 $\chi\omega^{-n}(a)=\chi(a)\omega^{-n}(a)$，一般情形下则并不一定如此，$L_p(s,\chi)$ 也被称为 **Kubota-Leopoldtp 进 L 函数**。既然解析的函数已经到手了，那么下一步计划自然就是往代数那边靠拢了，我们有下面的定理。

定理 4.14 (Iwasawa)

设狄利克雷特征 $\chi:(\mathbb{Z}/N\mathbb{Z})\to\overline{\mathbb{Q}_p}^{\times}, n\neq 1, p^n(n\geq 2)$、完备赋值环 $O_\chi=\mathbb{Z}_p[\chi]$ 和 $1+2p\mathbb{Z}_p$ 的拓扑生成元 u，则存在函数 $\frac{1}{2}g_\chi(T)\in O_\chi[[T]]$ 满足

$$L_p(s,\chi)=g_\chi(u^s-1)$$

如果取平凡特征 $\chi=\mathbf{1}$ 就有 $O_{\mathbf{1}}=\mathbb{Z}_p[1]=\mathbb{Z}_p$，因此 $O_\chi[[T]]$ 实际就是我们之前讨论的 $\Lambda=\mathbb{Z}_p[[T]]$ 的推广，因此我们需要去扩展代数那边 Λ 和 X 的内涵，即有

$$\Lambda_\chi=O_\chi[[Gal(K_\infty/K)]], X_\chi=X\otimes_{\Lambda_{N_0}}\Lambda_\chi$$

其中 $\Lambda_{N_0}=\mathbb{Z}_p[[Gal(K_\infty/\mathbb{Q})]], N=N_0p^a, (N_0,p)=1$，并且可得 X_χ **是一个有限生成挠 Λ_χ-模**，其特征理想满足 $\mathrm{Char}(X_\chi)\subset\Lambda_\chi$。如果让特征 χ 加上一些限制使得 $\chi^{-1}\omega$ 满足上一个定理的条件，就能得到 $\frac{1}{2}g_{\chi^{-1}\omega}(u^s-1)\in\Lambda_\chi$，于是我们的核心猜想就呼之欲出了。

定理 4.15 (The Main Conjecture of Iwasawa Theory)

对任意的数域 K、分圆 $\mathbb{Z}_p$ 扩域 K_∞/K 和特征 $\chi(-1)=-1, \chi\neq\omega, \mathbb{Q}^\chi\cap\mathbb{Q}_\infty=\mathbb{Q}$ 有

$$\mathrm{Char}(X_\chi)=(\frac{1}{2}g_{\chi^{-1}\omega}(T))$$

这是一个主理想等式，也是岩泽主猜想的原始形式，实际在证明的时候主要是先找到左右两边的生成元 $f_\chi(T)$ 和 $g_\chi(T)$，再主要分析 $f_\chi(T)=g_\chi(T)$，这也是大多数书籍对主猜想的表述，但于我个人而言更喜欢原始的表达，主要是那种十分迷人的不对称感，和让人激动不已的 $\frac{1}{2}$，不过主要还是因为生成元的严格构造链有点小复杂，不太想在这里用太多笔墨来介绍，写的人烦看的人也烦，简简单单才是最好的理解，复杂的东西就留给证明吧，至于岩泽主猜想的另一种更迷惑表达 $\mathrm{Char}(X_\chi)=\mathrm{Char}(E_\infty/C_\infty)_\chi$ 就不拿出来献丑了。回想我们最开始的目的，是估计 A_K 的阶数，通过上面的主猜想，在一些特殊情况下，我们确实可以得到相应的数值。

定理 4.16 (阿贝尔复扩域的阶数估计)

设 F 是 $\mathbb{Q}$ 的阿贝尔复扩域，p 为奇素数 $(p,n)=1, n=[F,\mathbb{Q}]$，$\chi$ 为符合条件的狄利克雷特征，$g=[O_\chi:\mathbb{Z}_p]$，$A_F^\chi=A_F\otimes_{\mathbb{Z}_p[Gal(F/\mathbb{Q})]}O_\chi$，则有

$$|A_F^\chi|=|O_\chi/(L_p(0,\chi^{-1}\omega))|$$

$$v_p(|A_F^\chi|)=gv_p(L_p(0,\chi^{-1}))$$

这下懂了什么叫代数与分析的联系吗？实际上，在岩泽理论之前就有过复分析函数与数论对象阶数估计的关系研究定理，没错，就是类数公式，不如说岩泽理论就是在此基础上发展出来的，我们接下来就讨论这个。

$$\prod_{\chi\in X,\chi\neq\mathbf{1}}L(1,\chi)=\frac{2^{r_1}(2\pi)^{r_2}R_Kh_K}{\omega_K\sqrt{|d_K|}}, \zeta_K(s)=\prod_{\chi\in X}L(1,\chi)$$

$$\prod_{\chi\in X,\chi\neq\mathbf{1}}(1-\frac{\chi(p)}{p})^{-1}L_p(1,\chi)=\frac{2^{n-1}R_p(K)h_K}{\sqrt{\Delta_K}}$$

第五章 解析数论

解析数论就是用分析学的方法来研究数论，这就是一句废话，但事实就是如此，而我们使用的分析学函数，无非就是 L 函数家族。对于解析数论我们很难有一套成型而且系统的理论，其基本都是都是对 L 函数家族家族的研究，并分布在各个定理的证明中，学习解析数论的最好办法就是以各种著名定理为导向，当然前提是分析水平不能落下。我们将以解析数论的几个著名问题来构成本章，大部分内容来自这本书 [43]。

5.1 类数公式

狄利克雷 (Dirichlet) 级数通论

我们先来承接上一部分内容，讨论类数公式的证明，L 函数家族或者狄利克雷级数是整个解析数论的核心，所以我们有必要来熟悉一下它们。什么是**数论函数**、**积性函数**和**完全积性函数**? 就没必要再重复讨论了。但我们需要把数论函数的概念进一步扩展为 $f:\mathbb{P}\to R$，其中 $\mathbb{P}=\mathbb{N}-\{0\}$ 是正整数，R 是交换幺环，我们把这样的数论函数 f 对应一个形式表达式

$$L(s,f)=\sum_{n=1}^{\infty}\frac{f(n)}{n^s}$$

并称为 f 的**狄利克雷级数** (简称为 **D-级数**)，反过来也是一样的，记 $D(R)=\{\sum_{n=1}^{\infty}\frac{a_n}{n^s}\mid a_n\in R\}$ 为交换幺环 R 上的所有形式 D-级数，则它与所有数论函数是一一对应的。我们知道所有数论函数 $R^{\mathbb{P}}=\{f:\mathbb{P}\to R\}$ 关于普通加法 $+$ 和狄利克雷卷积 $*$ 形成一个整环，故我们也在 $D(R)$ 上构造环结构

$$\sum_{n=1}^{\infty}\frac{a_n}{n^s}+\sum_{n=1}^{\infty}\frac{b_n}{n^s}=\sum_{n=1}^{\infty}\frac{a_n+b_n}{n^s}$$

$$\sum_{n=1}^{\infty}\frac{a_n}{n^s}\sum_{n=1}^{\infty}\frac{b_n}{n^s}=\sum_{r=1}^{\infty}\sum_{k=1}^{\infty}\frac{a_rb_k}{r^sk^s}=\sum_{n=1}^{\infty}\frac{\sum_{rk=n}a_rb_k}{n^s}=\sum_{n=1}^{\infty}\frac{\sum_{d|n}a_db_{\frac{n}{d}}}{n^s}$$

显然可以得到整环同构 $R^{\mathbb{P}}\cong D(R)$，我们把 $D(R)$ 称为形式 **D-级数环**。一个十分简单的例子是取整数环 $R=\mathbb{Z}$ 和恒值函数 $\mathbf{1}(n)=1$，则可得到黎曼 zeta 函数 $\zeta(s)$，取复数环 $R=\mathbb{C}$ 和狄利克雷特征 χ 就得到了狄利克雷 L 函数 $L(s,\chi)$，它们的倒函数或在环中的逆元为

$$L^{-1}(s,\chi)=\sum_{n=1}^{\infty}\frac{\mu(n)\chi(n)}{n^s}$$

这里的 $\mu(n)$ 是老朋友莫比乌斯函数。对于 D-级数的另一个重要性质是乘积分解，即下面定理。

定理 5.1 (无限和与无限积的转化)

设 $L(s,f)$ 是 $f(n)$ 的形式 D-级数，则相应的欧拉 (Euler) 乘积有下述性质

(1)f 为积性函数当且仅当，$L(s,f)$ 满足

$$L(s,f)=\prod_{p\ is\ prime}(1+f(p)p^{-s}+f(p^2)p^{-2s}+\ldots+f(p^m)p^{-ms}+\ldots)=\prod_{p\ is\ prime}(\sum_{m=0}^{\infty}f(p^m)p^{-ms})$$

(2)f 为完全积性函数当且仅当，$L(s,f)$ 满足

$$L(s,f)=\prod_{p\ is\ prime}(1-f(p)p^{-s})^{-1}$$

♡

黎曼 zeta 函数的重要性体现在不少数论函数的 D-级数都能由它生成，下面是一些简单的对应例子，我们均取 $\mathbb{Z}$ 作为基环。

$$\mathbf{1}\leftrightarrow\zeta(s),\mu\leftrightarrow\zeta(s)^{-1},\sigma_k\leftrightarrow\zeta(s)\zeta(s-k),\varphi\leftrightarrow\frac{\zeta(s-1)}{\zeta(s)}$$

但说到底我们现在依旧只是在讨论形式级数，没有分析就谈不上解析数论了，所有我们先限定一个足够大的可以分析的环 $\mathbb{C}$，这样 $\zeta(s)$ 和 $L(s,\chi)$ 这两个著名的例子都在其中，而且 $L(s,f)$ 自己也变成了一个复级数。剩下的内容都是复分析的结果，我们一笔带过，能理解其意即可。

定理 5.2 (收敛区域存在定理)

对任意级数 $F(s)=\sum_{n=1}^{\infty}\frac{a_n}{n^s},a_n\in\mathbb{C}$，记局部和为 $A(N)=\sum_{n=1}^{N}a_n$，则

(1) 存在扩充实数 $-\infty\le\sigma_0\le+\infty$ 使得，当 $\mathrm{Re}(s)>\sigma_0$ 时，级数 $F(s)$ 收敛，当 $\mathrm{Re}(s)<\sigma_0$ 时，级数 $F(s)$ 发散。并且 $F(s)$ 在 $\mathrm{Re}(s)>\sigma_0$ 的任意紧子集内一致收敛，在 $\mathrm{Re}(s)>\sigma_0$ 内收敛的函数是解析函数。

(2) 若局部和数列 $A(N)$ 发散，则

$$\sigma_0=\inf_a\{A(N)=O(N^a)\}=\overline{\lim_{N\to\infty}}\log_N|A(N)|$$

♡

定理 5.3 (极点位置定理)

若 $F(s)=\sum_{n=1}^{\infty}\frac{a_n}{n^s},a_n\ge0$ 在 $\mathrm{Re}(s)>c$ 内收敛且存在 $\varepsilon>0$ 使得 $F(s)$ 在 $B(c,\varepsilon)$ 内解析[a]，则存在 $0<\varepsilon_0<\varepsilon$ 使得 $F(s)$ 在 $\mathrm{Re}(s)>c-\varepsilon_0$ 内收敛。换言之，$s=\sigma_0$ 是 $F(s)$ 的奇点。

[a]此处主要由复变函数的唯一解析延拓做担保，与之前的收敛没什么关系

♡

定理 5.4 (欧拉乘积存在定理)

(1) 我们把 σ_0 称为 $\sum_{n=1}^{\infty}\frac{a_n}{n^s}$ 的收敛横坐标。并把 $\sum_{n=1}^{\infty}\frac{|a_n|}{n^s}$ 的收敛横坐标 σ_1 称为 $\sum_{n=1}^{\infty}\frac{a_n}{n^s}$ 的绝对收敛横坐标。它们的基本关系是

$$\sigma_0\le\sigma_1$$

(2) 设 $f(n)$ 是积性函数，$L(s,f)$ 的绝对收敛横坐标为 σ_1，则当 $\mathrm{Re}(s)>\sigma_1$ 时欧拉乘积

$$\prod_p(1+f(p)p^{-s}+f(p^2)p^{-2s}+\ldots+f(p^m)p^{-ms}+\ldots)$$

绝对收敛，并且收敛到 $L(s,f)$。

♡

关于 $\zeta(s)$ 和 $L(s,\chi)$ 的性质有必要复述一下吗？由于 $\zeta(s)=L(s,\mathbf{1})$，所以有些通性可以只看后者，由于 $|\chi|=1$ 所以 $|A(N)|=O(N),\sigma_0=1$，即 $L(s,\chi)$ 在 $\mathrm{Re}(s)>1$ 内收敛。设 G 是有限交换群，我们把群同态 $\chi:G\to\mathbb{C}^\times$ 称为 G 的**特征**，取加法群时 $\lambda:\mathbb{Z}/m\mathbb{Z}\to\mathbb{C}^\times$ 称为**模 m 的加法特征**，它的形式是唯一且单调 $\lambda_k(n)=e^{\frac{2\pi ikn}{m}},0\le k\le m-1$，平凡情形为 $\mathbf{1}=\lambda_0$。取乘法群时 $\chi:(\mathbb{Z}/m\mathbb{Z})^\times\to\mathbb{C}^\times$ 称为**模 m 的 D 特征**，对 D 特征 χ 做下面的嵌入

$$\chi_m:\mathbb{P}\to\begin{cases}(\mathbb{Z}/m\mathbb{Z})^\times\to\mathbb{C}^\times & a\mapsto(a \bmod m)\in(\mathbb{Z}/m\mathbb{Z})^\times\\ 0 & \text{other}\end{cases}$$

则 χ_m 就是我们之前所说的模 m 的狄利克雷特征。若不存在 $d\mid m$ 和模 d 的 D 特征 χ'，使得 $(m,a)=1$ 时 $\chi'(a)=\chi(a)$，就把 χ 称为**本原 D-特征**。对于非本原模 m 的 D-特征，可以不断地找到更小的模 d 特征，我们把最小的 d 称为 χ 的**导子**，也就是我们之前所说的狄利克雷特征的周期。方便起见，以后我们说的狄利克雷特征 $\chi:(\mathbb{Z}/N\mathbb{Z})^\times\to\mathbb{C}^\times$ 均指本原 D-特征，其中的 N 就是导子或周期。接下来我们引入**模 m 的高斯和**为

$$G(k,\chi)=\sum_{n=0}^{m-1}\chi(n)\lambda_k(n)=\sum_{n=0}^{m-1}\chi(n)e^{\frac{2\pi ikn}{m}}$$

当 $(k,m)=1$ 互素时，我们可以抽出其中的加法特征得到 $G(k,\chi)=\overline{\chi}(k)G(1,\chi)$。接着我们引入奇偶特征指标和 L 函数对应的 xi 函数

$$\delta(\chi)=\begin{cases}0 & \chi(-1)=1\\ 1 & \chi(-1)=-1\end{cases}$$

$$\xi(s,\chi)=(\frac{N}{\pi})^{\frac{s}{2}}\Gamma(\frac{s+\delta}{2})L(s,\chi)$$

这里的 $\Gamma(z)=\int_0^{+\infty}t^{z-1}e^{-t}dt,\mathrm{Re}(z)>0$ 是在复平面上解析延拓后的伽马函数。于是可以得到 L 函数最重要的函数方程为

$$\xi(s,\chi)=\frac{G(1,\chi)}{i^{\delta(\chi)}\sqrt{N}}\xi(1-s,\overline{\chi})$$

当 $N\geq 2$ 即 $L(s,\chi)\neq\zeta(s)$ 时，我们发现 $\sigma_0=0,\sigma_1=1$，借助延拓和补充定义 $L(s,\chi)$ 是整个复平面上的解析函数且没有奇点。其中 $s=-\delta(\chi)-2n(n=0,1,2,...)$ 是 $L(s,\chi)$ 的零点，称为它的**平凡零点**，而 $L(s,\chi)$ 的所有非平凡零点均位于 $0<\mathrm{Re}(s)<1$ 内，所谓的**黎曼猜想**为：$L(s,\chi)$ 的所有非平凡零点均在直线 $\mathrm{Re}(s)=\frac{1}{2}$ 上。

戴德金 zeta 函数

在上面狄利克雷级数的基础上，我们更进一步，对于数域 K，其代数整数环 O_K 的所有非零理想构成的集合为 I_K^o，我们把

$$\zeta_K(s)=\sum_{I\in I_K^o}\frac{1}{N(I)^s}$$

称为**戴德金 zeta 函数**。如果取 $K=\mathbb{Q}$，则 $I_K^o=\{(n)\mid n=1,2,...\}$，$N((n))=N_{\mathbb{Q}/\mathbb{Q}}((n))=|\mathbb{Z}/(n)|=n$，即退化为黎曼 zeta 函数 $\zeta(s)=\zeta_{\mathbb{Q}}(s)$。虽然 $\zeta_K(s)$ 的定义看起来有点奇怪，但从范 $N(I)$ 只能取整值上来看 $\zeta_K(s)$ 依旧是一个复级数，那么我们自然就应该讨论其收敛性和复解析性了。

定理 5.5

(1)$\zeta_K(s)$ 在 $\mathrm{Re}(s)>1$ 中收敛且在 $\mathrm{Re}(s)>1$ 的每个紧子集中一致绝对收敛。

(2) 当 $\mathrm{Re}(s)>1$ 时，欧拉乘积 $\prod_{\mathfrak{p}}(1-N(\mathfrak{p})^{-s})^{-1}$($\mathfrak{p}$ 过 O_K 的所有素理想) 收敛，且等于 $\zeta_K(s)$。

♡

证明 (1) 我们只需借助级数 $\sum_{n=1}^{\infty}\frac{1}{n^s}$ 在实数 $s>1$ 处进行比较审敛即可 ($|\frac{1}{N(I)^s}|=\frac{1}{N(I)^{\mathrm{Re}(s)}}$)。设定一个上限 x，我

们有

$$\sum_{N(\mathfrak{p})\le x}\frac{1}{N(\mathfrak{p})^s}\le\sum_{p\le x}\sum_{\mathfrak{p}|(p)}\frac{1}{N(\mathfrak{p})^s}\le n\sum_{p\le x}\frac{1}{p^s}\le n\sum_{m\le x}\frac{1}{m^s}$$

如果我们使用 $\mathfrak{p}$ 和 p 则它们分别代表素理想和素数，而其它则是一般的理想和整数。注意到 $\sum x_n$ 收敛等价于 $\prod(1+x_n)$ 收敛，我们有

$$\prod_{N(\mathfrak{p})\le x}\frac{1}{1-N(\mathfrak{p})^{-s}}=\prod_{N(\mathfrak{p})\le x}(1+\frac{1}{N(\mathfrak{p})^s-1}),\sum_{N(\mathfrak{p})\le x}\frac{1}{N(\mathfrak{p})^s-1}\le 2\sum_{N(\mathfrak{p})\le x}\frac{1}{N(\mathfrak{p})^s}$$

最后，我们借助局部的欧拉积就是所求收敛级数

$$\sum_{N(\mathfrak{a})\le x}\frac{1}{N(\mathfrak{a})^s}\le\prod_{N(\mathfrak{p})\le x}(1+N(\mathfrak{a})^{-s}+N(\mathfrak{a})^{-2s}+\ldots)=\prod_{N(\mathfrak{p})\le x}\frac{1}{1-N(\mathfrak{p})^{-s}}$$

(2) 在 (1) 中已经证明了欧拉积 $\prod_{\mathfrak{p}}\frac{1}{1-N(\mathfrak{p})^{-s}}$ 的收敛性。注意到

$$|\prod_{N(\mathfrak{p})\le x}\frac{1}{1-N(\mathfrak{p})^{-s}}-\sum_{N(\mathfrak{a})\le x}\frac{1}{N(\mathfrak{a})^s}|=|\sum_{N(\mathfrak{a})>x}\frac{1}{N(\mathfrak{a})^s}|\le\sum_{N(\mathfrak{a})\ge x}\frac{1}{N(\mathfrak{a})^{\mathrm{Re}(s)}}$$

由 $\mathrm{Re}(s)>1$ 可得右边是无限小的，从而 $\prod_{\mathfrak{p}}(1-N(\mathfrak{p})^{-s})^{-1}=\zeta_K(s)$。

上面都是些比较基础的东西，我们来考虑更复杂点的函数，令 $C\in C_K$ 是数域 K 的任意一个理想类，于是我们可以定义函数

$$\zeta_K(C,s)=\sum_{\mathfrak{a}\in C}\frac{1}{N(\mathfrak{a})^s}$$

如果我们令 $a_n(C)$ 表示满足 $N(\mathfrak{a})=n,\mathfrak{a}\in C$ 的整理想个数，则有

$$\zeta_K(C,s)=\sum_{n=1}^{\infty}\frac{a_n(C)}{n^s},\zeta_K(s)=\sum_{C\in C_K}\zeta_K(C,s)$$

也就是说上面的函数实际是戴德金 zeta 函数的一部分，而这种成分的个数就是我们的类数 h_K，第一步是得到一个计数函数的估计。

定理 5.6

设 $C\in C_K$ 是代数数域 K 的一个理想类，$n=[K,\mathbb{Q}]$，定义函数

$$f(C,x)=\sum_{\mathfrak{a}\in C,N(\mathfrak{a})\le x}1=\sum_{n\le x}a_n(C)$$

这里的 $\mathfrak{a}$ 只过 C 中的整理想，则有

$$\lim_{x\to\infty}f(C,x)=\rho_K x+O(x^{1-\frac{1}{n}})$$

$$\rho_K=\frac{2^{r_1}(2\pi)^{r_2}R_K}{\omega_K\sqrt{|d_K|}}$$

其中 r_1 和 r_2 分别是 K 到 $\mathbb{C}$ 中的实嵌入个数和复嵌入个数，d_K 是 K 的判别式，ω_K 是单位根群 W_K 的阶数，R_K 是 K 的正则化子 (regulator)，定义为 (等于 r 阶主子式)

$$R_K=\begin{vmatrix}\ln|\sigma_1(u_1)| & \cdots & \ln|\sigma_{r_1}(u_1)| & 2\ln|\sigma_{r_1+1}(u_1)| & \cdots & 2\ln|\sigma_{r_1+r_2}(u_1)|\\ \ln|\sigma_1(u_2)| & \cdots & \ln|\sigma_{r_1}(u_2)| & 2\ln|\sigma_{r_1+1}(u_2)| & \cdots & 2\ln|\sigma_{r_1+r_2}(u_2)|\\ \vdots & & \vdots & \vdots & & \vdots\\ \ln|\sigma_1(u_r)| & \cdots & \ln|\sigma_{r_1}(u_r)| & 2\ln|\sigma_{r_1+1}(u_r)| & \cdots & 2\ln|\sigma_{r_1+r_2}(u_r)|\\ (r+1)^{-1} & \cdots & (r+1)^{-1} & (r+1)^{-1} & \cdots & (r+1)^{-1}\end{vmatrix}$$

其中 $\{u_1,...,u_r\}, r=r_1+r_2-1$ 为基本单位组，$\sigma_1,...,\sigma_{r_1}$ 为实嵌入，$\sigma_{r_1+1},...,\sigma_{r_1+r_2}$ 为复嵌入。 ♡

证明 (1) 首先我们要将 $f(C,x)$ 的表达式进行转化。其表示范不超过 x 的 C 中整理想的个数，我们任取逆元 C^{-1} 中的一个整理想 $\mathfrak{B}$，则 $\mathfrak{aB}=(a), a\in\mathfrak{B}, N(a)=N(\mathfrak{a})N(\mathfrak{B})\leq xN(\mathfrak{B})$，换言之我们有

$$f(C,x)=|\{(a)=aO_K \mid a\in\mathfrak{B}, N(a)\leq xN(\mathfrak{B})\}|$$

由于 $(a)=(b)\Leftrightarrow \frac{a}{b}\in U_K$(即主理想的生成元可以差一个可逆元)，我们设 $\mathfrak{B}$ 的一组基为 $a_1,...,a_n$，对每个元素 $a=\sum_{i=1}^{n} m_i a_i (m_i\in\mathbb{Z})$ 做映射

$$\varphi:\mathfrak{B}\to\mathbb{R}^n, a\mapsto(m_1,...,m_n)$$

这使得 $\varphi(\mathfrak{B})$ 是 $\mathbb{R}^n$ 的一个格，从而进一步有

$$f(C,x)=|\{(x_1,...,x_n)\in\mathbb{Z}^n \mid a=\sum_{i=1}^{n} x_i[a_i], 0<N(a)\leq xN(\mathfrak{B})\}|$$

(2) 接着我们来看一下基 $\{a_i\}$ 在单位群 U_K 中代表元 $[a_i]$ 的情况。由单位分解定理可得 $U_K=\mathbb{Z}^r\oplus W_K$，设自由部分的基本单位组为 $\{\varepsilon_1,...,\varepsilon_r\}$，并定义复合映射

$$K^\times\xrightarrow{\sigma}\underbrace{\mathbb{R}^\times\times...\times\mathbb{R}^\times}_{r_1}\times\underbrace{\mathbb{C}^\times\times...\times\mathbb{C}^\times}_{r_2}\xrightarrow{l}\mathbb{R}^{r_1+r_2}$$

其中 $\sigma(a)=(\sigma_1(a),...,\sigma_{r_1+r_2}(a)), a\in K^\times$，$l(y_1,...,y_{r_1+r_2})=(\ln|y_1|,...,\ln|y_{r_1}|,2\ln|y_{r_1+1}|,...,2\ln|y_{r_1+r_2}|)$，故我们只需令 $t=(\underbrace{1,...,1}_{r_1},\underbrace{2,...,2}_{r_2})$ 则有

$$\omega_K f(C,x)=|\{(x_1,...,x_n)\in\mathbb{Z}^n \mid a=\sum_{i=1}^{n} x_i a_i, 0<N(a)\leq xN(\mathfrak{B}), l\sigma(a)=ct+\sum_{i=1}^{r} c_i l\sigma(\varepsilon_i)\}|$$

(3) 然后我们进行计算转化。对于 $y=(y_1,...,y_{r_1+r_2})\in\mathbb{R}^{r_1}\times\mathbb{C}^{r_2}$，定义

$$N(y)=|y_1...y_{r_1}y_{r_1+1}^2...y_{r_1+r_2}^2|$$

此时可以给出 $N(\sigma(a))=N(\sum x_i\sigma(a_i))$，其相当于 K 中元素范的扩展，此时我们给出一个几何体

$$\Gamma_0=\{(x_1,...,x_n)\in\mathbb{R}^n \mid 0<N(\sum x_i\sigma(a_i))\leq 1, l(\sum x_i\sigma(a_i))=ct+\sum_{i=1}^{r} c_i l\sigma(\varepsilon_i), 0\leq c_i<1\}$$

因此 $\omega_K f(C,x)$ 等于 $(xN(\mathfrak{B}))^{\frac{1}{n}}\Gamma_0$ 中整点的个数，即有

$$\omega_K\lim_{x\to\infty} f(C,x)=V((xN(\mathfrak{B}))^{\frac{1}{n}}\Gamma_0)=xN(\mathfrak{B})V(\Gamma_0)+O(x^{1-\frac{1}{n}})$$

(4) 再者我们来计算体积。我们先记 $y_k=\sum_{j=1}^{n} x_j\sigma_k(a_j)(1\leq k\leq r_1), y_k+iy_{r_2+k}=\sum_{j=1}^{n} x_j\sigma_k(a_j)(r_1+1\leq k\leq r_1+r_2)$，再进行一次坐标转化来计算

$$\Gamma=\{(y_1,...,y_n)\in\Gamma_0 \mid y_i>0(1\leq i\leq r_1)\}$$

$$V(\Gamma_0)=\int_{\Gamma_0} dx_1...dx_n=2^{r_2}\int_{\Gamma} J(\frac{y_1,...,y_n}{x_1,...,x_n})^{-1}dy_1...dy_n=2^{r_2}J(\frac{y_1,...,y_n}{x_1,...,x_n})^{-1}V(\Gamma)=\frac{2^{r_1+r_2}}{N(\mathfrak{B})\sqrt{|d_K|}}V(\Gamma)$$

再接再厉，我们考虑 $\mathbb{R}^n$ 中的极坐标 $\rho_i=y_i(1\leq i\leq r_1), \rho_{r_1+j}(\cos\theta_j+i\sin\theta_j)(1\leq j\leq r_2)$，则易得

$$V(\Gamma)=\int_{\Gamma}\rho_{r_1+1}...\rho_{r_1+r_2}d\rho_{r_1+1}...d\rho_{r_1+r_2}d\theta_1...d\theta_{r_2}=(2\pi)^{r_2}\int\rho_{r_1+1}...\rho_{r_1+r_2}d\rho_{r_1+1}...d\rho_{r_1+r_2}$$

我们记 $P=\rho_1...\rho_{r_1}\rho_{r_1+1}^2...\rho_{r_1+r_2}^2$，则有 $cn=\ln P$，并进行下面的坐标转化

$$J(\frac{\rho_1,...,\rho_{r_1+r_2}}{P,c_1,...,c_r})=\frac{\rho_1...\rho_{r_1+r_2}}{nP2^{r_2}}R_K n=\frac{R_K}{2^{r_2}\rho_{r_1+1}...\rho_{r_1+r_2}}$$

其给出了

$$V(\Gamma)=(2\pi)^{r_2}\int \rho_{r_1+1}\dots\rho_{r_1+r_2}\frac{R_K}{2^{r_2}\rho_{r_1+1}\dots\rho_{r_1+r_2}}dPdc_1\dots dc_r=\pi^{r_2}R_K$$

(5) 最后，我们结合 (3) 和 (4) 的结论即可得

$$\lim_{x\to\infty}f(C,x)=\frac{2^{r_1}(2\pi)^{r_2}R_K}{\omega_K\sqrt{|d_K|}}x+O(x^{1-\frac{1}{n}})$$

第二步是在估计函数的基础上计算复解析函数的留数，有下面的定理。

定理 5.7

复变函数 $\zeta_K(C,s)$ 和 $\zeta_K(s)$ 均可以解析延拓到半平面 $\mathrm{Re}(s)>1-\frac{1}{n}$，并且仅在 $s=1$ 处有单极点，相应的留数为

$$\mathrm{Res}[\zeta_K(C,s),1]=\rho_K,\mathrm{Res}[\zeta_K(s),1]=h_K\rho_K$$

♡

证明 我们只需证明前半部分，后面的由 $\zeta_K(s)=\sum_{C\in C_K}\zeta_K(C,s)$ 可以自然得到。我们先进行一些简单的变形

$$\zeta_K(C,s)=\rho_K\zeta(s)+\sum_{n=1}^{\infty}\frac{a_n(C)-\rho_K}{n^s}$$

左半部分级数 $\rho_K\zeta(s)$ 仅在 $s=1$ 处有单极点，此时考虑右半部分的级数有

$$\sum_{n=1}^{\infty}(a_n(C)-\rho_K)=\lim_{N\to\infty}f(C,N)-\rho_K N=O(N^{1-\frac{1}{n}})$$

于是由形式 D-级数的收敛区域存在定理可知，其在 $\mathrm{Re}(s)>1-\frac{1}{n}$ 内解析。综合可得，$\zeta_K(C,s)$ 仅在 $s=1$ 处有单极点，并且有

$$\mathrm{Res}[\zeta_K(C,s),1]=\mathrm{Res}[\rho_K\zeta(s),1]+0=\rho_K\mathrm{Res}[\zeta(s),1]=\rho_K$$

没有第三步了，因为想要的东西已经在其中了。借助一阶极点的留数性质 $\mathrm{Res}[f(z),z_0]=\lim_{z\to z_0}(z-z_0)f(z)$，就可以推出想要的类数公式了

$$\lim_{s\to1}(s-1)\zeta_K(s)=\frac{2^{r_1}(2\pi)^{r_2}R_Kh_K}{\omega_K\sqrt{|d_K|}}$$

我们必需知道一点，类数公式并不是直接由一些基础量计算出类数，而是通过一些基础量将类数的计算转化为亚纯函数留数的计算，或者我们也可以持有反过来的观点。其实讲得也差不多了，最后给个函数方程观赏一下吧。

$$\Lambda_K(s)\overset{def}{:=}(\frac{\sqrt{|d_K|}}{2^{r_2}\pi^{\frac{n}{2}}})^s\Gamma(\frac{s}{2})^{r_1}\Gamma(s)^{r_2}\zeta_K(s)$$

$$\Lambda_K(s)=\Lambda_K(1-s)$$

对于 $\zeta_K(s)$，$s=0$ 是 r 阶零点，$s=-2,-4,-6,\dots$ 是 $r+1$ 阶零点，($r_2\neq0$ 时)$s=-1,-3,-5,\dots$ 是 r_2 阶零点，它们通通称为平凡零点，而非平凡的零点在 $0<\mathrm{Re}(s)<1$ 内且关于 $\mathrm{Re}(s)=\frac{1}{2}$ 对称，自然可以导出对应的推广黎曼猜想，不过有点“食之无味，弃之可惜”之感。

5.2 素数分布

素数定理

解析数论的初衷或许就是对素数的研究了，从各种 L 函数的欧拉积中自然能感受到这点，因此介绍素数定理就是无可厚非的了。我们先定义一个用来指示素数的数论函数

$$f(n)=\begin{cases}1 & n\text{是素数}\\ 0 & \text{其它}\end{cases},\pi(x)=\sum_{n\le x}f(n)=\sum_{p\le x}1$$

即 $\pi(x)$ 表示不超过 x 的素数的个数，由于素数有无穷多个所以显然有 $\lim_{x\to+\infty}\pi(x)=+\infty$。而**素数定理**实际就是找一个较熟悉较好研究的函数 $f(x)$ 使得有极限

$$\lim_{x\to+\infty}\frac{\pi(x)}{f(x)}=1$$

较为常见的函数有两个，它们分别是

$$f(x)=\frac{x}{\ln x},f(x)=\mathrm{Li}(x)=\lim_{\eta\to0}(\int_0^{1-\eta}+\int_{1+\eta}^{x})\frac{du}{\ln u}$$

也就是说我们实际上是在找素数的分布规律，如果我们能完全解析 $\pi(x)$ 那么素数的各种性质就都不在话下。比如**素数的间隔问题**，我们知道最大间隔是可以任意的，即

$$n!+2,n!+3,n!+4,...,n!+n$$

最小间隔是 1 且只有 $(2,3)$ 是没什么好说的，于是我们可以进一步考虑间隔为 2 的素数个数，**孪生素数猜想**说它有无数个。如果我们用 $[x]$ 表示向上取整函数，则 $[x]$ 表示不超过 x 的整数的个数，则通过简单的古典筛法很容易证明素数的稀疏性

$$\lim_{x\to+\infty}\frac{\pi(x)}{[x]}=\lim_{x\to+\infty}\frac{\pi(x)}{x}=0$$

证明 我们还是用分析语言稍微证明一下吧，首先是一些基础的不等转化

$$\begin{aligned}|\frac{\pi(x)}{[x]}-0|&\le\frac{\varphi(x)}{x}\\&=\prod_{p|x}(1-\frac{1}{p})\\&\le\prod_{p\le N}(1-\frac{1}{p})\end{aligned}$$

我们解释一下上面的不等式，$[x]\ge x$ 由定义显然导出，$\pi(x)\le\varphi(x)$ 也是显然的因为素数一定与任何数互素，N 表示所有满足 $p\mid x$ 的素数相乘，即项数变多了，由 $p>1\Rightarrow(1-\frac{1}{p})<1$ 即可得到相应的不等式。于是我们只需要证明下面的极限

$$\lim_{x\to+\infty}\prod_{p\le x}(1-\frac{1}{p})=0$$

我们考虑倒数的情况，可得

$$\begin{aligned}\frac{1}{\prod_{p\le x}(1-\frac{1}{p})}&=\prod_{p\le x}\frac{1}{(1-\frac{1}{p})}\\&=\prod_{p\le x}(1+\frac{1}{p}+\frac{1}{p^2}+...)\\&>1+\frac{1}{2}+\frac{1}{3}+...+\frac{1}{x-1}\\&>\ln(x)\end{aligned}$$

即有 $\prod_{p\le x}(1-\frac{1}{p})<\frac{1}{\ln(x)}$，由 $\lim_{x\to+\infty}\frac{1}{\ln(x)}=0$ 即可得到我们的结论。(还可以通过证明估计式 $\exists C,\frac{\pi(x)}{x}\le\frac{C}{\ln\ln x},x\ge$

3 来证明，又或者可以通过等式 $\pi(x)=\pi(\sqrt{x})-1+\sum_d \mu(d)[\frac{N}{d}]$ 来得到结论)

有关数论的研究我们还要指出“对数的底数无关性”，即 $\log_a x$ 中的底数 a 是什么都无关紧要，它们无非只是差一个常数 $\log_a x = C\ln x, C=\frac{1}{\ln a}$，但由于自然底数 e 的对数有诸多优良性质，故基本在没指出的情况下，无论写成 $\log x$ 还是 $\lg x$ 均指 e 作为底数。回到我们的素数定理，我们不会讨论复杂的“初等证明”，因为解析数论中的“初等”指不用复分析，以笔者的观点来看既然都用了实分析，也就没必要拐弯抹角地绕过复分析，除非你对证明有一定的追求，可以参考这本书 [41]。首先是两个常见的符号

$$f(x)=O(g(x)) \Leftrightarrow \exists C, x_0, \forall x \geq x_0, |f(x)| \leq Cg(x)$$

$$f(x)=o(g(x)) \Leftrightarrow \lim \frac{f(x)}{g(x)}=0$$

$$f(x)\sim g(x) \Leftrightarrow \lim \frac{f(x)}{g(x)}=1$$

大 O 表示数量级的估计，常用于算法理论中，小 o 表示高阶无穷小，我们在数学分析中会经常见到，最后一个是等价无穷小，常用于极限的计算中。通过上面的这些符号，我们可以把素数的稀疏性和素数定理写为

$$\pi(x)=o(x), \pi(x)\sim \frac{x}{\ln x}$$

我们的目标就是证明 $x\to+\infty$ 时的后面的等价无穷小，至于 $\frac{x}{\ln x}\sim \mathrm{Li}(x)$ 就留给读者自己去考虑了，定理证明的起点是下面的估计不等式。

定理 5.8 (切比雪夫)

存在常数 $0<A<B$ 使得当 $x\geq 2$ 时有

$$A\frac{x}{\ln x} \leq \pi(x) \leq B\frac{x}{\ln x}$$

♡

证明 (1) 我们记 $\alpha=\alpha(p,n)$ 表示 $p^\alpha \,||\, n!$，即 $n!$ 中素因子 p 的指数，我们先要给出它的计算式。我们直接猜测为

$$\alpha(p,n)=\sum_{i=1}^{\infty}[\frac{n}{p^i}]$$

并用归纳法来证明，其中 $[x]$ 在这里及以后都表示向下取整 (对应不等式为 $[x]\leq x$)。当 $n=1$ 时，$\forall p, \alpha(p,n)=0$ 显然成立，假设小于 n 时结论成立，并设 $p^\lambda \,||\, n$，则有

$$\begin{aligned}\alpha(p,n)&=\alpha(p,n-1)+\lambda\\&=\alpha(p,n-1)+\sum_{i=1}^{\lambda}([\frac{n}{p^i}]-[\frac{n-1}{p^i}])+\sum_{i=\lambda+1}^{\infty}([\frac{n}{p^i}]-[\frac{n-1}{p^i}])\\&=\alpha(p,n-1)+\sum_{i=1}^{\infty}[\frac{n}{p^i}]-\sum_{i=1}^{\infty}[\frac{n-1}{p^i}]\\&=\sum_{i=1}^{\infty}[\frac{n}{p^i}]\end{aligned}$$

结合唯一分解式 $n!=\prod_{p\leq n} p^{\alpha(p,n)}$ 可得

$$\ln n!=\sum_{p\leq n}\alpha(p,n)\ln p$$

(2) 我们记一个特殊数 $N=\frac{(2n)!}{(n!)^2}$，则有下面的不等式

$$2^n \leq \frac{2n}{n}\frac{2n-1}{n-1}\cdots\frac{n+1}{1}=N=C_{2n}^n \leq (1+1)^{2n}=2^{2n}$$

左边是因为 $\frac{2n-i}{n-i}=2+\frac{i}{n-i}>2$，而右边是因为 C_{2n}^{n} 为 $(1+x)^{2n}$ 展开中 x^n 的系数。取对数可得

$$n\ln 2\le \ln N\le 2n\ln 2$$

(3) 然后再试着联系到我们的函数，先进行等式转化

$$\ln N=\ln(2n)!-2\ln n!=\sum_{p\le 2n}\alpha(p,2n)\ln p-2\sum_{p\le n}\alpha(p,n)\ln p=\sum_{p\le n}(\alpha(p,2n)-2\alpha(p,n))\ln p+\sum_{n<p\le 2n}\alpha(p,2n)\ln p$$

接着计算等式最后的两项

$$\sum_{n<p\le 2n}\alpha(p,2n)\ln p=\sum_{n<p\le 2n}\ln p>n\sum_{n<p\le 2n}1=(\pi(2n)-\pi(n))\ln n$$

$$0\le\sum_{p\le n}(\alpha(p,2n)-2\alpha(p,n))\ln p=\sum_{p\le n}\sum_{i=1}^{[\frac{\ln 2n}{\ln p}]}([\frac{2n}{p^i}]-2[\frac{n}{p^i}])\ln p\le\sum_{p\le n}\sum_{i=1}^{[\frac{\ln 2n}{\ln p}]}\ln p=\sum_{p\le n}[\frac{\ln 2n}{\ln p}]\ln p$$

$$\sum_{p\le n}[\frac{\ln 2n}{\ln p}]\ln p+\sum_{n<p\le 2n}\alpha(p,2n)\ln p\le\sum_{p\le 2n}[\frac{\ln 2n}{\ln p}]\ln p\le\sum_{p\le 2n}\frac{\ln 2n}{\ln p}\ln p=\ln 2n\sum_{p\le 2n}1=\pi(2n)\ln 2n$$

综合上面的不等式可得

$$(\pi(2n)-\pi(n))\ln n\le\ln N\le\pi(2n)\ln 2n$$

(4) 最后，当 $x>9$ 时考虑奇偶性即可得 $\pi(x)\le\frac{x}{2}$，结合 $\pi(8)=4,\pi(4)=2,\pi(2)=1$ 即可得 $\forall k\ge 0,\pi(2^{k+1})\le 2^k$。接着我们取 $n=[\frac{x}{2}]$，则可算出

$$\pi(x)\ln x\ge\pi(2n)\ln 2n\ge\ln N\ge n\ln 2>\frac{\ln 2}{3}x\Rightarrow A=\frac{\ln 2}{3}$$

容易证明 $m\pi(2^m)<3\times 2^m$，于是对于 x 我们可以找到一个 m 使得 $2^{m-1}\le x<2^m$，因此有

$$\pi(x)\le\pi(2^m)<\frac{3\times 2^m}{m}=6\times 2^{m-1}\frac{1}{m}\le 6x\frac{\ln 2}{\ln x}=6\ln 2\frac{x}{\ln x}\Rightarrow B=6\ln 2$$

我们所给的两个上下界常数其实还不够靠近 $A=\frac{\ln 2}{3}\approx 0.23105, B=6\ln 2\approx 4.15888$，在 [22] 中给出了一个十分靠近 1 的上下界常数 $A\approx 0.95695, B\approx 1.04423$，不过它要求 x 足够大才行，因此与其继续寻找更靠近的常数，不如去考虑素数定理。我们先来对函数进行转化，目的是为了靠近 $\zeta(s)$ 函数，引入函数

$$\theta(x)=\sum_{p\le x}\ln p,\psi(x)=\sum_{p^m\le x}\ln p=\sum_{n\le x}\Lambda(n),\Lambda(n)=\begin{cases}\ln p & n=p^m(m\ge 1)\\ 0 & \text{other}\end{cases}$$

于是有下面的等价转化关系

$$\pi(x)\sim\frac{x}{\ln x}\Leftrightarrow\theta(x)\sim x\Leftrightarrow\psi(x)\sim x$$

证明 (1) 我们先来证明右半部分，只需说明 $\theta(x)\sim\psi(x)$ 即可。我们需要先得到几个常用的不等式，给定 $0<\lambda<1, x\ge 1$ 我们有

$$\theta(x)=\sum_{p\le x}\ln p\le\sum_{p\le x}\ln x=\pi(x)\ln x$$

$$\theta(x)=\sum_{p\le x^\lambda}\ln p+\sum_{x^\lambda<p\le x}\ln p\ge 0+\sum_{x^\lambda<p\le x}\ln x^\lambda=\lambda\ln x(\pi(x)-\pi(x^\lambda))>\lambda\ln x(\pi(x)-x^\lambda)$$

注意到和的有限性 $m>[\frac{\ln x}{\ln 2}]=M\Rightarrow\theta(x^{\frac{1}{m}})=0$ 则可以得到

$$\psi(x)=\sum_{p^m\le x}\ln p=\sum_{m=1}^{M}\sum_{p\le x^{\frac{1}{m}}}\ln p=\sum_{m=1}^{M}\theta(x^{\frac{1}{m}})=\sum_{m=1}^{\infty}\theta(x^{\frac{1}{m}})$$

通过上面的联系式，可以直接估计出两个函数间的不等关系

$$\theta(x) \le \psi(x) = \sum_{m=1}^{M} \theta(x^{\frac{1}{m}}) \le \theta(x) + \sum_{m=2}^{M} x^{\frac{1}{m}} \ln x^{\frac{1}{m}} \le \theta(x) + Mx^{\frac{1}{2}} \ln x^{\frac{1}{2}} \le \theta(x) + \frac{\ln x}{\ln 2} x^{\frac{1}{2}} \ln x^{\frac{1}{2}} < \theta(x) + x^{\frac{1}{2}} (\ln x)^2$$

由于上面的函数均是正的，故有

$$1 \le \frac{\psi(x)}{\theta(x)} \le 1 + \frac{x^{\frac{1}{2}} (\ln x)^2}{\theta(x)} \le 1 + \frac{x^{\frac{1}{2}} (\ln x)^2}{\frac{1}{2}(\pi(x) - x^{\frac{1}{2}}) \ln x} \le 1 + \frac{2x^{\frac{1}{2}} \ln x}{A\frac{x}{\ln x} - x^{\frac{1}{2}}} = 1 + \frac{2}{A\frac{x^{\frac{1}{2}}}{(\ln x)^2} - 1}$$

我们显然有 $\lim_{x\to+\infty} \frac{x^{\frac{1}{2}}}{(\ln x)^2} = +\infty$，由极限的迫敛性即可得到

$$\lim_{x\to+\infty} \frac{\psi(x)}{\theta(x)} = 1$$

(2) 接着证明左半部分，同样地我们需要先得到阿贝尔变换，令 $f(t)$ 在区间 $[1,x]$ 上连续可微，$s(x) = \sum_{n\le x} c_n$ 是复数列 $\{c_n\}$ 的和函数，则有

$$s(x)f(x) - \sum_{n\le x} c_n f(n) = \sum_{n\le x} c_n (f(x) - f(n)) = \sum_{n\le x} c_n \int_n^x f'(t)dt = \sum_{n\le x} c_n \int_1^x 1_{[n,t]} f'(t)dt = \int_1^x s(t)f'(t)dt$$

我们取 $f(t) = \ln t, c_n = \begin{cases} 1 & n\text{是素数} \\ 0 & \text{其它} \end{cases}$ 则有

$$\pi(x)\ln x - \theta(x) = \int_1^x \frac{\pi(t)}{t}dt = (\int_1^{\sqrt{x}} + \int_{\sqrt{x}}^x)\frac{\pi(t)}{t}dt \le \int_1^{\sqrt{x}} dt + \int_{\sqrt{x}}^x \frac{B\frac{t}{\ln t}}{t}dt \le \sqrt{x} + \frac{B}{\ln\sqrt{x}}\int_{\sqrt{x}}^x dt = O(\frac{x}{\ln x})$$

注意到我们有 $\pi(x) = O(\frac{x}{\ln x})$，由此可以得到

$$\ln x - \frac{\theta(x)}{\pi(x)} = \frac{1}{O(\frac{x}{\ln x})} O(\frac{x}{\ln x}) = O(1)$$

由定义可知存在 C, x_0 使得 $|\ln x - \frac{\theta(x)}{\pi(x)}| \le C$，解得

$$|1 - \frac{\theta(x)}{\pi(x)\ln x}| \le \frac{C}{\ln x}, x \ge \max\{x_0, 1\}$$

因为 $\lim_{x\to+\infty} \frac{1}{\ln x} = +\infty$，由极限的迫敛性可得

$$\lim_{x\to+\infty} \frac{\theta(x)}{\pi(x)\ln x} = 1$$

注 如果你不熟悉极限的话，可能会觉得 (2) 的证明有点小“奇怪”，这是我故意这么做的，目的是想要读者理解分析意味着什么？在解析数论中我们通常都会回归极限的基本定义，即 $\varepsilon - \delta$ 语言，因此不等是分析证明中的核心，而在解析数论中引入的大 O 记号则是简化不等式计算的工具，它们的基础关系如下。

$$\text{分析} \overset{\varepsilon-\delta}{\longleftrightarrow} \text{不等式} \overset{f=O(g)}{\longleftrightarrow} \text{大 O 代数}$$

之前切比雪夫不等式的含义就是 $\pi(x) = O(\frac{x}{\ln x})$，在证明某些极限的时候，我们经常会使用极限的迫敛性来进行极限的证明，大 O 代数给出了我们想要的不等式，它与等价无穷小 (或无穷大) 的基本关系为

$$f(x) - g(x) = O(1) \Leftrightarrow f \sim g$$

我们再来稍微重塑一下上面 (2) 的证明，它的关键步骤是

$$\text{(i)}\ \ \pi(x)\ln x - \theta(x) = (\int_1^{\sqrt{x}} + \int_{\sqrt{x}}^x)\frac{\pi(t)}{t}dt = O(\frac{x}{\ln x})$$

$$\text{(ii)}\ \ \pi(x) = O(\frac{x}{\ln x}) \Rightarrow \ln x - \frac{\theta(x)}{\pi(x)} = O(1)$$

对于 (i) 式，主要是后半部分，借助基本不等式 $\int_1^{\sqrt{x}} \frac{\pi(t)}{t}dt = O(\sqrt{x}), \pi(x) = O(\frac{x}{\ln x})$ 可得

$$(\int_1^{\sqrt{x}} + \int_{\sqrt{x}}^x)\frac{\pi(t)}{t}dt = O(\sqrt{x}) + \int_{\sqrt{x}}^x \frac{O(\frac{t}{\ln t})}{t}dt$$

利用基本不等式 $f = hO(g) = O(hg)$ 和积分的不等式性质有

$$\int_{\sqrt{x}}^{x} \frac{O(\frac{t}{\ln t})}{t} dt = \int_{\sqrt{x}}^{x} O(\frac{1}{\ln t}) dt = O(\int_{\sqrt{x}}^{x} \frac{1}{\ln t} dt)$$

再利用基本不等式 $h \le g \Rightarrow f = O(h) = O(g)$ 和 $f(x) = O(kg(x)) = O(g(x)), k \in \mathbb{R}$ 可得

$$O(\int_{\sqrt{x}}^{x} \frac{1}{\ln t} dt) = O(\int_{\sqrt{x}}^{x} \frac{1}{\ln \sqrt{x}}) = O(\frac{2}{\ln x}(x - \sqrt{x})) = O(\frac{x - \sqrt{x}}{\ln x})$$

最后利用 $\forall x > 1, \frac{x-\sqrt{x}}{\ln x} < \frac{x}{\ln x}$ 和基本不等式 $f < g \Rightarrow O(f) + O(g) = O(g)$ 可得

$$O(\sqrt{x}) + O(\frac{x - \sqrt{x}}{\ln x}) = O(\sqrt{x}) + O(\frac{x}{\ln x}) = O(\frac{x}{\ln x})$$

实际上，这里的 $\sqrt{x} < \frac{x}{\ln x}$ 我们只需足够大的 x 成立即可，用极限来表达就是

$$\lim_{x \to +\infty} \frac{\sqrt{x}}{\frac{x}{\ln x}} = 0$$

对于 (ii) 式，有了上面的准备就十分简单了

$$\ln x - \frac{\theta(x)}{\pi(x)} = \frac{\pi(x) \ln x - \theta(x)}{\pi(x)} = O(\frac{x}{\pi(x) \ln x}) = O(1)$$

这里用了一个显然的结论

$$f = O(h) \Rightarrow k = O(gf) = fO(g) = O(h)O(g) = O(gh)$$

有关大 O 记号读者可以自行探索更多的东西，我们就点到为止了。

$$\frac{\theta(x)}{\pi(x)} \sim \ln x \quad \frac{\psi(x)}{\pi(x)} \sim \ln x$$

因此我们证明素数定理实际上有两条路径，但它们本质上是一样的，我们先来说明一些共性的内容，也就是黎曼 zeta 函数的几个重要基本性质 (不包括我们已经讲过的欧拉积、单极点、平凡零点、函数方程等[1])。

定理 5.9 (基本性质)

(1) 当 $\mathrm{Re}(s) > 1$ 时，$\frac{\zeta'(s)}{\zeta(s)} = -\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}$。

(2) 当 $\mathrm{Re}(s) = 1$ 时，$\zeta(s) \neq 0$。

(3) 记 $s = \sigma + it, \sigma, t \in \mathbb{R}$。$\forall 0 \le \sigma_0 \le 1, \varepsilon > 0, \exists c_\varepsilon$

(i)$\sigma \ge \sigma_0, |t| \ge 1 \Rightarrow |\zeta(s)| \le c_\varepsilon |t|^{1-\sigma_0+\varepsilon}$

(ii)$\sigma \ge 1, |t| \ge 1 \Rightarrow |\zeta'(s)| \le c_\varepsilon |t|^{\varepsilon}$

(iii)$\sigma \ge 1, |t| \ge 1 \Rightarrow |\frac{1}{\zeta(s)}| \le c_\varepsilon |t|^{\varepsilon}$

♡

证明 (1) 我们先利用在 $\mathrm{Re}(s) > 1$ 内的欧拉积和一致收敛性可得

$$\ln \zeta(s) = \ln \prod_p (1 - p^{-s})^{-1} = -\sum_p \ln(1 - p^{-s}) = \sum_p \sum_{m=1}^{\infty} \frac{(p^{-s})^m}{m} = \sum_p \sum_{m=1}^{\infty} \frac{1}{(p^m)^s m} = \sum_{n=1}^{\infty} \frac{c_n}{n^s}$$

其中 $c_n = \begin{cases} \frac{1}{m} & n = p^m \\ 0 & \text{其它} \end{cases}$，考察所有的求和项，n 只会取到所有素数及其相应的幂 $p, p^1, p^2, ..., p^m, ...$，接着我们对左右两边求导可得

$$\frac{\zeta'(s)}{\zeta(s)} = H'(s), H(s) = \sum_{n=1}^{\infty} \frac{c_n}{n^s}$$

(引理：) 对于收敛 D-级数 $F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$，有 $F'(s) = -\sum_{n=1}^{\infty} \frac{a_n \ln n}{n^s}$(参考 [19] 的 Proposition 1.7.10，这并非平凡

[1]欧拉积：$\zeta(s) = \prod_p (1 - p^{-s})^{-1}$，$s = 1$ 是留数为 1 的一阶极点，平凡零点为负偶数：-2,-4,...，函数方程：$\zeta(1-s) = 2(2\pi)^{-s}\Gamma(s)\cos(\frac{\pi s}{2})\zeta(s)$

结论)。运用这个的引理可得

$$\frac{\zeta'(s)}{\zeta(s)} = -\sum_{n=1}^{\infty}\frac{c_n \ln n}{n^s} = -\sum_{n=1}^{\infty}\frac{1_{\{p^m\}}\frac{1}{m}\ln p^m}{n^s} = -\sum_{n=1}^{\infty}\frac{1_{\{p^m\}}\ln p}{n^s} = -\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s}$$

(2) 换个说法就是 $\forall t \in \mathbb{R}, \zeta(1+it) \neq 0$。首先要证明一个引理不等式，对任意的 $\sigma > 1, t \in \mathbb{R}$ 有

$$\begin{aligned}
\ln|\zeta^3(\sigma)\zeta^4(\sigma+it)\zeta(\sigma+2it)| &= 3\ln|\zeta(\sigma)| + 4\ln|\zeta(\sigma+it)| + \ln|\zeta(\sigma+2it)| \\
&= 3\mathrm{Re}(\ln\zeta(\sigma)) + 4\mathrm{Re}(\ln\zeta(\sigma+it)) + \mathrm{Re}(\zeta(\sigma+2it)) \\
&= \sum_{n=1}^{\infty} c_n n^{-\sigma}(3 + 4\cos(t\ln n) + \cos(2t\ln n)) \\
&= \sum_{n=1}^{\infty} c_n n^{-\sigma} 2(1+\cos(t\ln n))^2 \\
&\geq 0
\end{aligned}$$

其中黎曼 zeta 函数取对数后的实部，运用了 (1) 中的第一个式子

$$\mathrm{Re}(\ln\zeta(a+ib)) = \sum_{n=1}^{\infty}\frac{c_n}{\mathrm{Re}(n^{a+ib})} = \sum_{n=1}^{\infty} c_n \mathrm{Re}(e^{-(a+ib)\ln n}) = \sum_{n=1}^{\infty} c_n e^{-a\ln n}\cos(t\ln n) = \sum_{n=1}^{\infty} c_n n^{-a}\cos(t\ln n)$$

然后使用反证法，假设存在 $t_0 \neq 1$ 使得 $\zeta(1+it_0) = 0$。运用上述不等式可得

$$|(\sigma-1)\zeta(\sigma)|^3 |\frac{\zeta(\sigma+it_0)}{\sigma-1}|^4 |\zeta(\sigma+2it_0)| \geq \frac{1}{\sigma-1}, \sigma > 1$$

在上述的三个成分中，我们取 $\sigma \to 1$，则有 $(\sigma-1)\zeta(\sigma) \to C$(一阶极点 $\sigma = 1$)、$\frac{\zeta(\sigma+it_0)}{\sigma-1} \to 0, C$(至少一阶零点 $1+it_0$)、$\zeta(\sigma+2it_0) \to \zeta(1+2it_0) \neq 0$(全纯且正项级数大于零)，由于 $4-3=1>0$，故左边整体上至少还有一阶零点，即 左边 $\to 0$，但显然 右边 $\to \infty$。矛盾，从而定理得证。

(3) 最后的三个是 zeta 函数的相关估计性质，我们就随便算算，细节留给读者自行补充

$$\begin{aligned}
|\zeta(s)| = |\frac{1}{1-s} + \lim_{N\to\infty}(\sum_{1\leq n<N}\frac{1}{n^s} - \int_1^N \frac{dx}{x^s})| &\leq |\frac{1}{1-s}| + |\lim_{N\to\infty}\sum_{1\leq n<N}\delta_n(s)| \\
&\leq |\frac{1}{1-s}| + \lim_{N\to\infty}\sum_{1\leq n<N}|\delta_n(s)| \\
&\leq |\frac{1}{1-s}| + \lim_{N\to\infty}\sum_{1\leq n<N}(\frac{|s|}{n^{\sigma_0+1}})^{1-(\sigma_0-\varepsilon)}(\frac{2}{n^{\sigma_0}})^{\sigma_0-\varepsilon} \\
&= |\frac{1}{1-s}| + \sum_{n=1}^{\infty}\frac{2|s|^{1-(\sigma_0-\varepsilon)}}{n^{1+\varepsilon}} \\
&= |\frac{1}{1-s}| + c|s|^{1-\sigma_0+\varepsilon}, c = 2\zeta(1+\varepsilon)
\end{aligned}$$

$$\begin{aligned}
|\zeta'(s)| = |\frac{1}{2\pi\varepsilon}\int_0^{2\pi}\zeta(s+\varepsilon e^{i\theta})e^{i\theta}d\theta| &\leq \frac{1}{2\pi\varepsilon}\int_0^{2\pi}|\zeta(s+\varepsilon e^{i\theta})|d\theta \\
&\leq \frac{1}{2\pi\varepsilon}\int_0^{2\pi} c_\varepsilon |t|^{1-\sigma_0+\varepsilon}d\theta \\
&\leq \frac{1}{2\pi\varepsilon}\int_0^{2\pi} c_\varepsilon |t|^{\varepsilon-1}d\theta \quad (|t|\geq 1, \sigma_0 \geq 2\sigma \geq 2 \Rightarrow 1-\sigma_0 \leq -1) \\
&= \frac{1}{2\pi\varepsilon}2\pi c_\varepsilon \varepsilon |t|^{\varepsilon} \\
&= c_\varepsilon |t|^{\varepsilon}
\end{aligned}$$

$$
\begin{aligned}
|\zeta(s)| = |\zeta(\sigma'+it) - \zeta(\sigma'+it) + \zeta(\sigma+it)| &\geq |\zeta(\sigma'+it)| - |\zeta(\sigma'+it) - \zeta(\sigma+it)| \quad (\sigma' = 1 + A|t|^{-5\varepsilon}) \\
&\geq |\zeta(\sigma'+it)^4|^{\frac{1}{4}} - c''|\sigma' - \sigma||t|^{\varepsilon} \\
&\geq |\zeta(\sigma')^{-3}\zeta(\sigma'+2it)^{-1}|^{\frac{1}{4}} - c''|\sigma' - 1||t|^{\varepsilon} \\
&\geq ((\sigma-1)^3 c_{\varepsilon}|t|^{-\varepsilon})^{\frac{1}{4}} - c''|\sigma' - 1||t|^{\varepsilon} \\
&\geq c'(\sigma'-1)^{\frac{3}{4}}|t|^{-\frac{\varepsilon}{4}} - c''|\sigma' - 1||t|^{\varepsilon} \\
&= 2c''|\sigma' - 1||t|^{\varepsilon} - c''|\sigma' - 1||t|^{\varepsilon} \quad (A = (\frac{c'}{2c''})^4) \\
&= c|t|^{-(4\varepsilon)}, c = c''A
\end{aligned}
$$

注 对于第一个性质，在 [43] 的第六章第 3 节的定理 4 中给出了一般性的结论：对于收敛 D-级数 $F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}$，如果 $f(n)$ 是完全积性函数，则有

$$
\frac{F'(s)}{F(s)} = -\sum_{n=1}^{\infty} \frac{\Lambda(n) f(n)}{n^s}
$$

对于第二个性质，关键点是不等式 $|\zeta^3(\sigma)\zeta^4(\sigma+it)\zeta(\sigma+2it)| > 1$，实际上还可以构造更多的不等式，但我们所给的这个 (也是大多数书所采用的) 算是目前来看最简便的。

最后就可以来证明素数定理了，我们采用的是最常见的 $\psi(x) \sim x$ 路径，至于另一条路径也是可行的，只是大多数书籍都不采用，我也就只能跟风学习了。

定理 5.10 (素数定理证明路径)

定义函数 $\psi_1(x) = \int_1^x \psi(t)dt$ 则有

$$
(1)\ \forall c > 0, \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{a^s}{s(s+1)}ds = \begin{cases} 0 & 0 < a \leq 1 \\ 1 - \frac{1}{a} & 1 \leq a \end{cases}
$$

$$
(2)\ \forall c > 1, \psi_1(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})ds
$$

$$
\text{核心内容} \begin{cases} (3) & \psi_1(x) \sim \frac{x^2}{2} \\ & \Downarrow \\ (4) & \psi(x) \sim x \end{cases}
$$

♡

证明 (1) 回忆一下此处积分路径的含义 $\int_{c-i\infty}^{c+i\infty} f(s)ds = \lim_{R\to\infty}\int_{c-iR}^{c+iR} f(s)ds$，且容易知道 $f(s) = \frac{a^s}{s(s+1)}$ 只有两个极点 $s = 0, 1$。当 $a \geq 1$ 时，我们需要补充一个半圆区域 $C(R) = \{x \in \mathbb{C} \mid |x - c| = R, \mathrm{Re}(x) \leq c\}$ 以形成积分围道 $C_R = C(R) \cup \{c + it \in \mathbb{R} \mid |t| \leq R\}$，并计算相应的积分值

$$
|\int_{C(R)} \frac{a^s}{s(s+1)}ds| \leq \int_{C(R)} \frac{|a^s|}{|s(s+1)|}ds \leq \int_{C(R)} \frac{a^c}{\frac{1}{2}R^2}ds = 2a^c\pi\frac{1}{R} \to 0 (R \to \infty)
$$

于是可得

$$
\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{a^s}{s(s+1)}ds = \frac{1}{2\pi i}\int_{C_R} \frac{a^s}{s(s+1)}ds = \mathrm{Res}[f(s), 0] + \mathrm{Res}[f(s), -1] = 1 - \frac{1}{a}
$$

当 $0 < a \leq 1$ 时，我们只需补充另一个方向的半圆区域 $C(R) = \{x \in \mathbb{C} \mid |x - c| = R, \mathrm{Re}(x) \geq c\}$，即可得到

$$
\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{a^s}{s(s+1)}ds = \frac{1}{2\pi i}\int_{C_R} \frac{a^s}{s(s+1)}ds = 0 (\text{区域内全纯})
$$

两种情况的关键区别是 $a \geq 1, \mathrm{Re}(s) \leq c \Rightarrow |a^s| \leq a^c$ 和 $0 < a \leq 1, \mathrm{Re}(s) \geq c \Rightarrow |a^s| \leq a^c$，以保证我们后面可以使得半圆路径的积分为零。

(2) 我们使用 $1_A(x)=\begin{cases}1 & x\in A\\ 0 & x\notin A\end{cases}$ 来代表指示函数，可以得到两个重要的等式

$$\psi(x)=\sum_{n\le x}\Lambda(n)=\sum_{n=1}^{\infty}\Lambda(n)1_{n\le x}$$

$$\psi_1(x)=\int_1^x\psi(t)dt=\int_1^x\sum_{n=1}^{\infty}\Lambda(n)1_{n\le t}dt=\sum_{n=1}^{\infty}\Lambda(n)\int_1^x 1_{n\le t}dt=\sum_{n=1}^{\infty}\Lambda(n)(1_{n\le x}\int_n^x dt)=\sum_{n\le x}\Lambda(n)(x-n)$$

最后我们只需结合前面的各种结论，即可得

$$\begin{aligned}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})ds&=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x^{s+1}}{s(s+1)}(\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s})ds\\&=x\sum_{n=1}^{\infty}\Lambda(n)\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x^s}{s(s+1)n^s}ds\\&=x\sum_{n=1}^{\infty}\Lambda(n)(\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{(\frac{x}{n})^s}{s(s+1)})ds\\&=x\sum_{n=1}^{\infty}\Lambda(n)(1-\frac{n}{x})\\&=\psi_1(x)\end{aligned}$$

(3) 我们实际要估计的就是 (2) 所给的积分式。我们记 $F(s)=\frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})$。我们知道 $s=1$ 是 $\zeta(s)$ 仅有的留数为 1 的一阶极点，于是可以在 $s=1$ 处展开洛朗级数得

$$\zeta(s)=\frac{1}{s-1}+H(s)$$

其中 $H(s)$ 是全纯函数，因此存在全纯函数 $h(s)$ 使得 $-\frac{\zeta'(s)}{\zeta(s)}=\frac{1}{s-1}+h(s)$，换言之 $-\frac{\zeta'(s)}{\zeta(s)}$ 也同样仅有留数为 1 的一阶极点 $s=1$，故 $F(s)$ 仅有 3 个极点 $s=-1,0,1$。我们构造积分路径

$$\gamma(T)=\{1+ix\mid|x|\ge T\}\cup\{a\pm iT\mid 1\le a\le c\}\cup\{c+ix\mid|x|\le T\}$$

其相对于原路径仅经过全纯区域，故两条路径上的积分值相等

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}F(s)ds=\frac{1}{2\pi i}\int_{\gamma(T)}F(s)ds$$

接着我们来把 $s=1$ 给圈起来，构造另外的一条路径以形成围道

$$\gamma(T,\delta)=\{1+ix\mid|x|\ge T\}\cup\{a\pm iT\mid 1-\delta\le a\le 1\}\cup\{1-\delta+ix\mid|x|\le T\}$$

实际上我们就是把 $\mathrm{Re}(s)=1$ 往左拉出了一个 $\delta>0$ 的小矩形，与 $\gamma(T)$ 正好形成了对应，这里的 δ 是足够小的，以保证围道内没有 $s=1$ 以外的极点，并且没有零点 (讨论 $F(s)$ 的可能零点即可)。于是有

$$\frac{1}{2\pi i}\int_{\gamma(T)}F(s)ds=\mathrm{Res}[F(s),1]+\frac{1}{2\pi i}\int_{\gamma(T,\delta)}F(s)ds=\frac{x^2}{2}+\frac{1}{2\pi i}\int_{\gamma(T,\delta)}F(s)ds$$

我们把最后一个积分的路径从下往上依次拆成 5 段直线 $\gamma(T,\delta)=\gamma_1+\gamma_2+\gamma_3+\gamma_4+\gamma_5$，然后一段段进行估计

$$|\int_{\gamma_1}F(s)ds|\le\int_{\gamma_1}|\frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})|ds\le\int_T^{\infty}\frac{x^2}{t^2}A|t|^{\frac{1}{2}}dt\le\frac{\varepsilon}{2}x^2$$

$$|\int_{\gamma_5}F(s)ds|\le\int_{\gamma_5}|\frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})|ds\le\int_{-\infty}^{-T}\frac{x^2}{t^2}A|t|^{\frac{1}{2}}dt\le\frac{\varepsilon}{2}x^2$$

$$|\int_{\gamma_3}F(s)ds|\le\int_{\gamma_3}|\frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})|ds\le\int_{-T}^{T}\frac{x^{2-\delta}}{t^2}A|t|^{\frac{1}{2}}dt\le C_Tx^{2-\delta}$$

$$|\int_{\gamma_2}F(s)ds|\le\int_{\gamma_2}|\frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})|ds\le C_T'\int_{1-\delta}^{1}x^{1+\sigma}d\sigma\le C_T'\frac{x^2}{\ln x}$$

$$|\int_{\gamma_4} F(s)ds| \le \int_{\gamma_4} |\frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})|ds \le C'_T \int_1^{1-\delta} x^{1+\sigma} d\sigma \le C'_T \frac{x^2}{\ln x}$$

综合上面的各种估计可得

$$|\psi_1(x) - \frac{x^2}{2}| \le |\frac{1}{2\pi i}\int_{\gamma(T,\delta)} F(s)ds| \le \varepsilon x^2 + C_T x^{2-\delta} + C'_T \frac{x^2}{\ln x}$$

相除可得

$$|\frac{2\phi_1(x)}{x^2} - 1| \le 2\varepsilon + 2C_T x^{-\delta} + 2C'_T \frac{1}{\ln x}$$

由于 $\lim_{x\to\infty} x^{-\delta} = 0, \lim_{x\to\infty} \frac{1}{\ln x} = 0$，故可得

$$\psi_1(x) \sim \frac{x^2}{2}$$

(4) 最后一步算是比较简单的了，我们先强行进行积分转化，往上一个命题靠近即可

$$\begin{aligned}
\frac{\psi(x)}{x} &\le \frac{1}{(\beta-1)x^2}\int_x^{\beta x} \psi(u)du \\
&= \frac{1}{(\beta-1)x^2}(\psi_1(\beta x) - \psi_1(x)) \\
&= \frac{1}{(\beta-1)}(\frac{\psi_1(\beta x)}{(\beta x)^2}\beta^2 - \frac{\psi_1(x)}{x^2}) \\
&\le \frac{1}{(\beta-1)}(\frac{1}{2}\beta^2 - \frac{1}{2}) \\
&= \frac{1}{2}(\beta+1) \to 1(\beta \to 1)
\end{aligned}$$

从而可得

$$\lim_{x\to\infty} \frac{\psi(x)}{x} = 1$$

上述证明来自 [31] 一书的第七章，个人觉得算是过程比较清晰地对素数定理的证明阐述。

狄利克雷定理

相较于素数定理，接下来要讨论的狄利克雷定理就简单多了，我们先定义一个算术级数集合

$$k \ge 2, 1 \le l < k, (l,k) = 1, A(k,l) = \{l + kt \mid t \in \mathbb{N}\} = \{n > 0 \mid n \equiv l(\mathrm{mod}k)\}$$

并且使用

$$\pi(x;k,l) = \sum_{p\le x, p\equiv l(\mathrm{mod}k)} 1$$

表示集合 $A(k,l)$ 中不超过 x 的素数的个数，于是可以得到下面的定理。

定理 5.11 (Dirichlet)

设 $k \ge 2, 1 \le l < k, (l,k) = 1$，则 $A(k,l)$ 中有无穷多个素数，或者

$$\lim_{x\to+\infty} \pi(x;k,l) = +\infty$$

♡

证明 (1) 我们用 X 表示所有模 k 的狄利克雷特征构成的集合。对每个 $\chi \in X$ 在区域 $\mathrm{Re}(s) > 1$ 内有

$$\ln L(s,\chi) = \sum_p \ln(1-\chi(p)p^{-s})^{-1} = \sum_p \sum_{m=1}^{\infty} \frac{(\chi(p)p^{-s})^m}{m} = \sum_p \chi(p)p^{-s} + \sum_p \sum_{m=2}^{\infty} \frac{\chi(p)^m}{m} p^{-sm}$$

我们记 $g_\chi(s)=\sum_p\sum_{m=2}^{\infty}\frac{\chi(p)^m}{m}p^{-sm}$，它在 $\mathrm{Re}(s)>\frac{1}{2}$ 内收敛，在 $s=1$ 处连续。另一方面我们有

$$\sum_{p\equiv l(\mathrm{mod}k)}\frac{1}{p^s}=\frac{1}{\varphi(k)}\sum_{\chi\in X}\overline{\chi}(l)\sum_p\chi(p)p^{-s}=\frac{1}{\varphi(k)}(\sum_{\chi\in X}\overline{\chi}(l)\ln L(s,\chi)-\sum_{\chi\in X}\overline{\chi}(l)g_\chi(s))$$

当特征非平凡 $\chi\neq\mathbf{1}$ 时 (排除黎曼 zeta 函数) 时，$L(s,\chi)$ 在 $s=1$ 处解析。而平凡时有

$$\ln L(s,\mathbf{1})=\ln\zeta(s)\sim-\ln(s-1),s\to1^+$$

同样地由于 $\sum_{\chi\in X}\overline{\chi}(l)g_\chi(s)$ 在 $s=1$ 处解析，故上述和式只有 $\zeta(s)$ 一项，即

$$\sum_{p\in A(k,l)}\frac{1}{p^s}\sim\frac{1}{\varphi(k)}(\mathbf{1}L(s,\mathbf{1})-0)\sim\frac{1}{\varphi(k)}(-\ln(s-1)),s\to1^+$$

因此当 $s\to1^+$ 时，$\sum_{p\in A(k,l)}\frac{1}{p^s}\to+\infty$，从而 $A(k,l)$ 中包含无穷多个素数。

$$\lim_{s\to1^+}\frac{\sum_{p\in A(k,l)}p^{-s}}{-\ln(s-1)}=\frac{1}{\varphi(k)}$$

可以看到这个证法实际上就是，欧拉使用 $\zeta(s)$ 来证明素数无限的推广。

5.3 零散内容

至此我们算是对解析数论有了一个初步的了解，通常就是两步流程：(1) 结论解析化 (2) 不等式估计。在“类数公式”一节中，我们专注于结论的解析化，即将类数的计算转化为一个复解析函数的留数计算，而在“素数分布”中，我们本身面对的就是一个分析结论，由于大多数的数论函数，都不是可微的，甚至不是连续的，导致我们只能回归极限的原始定义，进行不等式分析，进行估计，而这些内容正是“解析数论”的精髓所在。另一方面则是，数论函数 $f(n)$ 与相应的 D-级数 $L(s,f)$ 之间可建立的对应关系，使得用分析的方法讨论离散的数论成为可能。在这个大篇章的最后，我们就稍微地介绍一下，一些著名的东西是如何进行解析化的。

哥德巴赫猜想

所谓**哥德巴赫猜想**指：任一大于 2 的偶数，都可表示成两个素数之和。与之相对的有弱哥德巴赫猜想：任一大于 7 的奇数，都可表示成三个奇素数的和。由于只有 2 是偶素数且只有偶数加偶数或奇数加奇数等于偶数，故哥德巴赫猜想也可以表述成：任一大于 4 的偶数，都可表示成两个奇素数之和。于是通过-2+3 的操作就能从哥德巴赫猜想推出弱哥德巴赫猜想，而反过来是不一定可行的。

有关哥德巴赫猜想的第一个进展是通过圆法证明三素数定理 (Vinogradov’s Three Primes Theorem)：每个**充分大**的奇整数可以写成三个 (奇) 素数的和。这里的“充分大”，指存在一个下界 n_0 使得对所有的 $n\geq n_0$ 定理成立，我知道直接这样说，你肯定不好理解，所以我们就来简单阐述一下证明思路以加深理解。

圆法 (Circle Method) 的基本思想是简单的，如果我们想要 $n=n_1+\ldots+n_r,n_i\in N_i$ 的方法数，只需构造相应式子

$$\prod_{i=1}^r\sum_{t\in N_i}x^t=\sum_{i=1}^{\infty}a_nx^n$$

其中 a_n 的数值就是我们想要的，于是我们的想法是从和式中挑出 a_n，即找到 $a_n=f(\prod_{i=1}^r\sum_{t\in N_i}x^t)$，这时我们可以利用 x 的可变性来构造一个正交向量空间，使得 $1,x,\ldots,x^n,\ldots$ 是标准正交基，并且可以定义内积使得 $\langle x^i,x^j\rangle=\delta_{ij}$，这样我们就能得到 $a_n=\langle(\prod_{i=1}^r\sum_{t\in N_i}x^t),x^n\rangle=\delta_{ij}$。满足上述条件的基和内积都是存在的，即

$$(e^{2\pi ix})^n \quad \langle f(x),g(x)\rangle=\int_0^1 f(x)\overline{g(x)}dx \Rightarrow \langle (e^{2\pi ix})^n,(e^{2\pi ix})^m\rangle=\int_0^1 e^{2\pi inx}e^{-2\pi imx}dx=\begin{cases}1 & n=m\\ 0 & n\neq m\end{cases}$$

这样我们的圆法雏形实际上已经完成了，我们来试着用这个方法将三素数定理 $n=p_1+p_2+p_3$ 进行转化。对于素数列写成 $\sum_{p\leq n}e^{2\pi pi\alpha}$ 也不是不行，但各种实践表明，我们选取一个上限 $N\leq n$ 并使用下面的素数列和相应生成的待选取列会更好些

$$S_N(x)=\sum_{k\leq N}\Lambda(k)e^{2\pi kix}$$

$$S_N(x)^3=\sum_{k_1,k_2,k_3\leq N}\Lambda(k_1)\Lambda(k_2)\Lambda(k_3)e^{2\pi(k_1+k_2+k_3)ix}=\sum_n(\sum_{\substack{k_1+k_2+k_3=n\\ k_1,k_2,k_3\leq N}}\Lambda(k_1)\Lambda(k_2)\Lambda(k_3))e^{2\pi nix}$$

我们选一个上界 $N\leq n$ 就是想让项数不至于是无限的 (当然可以直接用 n)，在上述式子中，指数方的行为没有区别，但多了素数指数幂的项 p^m，于是

$$r(n)=r(n,N)=\int_0^1 S_N(x)^3e^{-2\pi nix}dx$$

在 $r(n)>0$ 时表示 n 可以写成三个“素数幂”之和。不过你先别急，后面我会告诉你如何将其转为正统的三素数定理，先来看一下 Vinogradov 的主要结论，其实就是对 $r(n)$ 的估计。

定理 5.12 (Vinogradov 定理)

对任意 $A>0$ 有

$$r(N)\sim\frac{1}{2}\mathfrak{S}(N)N^2+O(\frac{N^2}{\ln^A N})$$

其中

$$\mathfrak{S}(N)=\prod_{p|N}(1-\frac{1}{(p-1)^2})\prod_{p\nmid N}(1+\frac{1}{(p+1)^3})$$

♡

怎么证明不是我们重点，解释它并推出三素数定理才是我们的目的。上面的 $\mathfrak{S}(N)$ 是欧拉积表示，原来的形式应该是

$$\mathfrak{S}(N)=\sum_{q=1}^{\infty}\frac{\mu(q)}{\varphi(q)^3}c_q(N),c_q(N)=\sum_{\substack{a=1\\(a,q)=1}}^{q}e^{-2\pi(\frac{aN}{q})i}$$

分式的上下分别是莫比乌斯函数和欧拉函数，$c_q(N)$ 是一个乘性函数。$\mathfrak{S}(N)$ 的一个重要估计性质是，我们可以找到两个常数 $0<c_1<c_2$ 使得对所有的奇数 N 有

$$c_1<\mathfrak{S}(N)<c_2$$

而 N 为偶数时，$\mathfrak{S}(N)=0$，这些性质可以从欧拉积表示中得到。于是对于充分大的奇数 N，$\mathfrak{S}(N)\to 1$，$r(N)\to\frac{1}{2}N^2>0$，读者需要注意误差项 $O(\frac{N^2}{\ln^A N})$ 是可以为负的，但由于 $\frac{N^2}{\ln^A N}$ 被分子给放缓了增长速度，故其在充分大的情形下，比不上 $\frac{1}{2}N^2$ 项。在三素数定理的基础下，无非就只剩下两件事了，一是将充分大的下界不断地变小，二是用计算机来验证充分大之下的情况，在这篇论文 [15] 中完成了剩下的两步，从而证明了弱哥德巴赫猜想。

注 实际上，**圆法**并不是上述简单的积分转化，而是如何处理上述转化后的积分，很容易发现 $e^{ix},x\in\mathbb{R}$ 在复平面上的是一个圆，或者说 $e^{2\pi nix},x\in\mathbb{R}$ 是以 1 为周期的函数，这使得我们可以将原来的积分区间 $[0,1]$ 进行偏移

$[-\frac{1}{\tau}, 1-\frac{1}{\tau}], \tau \geq 1$，这木有啥用，但我们可以开始从中挑碎片，来将区间分成两个部分 $[-\frac{1}{\tau}, 1-\frac{1}{\tau}] = \mathfrak{M} \sqcup \mathfrak{m}$，这里的 $\mathfrak{M}$ 是所有

$$\mathfrak{M}(a,q) = \{\alpha \mid |\alpha - \frac{a}{q}| < \frac{1}{Q}\}, 1 \leq a \leq q \leq P$$

的并集，$P = \ln^B(n)$ 和 $Q = \frac{n}{\ln^{2B} n}$ 为选定常数，我们把 $\mathfrak{M}$ 称为**主要弧** (major arcs)，把 $\mathfrak{m}$ 称为**次要弧** (minor arcs)，这样我们就可以把 $r(n)$ 的估计变成两个部分的估计了。你问我为啥这样搞？除了方便难道会有其它理由吗？所以这些处理本质都是些分析技术。以我个人的观点来看，数论到分析的转化才是“圆法”的关键步骤，或许我给你说了以后，你会觉得“就这”，但是如果你没有相关的经验又怎么能知道这样转化？又怎么知道这样来与分析挂上关系？又怎么进行所谓的解析数论呢？“万事开头难”说的就是这个道理。

比起弱哥德巴赫猜想，我们更渴望证明的应该是哥德巴赫猜想本身，通过类似的思想我们很容易把命题写成

$$\forall n, r(n,N) = \int_0^1 (\sum_{p \leq N} e^{2\pi p i x})^2 e^{-2\pi n i x} dx > 0$$

但目前通过这种形式证明出的最好结论是：**几乎所有**的偶数都可被表为两个奇素数的和。这里的“几乎所有”，实际上就是指不满足的情形几乎可以忽略，我们用 $E(x)$ 表示不超过 x 的不满足哥德巴赫猜想的偶数的个数，则严格表示上述的结论就是

$$\lim_{x \to \infty} \frac{E(x)}{x} = 0, E(x) = o(x)$$

因此这实际上是一个严格的解析结论，哥德巴赫猜想实际也可以写成 $E(x) = 2, x \geq 6$(2,4=2+2 不符合)。使用哥德巴赫猜想的 $r(N)$，通过各种方法，可以与之前一样的估计出

$$r(n,N) = \mathfrak{S}_2(n)\frac{n}{\ln^2 n} + O(\frac{N(\ln\ln N)^3}{\ln^3 N}), \mathfrak{S}_2(N) = \sum_{q=1}^{\infty} \frac{\mu(q)^2}{\varphi(q)^3} c_q(-N) = \frac{N}{\varphi(N)} \prod_{p \nmid N} (1 - \frac{1}{(p-1)^2})$$

读者还需注意这是限制了 $\frac{N}{2} < n \leq N$ 的估计结果，于是可以得到

$$E(x) - E(\frac{x}{2}) = o(\frac{x}{\ln^A x}) \Rightarrow E(x) = o(\frac{x}{\ln^A x}) = o(x)$$

但是，陈景润证明哥德巴赫猜想 1+2 情形使用的是筛法。**筛法** (Sieve Method) 本身，特别是古典的 Eratosthenes 筛法，对于大多数人来说算是很熟悉的了，例如我们要找 n 以下的所有素数，我们只需从 2 到 $\sqrt{n}$ 进行 $k = 2, 3, \ldots$ 倍删数，剩下的就是我们想要的素数了。我们把素因子个数不超过 r 的整数称为 **r-殆素数**，则有

定理 5.13 (陈景润)

每个充分大的偶数都可以表示成一个素数与一个 2-殆素数之和的形式。即每个充分大的偶数都是一个素数与一个不超过两个素数的乘积之和。

我肯定无法讨论怎么证明它，但是我可以稍微介绍一下证明所使用的工具。我们随便搞个整数集 $\mathcal{A} \subset \mathbb{N}$ 和素数集 $\mathcal{P} \subset \mathbb{N}$，定义函数

$$2 \leq w \leq z, P(w,z) = \prod_{\substack{w \leq p < z \\ p \in \mathcal{P}}} p, P(z) = P(2,z)$$

即 $P(z) = \prod_{p<z} p$ 表示所有小于 z 的素数之积，再给出**筛函数**

$$S(\mathcal{A}; \mathcal{P}, z) = \sum_{\substack{a \in \mathcal{A} \\ (a, P(z))=1}} 1$$

即 $S(\mathcal{A};\mathcal{P},z)$ 表示 $\mathcal{A}$ 中没有小于 z 的素因子的元素个数，换句话说就是，我们从 $\mathcal{A}$ 中删去所有小于 z 的素数的倍数所剩下的元素个数 $S(\mathcal{A};\mathcal{P},z)=|\{a\in\mathcal{A}\mid\forall p\in\mathcal{P}\cap(1,z),p\nmid a\}|$，它的一个经典结果是

$$S(\mathcal{A};\mathcal{P},z)=\sum_{a\in\mathcal{A}}\sum_{d|(a,P(z))}\mu(d)=\sum_{d|P(z)}\mu(d)\sum_{d|a\in\mathcal{A}}1=\sum_{d|P(z)}\mu(d)|\mathcal{A}_p|$$

这里的筛子 $\mathcal{P}$ 实际上是可以不断改进的，我们记 $\mathcal{P}_n\subset\mathcal{P}$ 表示所有不能整除 n 的素数，则 $\mathcal{P}_1=\mathcal{P}$ 表示所有的素数、$\mathcal{P}_2=\mathcal{P}-\{2\}$ 表示所有的奇素数。对于 $x\geq y>1,x-y\geq\sqrt{x}$，我们使用 $\mathcal{A}=\{a\mid x-y<a\leq x\}$ 则有

$$\pi(x)-\pi(x-y)=S(\mathcal{A};\mathcal{P},\sqrt{x})+r(x),r(x)=\begin{cases}-1 & \sqrt{x}\text{是素数}\\ 0 & \text{其它}\end{cases}$$

接着我们试着用筛函数来描述我们的猜想。对于任意偶数 N，我们取 $\mathcal{A}=\{N-p\mid p\in\mathcal{P},p\leq N\}$，即用 N 减去所有的素数所生成的整数集，接着我们只需要挑出所有非素数，证明剩下的元素个数大于零，也就是下面的不等式

$$\forall N,S(\mathcal{A};\mathcal{P},\sqrt{N})>0$$

即可证明哥德巴赫猜想。不过这个还没有被证明就是了，这里的筛子和上界具有可变性，如果我们想要它把 r-殆素数给剩下来，就可以将不等式变成

$$\forall N,S(\mathcal{A};\mathcal{P},N^{\frac{1}{r+1}})>0$$

准确来说，它只是放行了一些 r-殆素数，还有一些小于 $N^{\frac{1}{r+1}}$ 的素数生成的殆素数依旧被框住。有了之前的基础，剩下要做的事，我想读者应该心里有数了，就是对筛函数进行估计，当然在实际证明各种情形的时候，对内部的各种参数有不同程度的修改，以便于以后的分析，比如加权筛法就是其中一个重要的变种。筛法的另一个成果就是给出了一个接近孪生素数猜想的定理。

定理 5.14 (陈景润)

存在无穷多个素数 p 使得 p+2 是 2-殆素数。

♡

此时，我们可以选取 $\mathcal{A}=\{a\mid a=n(n-h),1\leq n\leq x\}$，并使用 $Z(x,h)$ 表示，使得 $n\leq x$ 且 n 和 $n-h$ 均为素数的 n 的个数，则孪生素数猜想实际就是说 $\lim_{x\to+\infty}Z(x,2)=+\infty$，而我们的筛函数可以给出它的区域上下界

$$S(\mathcal{A};\mathcal{P},2\sqrt{x})-2\leq Z(x,h)\leq S(\mathcal{A};\mathcal{P},x)+\pi(x)$$

于是我们就将素数间隔函数的估计变成了筛函数的估计，更多内容就留给读者自己去探索了。

孪生素数猜想

所谓**孪生素数猜想**指：存在无穷多个素数 p，使得 p+2 是素数。此时我们把 $(p,p+2)$ 称为一对孪生素数，我们还可以推广为 $(p,p+2k)$，如果我们更进一步，记 $\pi_2(x)=\sum_{p\leq x}(\pi(p+2)-\pi(p))$ 表示不超过 $x+2$ 的孪生素数个数，则孪生素数猜想表述为 $\lim_{x\to+\infty}\pi_2(x)=+\infty$，而 Hardy-Littlewood 第一猜想就是说

$$\pi_2(x)\sim C_2\frac{x}{\ln^2 x},C_2=2\prod_{p>2}(1-\frac{1}{(p-1)^2})$$

这里的 C_2 也称为孪生素数常数，更进一步其实还有 **k 素数组猜想** (k-Tuple Conjecture)，但有些过于遥远就别去想了。对应地有 Hardy-Littlewood 第二猜想

$$\pi(x)+\pi(y)\geq\pi(x+y)$$

但这和孪生素数猜想没啥关系，因为孪生素数猜想在这种情形下应该表述成存在无穷多个素数 p 使得 $\pi(p+2)-\pi(p)=1$，而上面的猜想只能推出一句废话 $\pi(x+2)-\pi(x)\leq\pi(2)=1$。相较于哥德巴赫猜想，孪生素数猜想的研究就相形见绌了，比较重要的就两个，一个是我们上面所讨论的陈景润的结论，但它的素数是不够纯净的殆素数，另一个就是张益唐，我们记 p_n 表示第 n 个素数，则这篇论文 [35] 给出了核心结论

$$\liminf_{n\to\infty}(p_{n+1}-p_n)<7\times10^7$$

下极限指收敛子列极限值的下确界，排除 $p_1=2$ 以后，数列 $p_{n+1}-p_n,n>1$ 的最小值一定是 2，因此孪生素数猜想也能写成

$$\liminf_{n\to\infty}(p_{n+1}-p_n)=2$$

实际上，就算我们不知道孪生素数猜想本身，下极限 $\liminf_{n\to\infty}(p_{n+2}-p_n)$ 的值也表示无穷多对素数的最小间隔。至于证明过程，我想读者应该是可以猜到的，就是找到待估计函数，然后进行大量的不等式估计来推出结论，算是分析的老传统了，而我就来讲一下没啥用的第一步，更详细的过程，就准备好你的不等式和分析能力，然后去读原论文吧，因为那才是证明的核心内容。我们先记一个整数列 $\mathcal{H}_k=\{h_1,...,h_k\}$，使用 $v_p(\mathcal{H})$ 表示 $\mathcal{H}$ 中模 p 剩余类的个数，并且 $\mathcal{H}_k$ 是**容许的** (admissible) 指：$\forall p,v_p(\mathcal{H})<p$，于是有下面的定理。

定理 5.15 (张益唐)

设 $k_0\geq3.5\times10^6$，$\mathcal{H}_{k_0}$ 是容许的。则存在无穷多个整数 n 使得

$$\{n+h_1,n+h_2,...,n+h_{k_0}\}$$

包含至少两个质数。

由简单的事实 $\pi(7\times10^7)-\pi(k_0=3.5\times10^6)>k_0=3.5\times10^6$，我们可以取到 k_0 个区间 $(k_0,7\times10^7)$ 内的素数，由它们构成的 $\mathcal{H}$ 一定是容许的 $(p\leq k_0\Rightarrow v_p(\mathcal{H})\leq p-1<p;p>k_0\Rightarrow v_p(\mathcal{H})\leq k_0<p)$，故我们可以找到无穷多对素数，它们之间的间隔不会超过 $\max\mathcal{H}<7\times10^7$。我们任意给定一个实函数 $\lambda(n)$，并定义

$$\theta(n)=\begin{cases}\ln n & n\text{是素数}\\ 0 & \text{其它}\end{cases}$$

如果我们能证明下面的不等式

$$S_2(x)-(\ln3x)S_1(x)>0,S_1(x)=\sum_{x\leq n\leq2x}\lambda(n)^2,S_2=\sum_{x\leq n\leq2x}(\sum_{i=1}^{k_0}\theta(n+h_i))\lambda(n)^2$$

则说明 $\{n+h_1,n+h_2,...,n+h_{k_0}\}$ 中至少存在两个素数。原理其实很好理解，我们可以通过选取适当的 $\lambda(n)$，来保证 S_1 是个正值，由于不考虑 $p_1=2$ 则 $\ln p\geq\ln3>1$，则 S_2 会由于增加的素数而被放大，于是 $S_2(x)-S_1(x)>0$ 表示至少有一个素数。接着我们需要通过增加系数来调整条件，至于为什么是 $\ln3x$ 可以说明至少有两个素数，需要看张益唐证明源头的两篇文章 [12, 13]，解释起来有点麻烦，我也懒得讲了。最后得到的估计式是

$$S_2(x)-(\ln3x)S_1(x)\geq\mathfrak{S}x(\ln D)^{k_0+2l_0+1}+o(x\mathcal{L}^{k_0+2l_0+1})$$

里面的一大堆常数都没有关注的必要，核心就是 $kx+o(x)$，因此对于充分大的 x 有 $S_2(x)-(\ln3x)S_1(x)>0$。当我们得到这个充分大的 x 时，只需不断地从区间 $[x,2x+\max\mathcal{H}],[2x+\max\mathcal{H},2(2x+\max\mathcal{H})+\max\mathcal{H}],...$ 中取出那至少的两个素数即可获得无穷多对满足定理的素数对，从而完成证明。

整数分拆

我们再来讲一个似乎没什么人知道的**华林问题** (Waring's problem)，它的核心就是研究数论函数 $g(k), k \geq 2$，其表示方程

$$\forall N, x_1^k + ... + x_m^k = N, x_j \geq 0$$

有整数解的最小值 $m = g(k)$。一些简单的结果如下

$$g(2) = 4(\text{四平方和定理}), g(3) = 9(\text{九立方和定理}), g(4) = 19, g(5) = 37$$

根据一些基础知识可以进一步猜想，它的精确表达式为

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

至于证明当然是没有的，不然也不会叫猜想了，但我们可以来估计函数的上下界，在这里 [8] 给出了一个上界估计，再结合显而易见的下界即可得

$$2^k + [(\frac{3}{2})^k] - 2 \leq g(k) \leq (2k+1)^{260(k+3)^{3k+8}}$$

所以关键在于函数上界的估计，哦，我们好像忘记了存在性定理，这有必要吗？连估计都出来了，就没必要去回头了。如何去研究它呢？稍微想想也能明白，就是十分成熟的我们之前所讨论的圆法了，想要继续深造的读者可以看这个 [38] 八百多页的大块头。而我更想讨论的是，与华林问题有那么一点相似的整数分拆问题，它的关键是研究**分拆函数** (partition function)$p(n)$，其表示下面整数拆分的个数

$$n = n_1 + n_2 + ... + n_l, n_1 \geq n_2 \geq ... \geq n_l > 0$$

一些简单的实例如下

$$p(1) = 1, p(2) = 2(2 = 1+1), p(3) = 3(3 = 1+1+1 = 2+1)$$

$$p(4) = 5, 4 = 1+1+1+1 = 2+1+1 = 3+1 = 2+2$$

$$p(5) = 7, 5 = 4+1 = 3+2 = 3+1+1 = 2+2+1 = 2+1+1+1 = 1+1+1+1+1$$

不过这个函数其实没有想象中那么复杂，我们很容易得到它的递归式和母函数

定理 5.16 (基本性质)

$$np(n) = \sum_{l=1}^{n} \sigma(l)p(n-l)$$

$$p(0) = 1, |z| < 1 \Rightarrow \sum_{n=0}^{\infty} p(n)z^n = \prod_{r=1}^{\infty}(1-z^r)^{-1} = P(z)$$

♡

但读者可别想着用圆法来研究这个问题，虽然我们看到了令人兴奋的加法，但是存在性不需要我们来证明，相应的积分估计就没什么意义了，而算出具体的值才是王道，这实际需要利用模形式的相关理论才行。有关模形式，我在以前讲太多了，不知道的可以去回顾这本书 [42]，在十一章的第二节中，可以得到母函数的函数方程为

$$e^{-\frac{2\pi i\tau}{24}} P(e^{2\pi i\tau}) = \eta^{-1}(\tau)$$

这里的 $\eta(\tau)=e^{\frac{2\pi i\tau}{24}}\prod_{r=1}^{\infty}(1-e^{2\pi ir\tau})$ 是狄利克雷 η 函数。既然我们有了强大无比的模形式，进行下面这样的近似

$$\ln p(n)\sim\pi\sqrt{\frac{2}{3}}n^{\frac{1}{2}}$$

就太掉价了，而应该直接就来一个渐进估计

$$p(n)=\frac{1}{4\sqrt{3}}\frac{e^{\lambda\sqrt{m}}}{m}(1+O(\frac{1}{\sqrt{m}})),\lambda=\pi\sqrt{\frac{2}{3}},m=n-\frac{1}{24}$$

至于用来计算的级数展开也是有的

$$p(n)=\sum_{k=1}^{\infty}\sqrt{k}A_k(n)\psi_k(n)$$

$$A_k(n)=\sum_{h=1}^{k}\delta_{gcd(h,k),1}e^{\pi i\sum_{j=1}^{k-1}\frac{j}{k}(\frac{hj}{k}-[\frac{hj}{k}]-\frac{1}{2})-\frac{2\pi ihn}{k}}$$

$$\psi_k(n)=\frac{1}{\pi\sqrt{2}}\frac{d}{dn}(\frac{1}{\sqrt{n-\frac{1}{24}}}\sinh(\frac{\pi}{k}\sqrt{\frac{2}{3}(n-\frac{1}{24})}))$$

充分大与几乎所有

不知读者有没有意识到这节的与众不同之处，没错，就是**充分大**笼罩着几乎所有的命题。如果一个命题不带有充分大几个字，这意味其本身就是一个分析命题，例如素数定理 $\pi(x)\sim\frac{x}{\ln x}$，或者某种性质的数有无穷个所对应到的计数函数 $C(x)$ 满足式 $\lim_{x\to\infty}C(x)=\infty$。另一种情况则是下界被充分优化到了可以计算的程度，弱哥德巴赫猜想就是典型，而哥德巴赫猜想的 1+2 情形证明又说明了下界不一定能得到优化，这篇文章 [34] 给出了这个巨大的下限是 $e^{e^{36}}$。当然所谓的验证并不是说得到一个小的下界就行，比如基于张益唐的理论，对 $\liminf_{n\to\infty}(p_{n+1}-p_n)$ 的上界估计压缩到了 246，但这种存在无穷性质的命题是无法用计算机来验证的。当然分析本身远不止分析那么简单，在这本书 [32] 的第三部分中，我们可以看到使用概率论来解决数论问题的方法，但实际上只是借助了一些概率论的术语和简单工具，本质的过程还是分析学，由这种方法得到的典型结论就是：(Green-Tao) 存在长度为 $k\geq 1$ 的仅由素数构成的等差数列。而它的核心就是证明一个期望估计式

$$\mathbb{E}_{n,r\in[N]}(f(n)f(n+r)...f(n+(k-1)r))\geq c(k,\delta)-o_{k,\delta}(1)$$

而估计的方法就是我们分析的那一套内容了。一个类似的词语是**几乎所有**，我们之前也见过，也是一种分析式的文字描述，比如下面这个定理：(Bohr-Landau) 函数 $\zeta(s)$ 的几乎全部非平凡零点都位于 $\mathrm{Re}(s)=\frac{1}{2}$ 附近。我们可以使用分析的语言把定理写成

$$\forall\sigma>\frac{1}{2},\lim_{T\to\infty}\frac{N(\sigma,T)}{N(T)}=0,N(T)=\{s\mid\zeta(s)=0,0<\mathrm{Re}(s)<1,0<\mathrm{Im}(s)\leq T\},N(\sigma,T)=\{s\in N(T)\mid\mathrm{Re}(s)\geq\sigma\}$$

我们知道古典概率的计算方法就是组合和排列，我们利用的是离散计数本身，因此在数论研究中也可能会用到组合的方法，但这玩意的系统性书籍比较少，在 [11] 或 [24] 的第二部分都有介绍，但算不上什么重点内容，或者说用数论的方法研究组合才比较恰当。既然谈到了计数，那么计算是不是也该来插一脚，例如 [7] 一书，虽然翻译成“计算数论”像一个新方向，但其探讨的无非就是一些数论的相关算法和密码学应用。“无论分析还是组合，计算的确是一个基本功，到处都要用到。”

第六章 实数与逼近

或许你听过各种各样的数论，像组合数论、概率数论、计算数论、几何数论、超越数论、堆垒素数论等等，但名字真的很重要吗？数学本身就是一个各分部之间紧密关联的学科，各种方法理论的互相渗透是见怪不怪的事了。或许有些人喜欢这种界线分明的感觉，比如 MSC[1] 就将数学各方向进行了十分详细地划分，但我个人并不喜欢这种隔阂感，认为“数理逻辑、集合论-序数公理、定理与证明”的划分就已经足够了，或许我无需强求你接受我的观点，但是我写书来传播我的观点也是我写书的初衷。通常情况下，我们基本都是在“定理与证明”这一层中研究数学，而公理基本都是被我们默默接受的东西，以我个人的视角来看理解“定理证明中的依赖关系”才是最重要的。看过我以前写的文章的读者应该能够理解，我是十分注重定义的依赖关系和符号统一性的，而不严谨的内容往往都是放到一段段的说明文字中去，但无论如何都要保证“定义”和“定理”词目中的每一个细节都在前面有过说明。好吧，我的废话好像说太多了，从这一章开始，我们都不会带上“数论”两字，而是以我们研究什么作为标题，代数和分析这两个巨头已经把数论的计算给玩坏了，几何什么的根本就插不了手，或者说它只是来添麻烦的。在这一章中，我们将讨论实数。

6.1 超越数

所谓**无理数**指集合 $\mathbb{R}-\mathbb{Q}$ 内的元素，由于它不具备完全确定的表示还精确的计算，使得无理数的研究在实数的讨论之间占据着主要地位。由实数的严格定义为有理数列极限可知，无理数自身具有分析性质，因此实分析才是无理数的四则运算，而不像有理数那样只有加减乘除。但无理数中还有一部分特殊的存在，它们是整系数一元多项式方程的根，称为**代数数**，由于这一特性导致对于代数数的分析可以稍微摆脱分析的束缚，使用一些初等的代数方法推出一些简单的结论，比如根式的研究，就给人一种十分明晰的感觉。因此真正十分需要分析手段的，当属于超越数的研究了，比如两大常数 π 和 e，因此许多无理数的结论都扎根于此，因为它有这个资本，我们这就来介绍一下这些结论。

几个重要常数

我们直接给出几个重要常数的定义，它们依次是：圆周率、自然底数、欧拉常数

$$\pi=2\int_0^1\frac{dt}{\sqrt{1-t^2}}$$

$$e=\lim_{n\to\infty}(1+\frac{1}{n})^n$$

$$\gamma=\lim_{n\to\infty}((\sum_{k=1}^{n}\frac{1}{k})-\ln n)$$

下面是两个十分基础的结论

定理 6.1 (无理性)

e 和 π 都是无理数。

♡

证明 (1) 我们先来证明比较简单的前一个，假设 $e=\frac{p}{q},(p,q)=1$ 是有理数，我们取 $n>q$ 则有

$$n!e=n!\frac{p}{q}=1\times2\times\ldots\times(q-1)\times(q+1)\times\ldots\times n\times p$$

[1] Mathematics Subject Classification, https://zbmath.org/classification/

为整数。另一方面，根据分析学的基本结论可知

$$e=\sum_{i=0}^{\infty}\frac{1}{i!}=1+\frac{1}{1!}+\frac{1}{2!}+\ldots$$

于是可得

$$n!e=N+R_{n+1},N=n!\sum_{i=0}^{n}\frac{1}{i!}\in\mathbb{N}$$

最后我们来估计剩余部分

$$0<R_{n+1}=n!\sum_{i=n+1}^{\infty}\frac{1}{i!}<\frac{1}{n+1}(\sum_{j=0}^{\infty}\frac{1}{(n+2)^j})=\frac{n+2}{(n+1)^2}<\frac{2}{n+1}\leq 1$$

故又可得 $n!e$ 非整数，矛盾，从而定理得证。

(2) 同样地我们也使用反证法，假设 $\pi=\frac{p}{q},(p,q)=1$ 是有理数，并取一个较大的 n 构造函数

$$f(x)=\frac{x^n(p-qx)^n}{n!}=q^n\frac{x^n(\pi-x)^n}{n!},F(x)=\sum_{i=0}^{n}(-1)^i f^{(2i)}(x)$$

注意到

$$\begin{aligned}
f^{(k)}(0)&=\frac{d^k}{dx^k}(\frac{x^n(p-qx)^n}{n!})|_{x=0}\\
&=\frac{d^k}{dx^k}(\frac{x^n(\sum_{i=0}^{n}C_n^i p^{n-i}(-qx)^i)}{n!})|_{x=0}\\
&=\frac{d^k}{dx^k}(\frac{\sum_{i=0}^{n}C_n^i p^{n-i}(-q)^i x^{n+i}}{n!})|_{x=0}\\
&=\frac{\sum_{i=0}^{n}C_n^i p^{n-i}(-q)^i(n+i)\ldots(n+i+1-k)x^{n+i-k}}{n!}|_{x=0}\\
&=\sum_{i=0}^{n}p^{n-i}(-q)^i x^{n+i-k}\frac{C_n^i(n+i)\ldots(n+i+1-k)}{n!}|_{x=0}\\
&=p^{2n-k}(-q)^{k-n}C_n^{k-n}\frac{k!}{n!}\quad,n+i-k=0\Rightarrow i=k-n,n\leq k\leq 2n
\end{aligned}$$

因此当 $0\leq k<n$ 或 $k>2n$ 时，$f^{(k)}(0)=0$；当 $n\leq k\leq 2n$ 时，$f^{(k)}(0)$ 显然也是一个整数。接着我们利用 $f(\pi-x)=f(x)$，不断求导，即可得

$$(-1)^k f^{(k)}(\pi)=f^{(k)}(0)$$

从而 $f^{(k)}(\pi)$ 也是整数，进而 $F(0)$ 和 $F(\pi)$ 是整数，更进一步有

$$\int_0^{\pi}f(x)\sin x dx=[F'(x)\sin x-F(x)\cos x]_0^{\pi}=F(\pi)+F(0)$$

是整数。另一方面，我们可以估计

$$\int_0^{\pi}f(x)\sin x dx<\int_0^{\pi}q^n\frac{\pi^{2n}}{n!}1dx=\pi\frac{(q\pi^2)^n}{n!}$$

由于 $\lim_{n\to\infty}\frac{a^n}{n!}=0,a>0$，故对于充分大的 n 有

$$0<\int_0^{\pi}f(x)\sin x dx<\pi\frac{(q\pi^2)^n}{n!}<1$$

是非整数，矛盾，从而定理得证。

定理 6.2 (超越性)

e 和 π 都是超越数。

证明 (1) 先从比较简单的前一个开始，假设 e 是代数数，即有整系数方程

$$\sum_{i=0}^{m} a_i e^i = a_0 + a_1 e + ... + a_m e^m = 0, a_0, ..., a_m \in \mathbb{Z}$$

我们选取一个较大的素数 $p > 2$ 构造函数

$$f(x) = \frac{x^{p-1}\prod_{i=1}^{m}(x-i)^p}{(p-1)!}, F(x) = \sum_{i=0}^{mp+p-1} f^{(i)}(x)$$

则有

$$\begin{aligned}
I = \sum_{i=1}^{m} a_i e^i \int_0^i f(x)e^{-x}dx &= \sum_{i=0}^{m} a_i e^i \int_0^i (F(x) - F'(x))e^{-x}dx \\
&= \sum_{i=1}^{m} a_i e^i (\int_0^i F(x)e^{-x}dx - (F(x)e^{-x}|_0^i - \int_0^i F(x)(-e^{-x})dx)) \\
&= \sum_{i=1}^{m} a_i e^i (F(0) - F(i)e^{-i}) \\
&= (0 - a_0)F(0) - \sum_{i=1}^{m} a_i e^i (F(i)e^{-i}) \\
&= -a_0 F(0) - \sum_{i=1}^{m} a_i F(i)
\end{aligned}$$

先考虑 $F(i), i = 1, ..., m$，显然当 $0 \le k < p$ 时，$f^{(k)}(i) = 0$；当 $k \ge p$ 时，我们有

$$\begin{aligned}
f^{(k)}(i) &= \frac{d^k}{dx^k}(\frac{x^{p-1}\prod_{i=1}^{m}(x-i)^p}{(p-1)!})|_{x=i} \\
&= \frac{d^k}{dx^k}(\frac{x^{p-1}\prod_{i=1}^{m}\sum_{j=0}^{p} C_p^j x^{p-j}(-i)^j}{(p-1)!})|_{x=i} \\
&= \frac{d^k}{dx^k}(\frac{\sum_{i=p-1}^{mp+p-1} n_i x^i}{(p-1)!})|_{x=i}, n_k \in \mathbb{N} \\
&= \frac{\sum_{j=p-1}^{mp+p-1} n_j j(j-1)...(j+1-k)x^{j-k}}{(p-1)!}|_{x=i} \\
&= \sum_{j=p-1}^{mp+p-1} n_j \frac{j(j-1)...(j+1-k)}{(p-1)!} i^{j-k} \\
&= \sum_{j=p-1}^{mp+p-1} n_j i^{j-k} C_j^k \frac{k!}{(p-1)!}
\end{aligned}$$

从而 $p \mid f^{(k)}(i)$ 是整数，进而 $p \mid F(i)$ 是整数。再考虑 $F(0)$，显然当 $0 \le k < p-1$ 时，$f^{(k)}(0) = 0$；当 $k = p-1$ 时，其相当于泰勒展开的第一个系数，从而有

$$f^{(p)}(0) = \prod_{i=1}^{m}(0-i)^p = (-1)^{pm}(m!)^p$$

选取 $p > m$ 则 $p \nmid f^{(p)}(0)$ 是整数；当 $k > p-1$ 时

$$\begin{aligned}
f^{(k)}(0) &= \sum_{j=p-1}^{mp+p-1} n_j 0^{j-k} C_j^k \frac{k!}{(p-1)!} \\
&= n_k C_k^k \frac{k!}{(p-1)!} \quad , j-k = 0 \Rightarrow j = k
\end{aligned}$$

从而有 $p \mid f^{(k)}(0)$ 是整数，进而 $p \nmid F(0)$ 是整数，我们进一步选取 $p > \max\{m, |a_0|\}$，则 $p \nmid I$ 是非零整数 $(p \mid 0)$。另一方面，我们可以估计

$$\begin{aligned}
|I| \leq \sum_{i=1}^{m} |a_i| e^m \left| \int_0^i f(x) e^{-x} dx \right| &\leq e^m \sum_{i=1}^{m} |a_i| \int_0^i |f(x) e^{-x}| dx \\
&\leq e^m \sum_{i=1}^{m} |a_i| \int_0^i \frac{m^{p-1}(m^p)^m}{(p-1)!} e^{-x} dx \\
&\leq e^m \sum_{i=1}^{m} |a_i| \frac{m^{mp+p-1}}{(p-1)!} \quad , \int_0^i e^{-x} dx = [-e^{-x}]_0^i = 1 - \frac{1}{e^i} < 1 \\
&= \frac{m^{(m+1)p-1}}{(p-1)!} e^m \sum_{i=1}^{m} |a_i|
\end{aligned}$$

由于 $\lim_{p\to\infty} \frac{m^{(m+1)p-1}}{(p-1)!} = 0, m > 0$，故对于充分大的素数 p 有

$$0 < |I| \leq \frac{m^{(m+1)p-1}}{(p-1)!} e^m \sum_{i=1}^{m} |a_i| < 1$$

是非整数，矛盾，从而定理得证。

(2) 我们之所以难处理 π，主要因为其没有优秀的解析式，此时我们只能借助欧拉公式 $e^{i\pi} = -1$ 来辅助证明。假设 π 是代数数，则 $\theta = i\pi$ 也是代数数 (封闭性)，故有整系数方程

$$\sum_{i=0}^{m} \alpha_i \theta^i = \alpha_0 + \alpha_1 \theta + \ldots + \alpha_m \theta^m = 0, \alpha_0, \ldots, \alpha_m \in \mathbb{Z}$$

设它的所有根为 $\theta = \theta_1, \ldots, \theta_m$，则里面至少有一个 $i\pi$，故有等式

$$\prod_{i=1}^{m} (1 + e^{\theta_i}) = (1 + e^{\theta_1}) \ldots (1 + e^{\theta_m}) = 0$$

我们把它展开，可以得到这样的形式

$$q + e^{a_1} + \ldots + e^{a_n} = 0, q = 2^m - n \in \mathbb{N}, a_i = \sum_{i=1}^{m} \varepsilon_i \theta_i \neq 0, \varepsilon_i = 0, 1$$

我们选取一个较大的素数 $p > 2$ 构造函数

$$f(x) = \frac{x^{p-1} \prod_{i=1}^{n} (l(x - a_i))^p}{(p-1)!}, F(x) = \sum_{i=0}^{np+p-1} f^{(i)}(x)$$

注意到在 $f(x)$ 中我们添加了一个调整系数 l，实际上由对称多项式的理论可知 $\prod_{i=1}^{n} (x - a_i) \in \mathbb{Q}[x]$，因此我们取 l 是展开中所有系数分母的最小公倍数，从而使得 $\prod_{i=1}^{n} l(x - a_i) \in \mathbb{Z}[x]$ 成立。此时有

$$\begin{aligned}
J = \sum_{i=1}^{n} e^{a_i} \int_0^{a_i} f(x) e^{-x} dx &= \sum_{i=1}^{n} e^{a_i} \int_0^{a_i} (F(x) - F'(x)) e^{-x} dx \\
&= \sum_{i=1}^{n} e^{a_i} (F(0) - F(a_i) e^{-a_i}) \\
&= (0 - q) F(0) - \sum_{i=1}^{n} e^{a_i} F(a_i) e^{-a_i} \\
&= -qF(0) - \sum_{i=1}^{n} F(a_i)
\end{aligned}$$

我们不做多余的论证，由 (1) 类似方法可得，当 $p > 2$ 足够大时，$p \nmid F(0)$ 是整数，$p \mid F(a_i)$ 是整数，从而 $p \nmid J$

是非零整数。另一方面，我们可以估计

$$\begin{aligned}|J| \leq \sum_{i=1}^{n}|e^{a_i}\int_0^{a_i} f(x)e^{-x}dx| &= \sum_{i=1}^{n}|e^{a_i}||\int_0^{a_i}\frac{l^{np}x^{p-1}\prod_{i=1}^{n}(x-a_i)^p}{(p-1)!}e^{-x}dx| \\ &\leq \sum_{i=1}^{n}|e^{a_i}|\frac{l^{np}|a_i|^{np+p-1}}{(p-1)!}|a_i| \\ &\leq \sum_{i=1}^{n}|e^{a_i}|\frac{(l^n|a_i|^{n+1})^p}{(p-1)!}\end{aligned}$$

读者需要注意，虽然 a_i 不一定是实的，但我们所给的函数是全纯的，因此复积分与路径无关，可以像实积分那样运算。这里的 n 是个可控的有限数，由于 $\lim_{p\to\infty}\frac{a^p}{(p-1)!}=0, a>0$，故对于充分大的素数 p 有

$$0<|J|\leq\sum_{i=1}^{n}|e^{a_i}|\frac{(l^n|a_i|^{n+1})^p}{(p-1)!}<1$$

是非整数，矛盾，从而定理得证。

虽然我们不能证明 γ 的无理性和超越性，但我们可以得到一些好康的公式。

定理 6.3 (欧拉常数 [29])

(1) $\gamma=-\Gamma'(1)$

(2) $\gamma=1-\ln\frac{3}{2}-\sum_{k=1}^{\infty}\frac{\zeta(2k+1)-1}{4^k(2k+1)}$

(3) $\gamma=-\int_0^1\ln\ln\frac{1}{t}dt$

(4) $e^{\gamma}=\lim_{n\to\infty}\frac{1}{\ln x}\prod_{p\leq n}(1-\frac{1}{p})^{-1}$

(5) $\gamma=\sum_{k=1}^{n}\frac{1}{k}-\ln n+r_n, r_n=-\frac{1}{2n}+\sum_{k=1}^{r}\frac{B_{2k}}{(2k)n^{2k}}+\theta\frac{B_{2r+2}}{(2r+2)n^{2r+2}}, \theta\in(0,1), \frac{z}{e^z-1}=\sum_{n=0}^{\infty}B_n\frac{z^n}{n!}$

♡

这些都是典型的分析结论，就当作习题给读者练习分析能力吧。

超越数判定

有了上面的基础，我们实际上能证明四个更广泛的超越数判定结论。

定理 6.4

(1) 若 r 是 n 次实代数数，则存在常数 $C>0$ 使得 $|r-\frac{p}{q}|\geq\frac{C}{q^n}$ 对任何整数对 $p,q(q>0)$ 成立。

(2) 若 $b_1,\ldots,b_n$ 为互不相等的代数数，$a_1,\ldots,a_n$ 是不全为零代数数，则 $\sum_{i=1}^{n}a_ie^{b_i}\neq 0$。

(3) 若 α,β 为代数数，$\alpha\neq 0,1$ 且 β 不是实有理数，则 α^{β} 是超越数。

(4) 若 $b_1,\ldots,b_n$ 是非零代数数，$\ln b_1,\ldots,\ln b_n$ 在有理数域上线性无关，则对任意不全为零的代数数 $a_0,a_1,\ldots,a_n$ 有 $a_0+\sum_{i=1}^{n}a_i\ln b_i\neq 0$。

♡

注 第一个结论是刘维尔构造超越数的基础定理，由此可以很容易推出 $\sum_{n=1}^{\infty}\lambda^{-n!}, \lambda>1$ 不是实代数数，从而是实超越数。第二个结论是 Lindemann-Weierstrass 定理，借助它、e 是超越数和欧拉公式 $e^{i\pi}=-1$ 就能直接得到 π 是超越数。第三个结论是 Gelfond-Schneider 定理，隶属于希尔伯特第七问，可以容易地推出 $2^{\sqrt{2}}$ 和 $\sqrt{2}^{\sqrt{2}}$ 之类的是超越数。第四个结论是 Baker 定理，虽然好像不太为人所知，但它是上一个结论的推广，由 $\beta\ln\alpha-\ln\alpha^{\beta}=0$ 可以直接推出第三个结论。

证明 [Liouville] 由 r 是 n 次实代数数，我们可以设它的极小多项式为

$$f(x)=\sum_{i=0}^{n}a_ix^i, f(r)=0, a_i\in\mathbb{N}$$

我们取区间 $J=\{x\mid |x-r|\leq 1\}$ 和常数 $A=\max_{x\in J}|f'(x)|$。对任意的有理数 $\frac{p}{q}$，如果 $\frac{p}{q}\notin J$，我们很容易找到有理数 $\frac{h}{k}\in J$ 使得 $|r-\frac{p}{q}|>|r-\frac{h}{k}|$(稠密性)，故我们假设 $\frac{p}{q}\in J$。由微分中值定理可得

$$f(r)-f(\frac{p}{q})=f'(\varepsilon)(r-\frac{p}{q}), \varepsilon\in J\Rightarrow |f'(\varepsilon)|\leq A$$

取绝对值并变换上式可得

$$|r-\frac{p}{q}|=|\frac{f(r)-f(\frac{p}{q})}{f'(\varepsilon)}|\geq\frac{|f(\frac{p}{q})|}{A}=\frac{|\sum_{i=0}^{n}a_ip^iq^{n-i}|}{Aq^n}\geq\frac{1}{Aq^n}=\frac{C}{q^n}, C=\frac{1}{A}$$

剩下三个定理的证明可以参考这本书 [36] 的第四、五、六章，一个定理占一个章节，写到这里来就太占篇幅了，同样还是那句话，留给感兴趣的读者自行探索。

数的分类

有关超越数的研究，其实基本都是在用各种手段来证明某些数的无理性或超越性，比如 $\ln n, \zeta(2n+1)$ 等等，在 [40] 一书中，就探讨了椭圆函数 $\mathscr{C}$ 和模形式 j 的相关超越性。而稍有些理论雏形的是，超越性度量理论，其不仅可以用来评价数到底有多超越，还对复数进行了不相交的分类。对于多项式 $f(x)=\sum_{i=0}^{n}a_ix^i\in\mathbb{Z}[x]$，我们定义它的**高**、**长度**和**次数**分别为

$$h(f)=\max_{0\leq i\leq n}|a_i|, L(f)=\sum_{i=0}^{n}|a_i|, \deg f=n$$

接着再引入 $\mathbb{Z}[x]$ 的一个子集

$$\mathfrak{M}(n,h)=\{f\in\mathbb{Z}[x]\mid \deg f\leq n, h(f)\leq h\}$$

对任意一个超越数 r，如果存在一个正值函数 $f(x,y)$ 使得

$$\forall n,h\in\mathbb{N}, P(x)\in\mathfrak{M}(n,h), |P(r)|>f(n,h)$$

则称 $f(x,y)$ 是 r 的**超越性度量**。比如对于代数数 $a\neq 0,\ln a$ 的一个超越性度量为

$$f(n,h)=e^{-Cn^2(\ln h+d\ln d)(1+\ln d)^{-1}}$$

类似地，e^a 的一个超越性度量为

$$f(n,h)=e^{-Cn^2(\ln dh)(\ln(d\ln h))^2(\ln\ln h+\ln\ln d)^{-2}}$$

不过这些估计都是在一个足够大的 h 下成立的，而最简单的估计应该是

$$f(h,h)\leq\begin{cases}e^{c_1n}h^{-n} & r\in\mathbb{R}\\ e^{c_2n}h^{-\frac{1}{2}(n+1)} & r\notin\mathbb{R}\end{cases}$$

其中 c_1,c_2 是只与 r 有关的常数。如果我们再引入符号

$$M(n,h)=\{r\in\mathbb{C}\mid f(r)=0, f\in\mathbb{Z}[x], \deg f\leq n, h(f)\leq h\}$$

对任意一个超越数 r，如果存在一个正值函数 $g(x,y)>0$ 使得

$$\forall n,h\in\mathbb{N},\min_{a\in M(n,h)}|a-r|>g(n,h)$$

则称 $g(x,y)$ 是 r 的**逼近度函数**。通常情况下它们的联系是简单的

$e^{-g(n,h)}$是逼近度函数，则$e^{-g(n,h)-3n\ln(n+1)-n\ln h}$是超越性度量。

$e^{-f(n,h)}$是超越性度量，则$e^{-f(n,h)-4n-n\ln h}$是逼近度函数。

首先，我们需要给出一个超越数的必要条件，即代数数的充分条件

定理 6.5

设 r 是超越数，$P_i(x)$ 是次数为 n_i 高为 h_i 的整系数多项式，且存在 C_1,C_2 使得 $n_i<n_{i+1}\le C_1n_i,\ln h_i<\ln h_{i+1}\le C_2\ln h_{i+1}$，则存在无穷多个 i 和常数 C_3 使得 $\ln|P_i(r)|\ge -C_3n_i(n_i+\ln h_i)$。

接着引入一些符号

$$\omega_n(h,r)=\min_{\substack{P(z)\in\mathfrak{M}(n,h)\\P(r)\neq0}}|P(r)|$$

$$\omega(r)=\limsup_{n\to\infty}\frac{\omega_n(r)}{n}=\limsup_{n\to\infty}\limsup_{h\to\infty}(-\frac{\ln\omega_n(h,r)}{n\ln h})$$

$$\nu(r)=\min_{\omega_n(r)=\infty}n$$

最后就可以进行分类了

$$\text{A 数：}\omega(r)=0,\nu(r)=\infty$$

$$\text{S 数：}0<\omega(r)<\infty,\nu(r)=\infty$$

$$\text{T 数：}\omega(r)=\infty,\nu(r)=\infty$$

$$\text{U 数：}\omega(r)=\infty,\nu(r)<\infty$$

其中最简单的 A 数就是与超越数相对应的代数数

定理 6.6

A 数 $=\mathbb{A}=\overline{\mathbb{Q}}$

接着我们可以讨论这四类数的占比问题，$\mathbb{A}$ 是可数的，$\mathbb{C}$ 是连续的，我们以 $\mathbb{C}=\mathbb{R}^2$ 的观点可以得到复数上的勒贝格测度，由此 A 数构成零测度集，同样地我们还能证明：T 数和 U 数也是零测度集，因此几乎所有的复数都是 S 数。所以我们需要给出一个 S 数判定的充要条件。

定理 6.7

(1)r 是 S 数当且仅当，存在 $\theta_0>0$ 使得 $\forall\varepsilon>0\exists c_n,\omega_n(h,r)>c_nh^{-(\theta_0+\varepsilon)n},n=1,2,...,h=1,2,...$

(2) 我们把 (1) 中 $\theta_0+\varepsilon$ 的下界称为 r 的型，即 $\theta=\inf\{\theta_1\mid\forall n\exists c_n,\omega_n(h,r)>c_nh^{-\theta_1n},h=1,2,...\}$，则有 $\theta=\sup_n\frac{\omega_n(r)}{n}$。

我们之前所讨论的刘维尔数 $\sum_{n=1}^{\infty}\lambda^{-n!},\lambda>1$ 实际就是 $\nu=1$ 的 U 数，e 是型为 1 的 S 数，实际上这种数的分类有我们之前所讨论的一个至关重要的性质。

定理 6.8

代数相关的数属于同一类数，即如果存在不为零的代数数 $a_1,...,a_n$ 使得 $a_1b_1+...+a_nb_n=0$，则 $b_1,...,b_n$ 是同一类数。

♡

6.2 逼近原理

近似三宝

我们先来给出一个最基础的逼近原理，它是我们能讨论逼近问题的基础。

定理 6.9 (Dirichlet)

对任意实数 r,Q 且 $Q>1$，存在整数 p,q 满足 $1\le q<Q, |r-\frac{p}{q}|\le\frac{1}{Qq}$。

♡

证明 假设 Q 是整数，考虑下面区间 $[0,1]$ 上的 $Q+1$ 个实数

$$0,\{r\},\{2r\},...,\{(Q-1)r\},1$$

这里的 $\{r\}=r-[r]$ 表示取小数部分。接着再把 $[0,1]$ 划分成 Q 个区间

$$[0,\frac{1}{Q}),[\frac{2}{Q},\frac{3}{Q}),...,[\frac{Q-1}{Q},1]$$

由抽屉原理可知，存在整数 $0\le r_1\ne r_2\le Q-1$ 使得

$$|\{r_1r\}-\{r_2r\}|\le\frac{1}{Q}\Rightarrow|(r_1r-[r_1r])-(r_2r-[r_2r])|\le\frac{1}{Q}$$

此时我们取整数 $q=r_1-r_2, p=[r_1r]-[r_2r]$ 即可得

$$|r-\frac{p}{q}|=|r-\frac{[r_1r]-[r_2r]}{r_1-r_2}|=|\frac{(r_1-r_2)r-([r_1r]-[r_2r])}{q}|\le\frac{1}{Qq}$$

若 Q 非整数，则令 $Q'=[Q]+1>Q$，同上可得

$$|r-\frac{p}{q}|\le\frac{1}{Q'q}<\frac{1}{Qq}$$

这个定理虽然简单，但它却告诉了我们任何实数都可以以任意精度被有理数逼近，实际上，一个简单的逼近例子，就是拿出实数的十进制小数，不断截取更高的位数就是一种逼近。为什么我们会找有理数逼近呢？因为有理数在工程实践中能够计算，比较有意义，基本各种自然科学都喜欢取多少位小数点的近似，在计算机中，能运算的也确实是有理数，或者说有限小数，典型实例就是 js 里面的 $0.1+0.2\ne0.3$，不过对于像我这样的数学人来说，有理数是放在手里最安心且最能把控的数了。在上面定理的基础上，我们还可以讨论逼近有理数的个数是否无穷，即有

定理 6.10 (Hurwitz)

对每个无理数 r，存在无穷多个不同的有理数 $\frac{p}{q}$ 满足 $|r-\frac{p}{q}|<\frac{1}{\sqrt{5}q^2}$。并且常数 $\sqrt{5}$ 是最好的，换言之，如果 $A>\sqrt{5}$，则存在无理数 r 使得 $|r-\frac{p}{q}|<\frac{1}{Aq^2}$ 只有有限多个有理数解 $\frac{p}{q}$。

♡

证明 (1) 不失一般性，我们假设 $0<r<1$，首先我们要试图找到 $(0,1)$ 间的所有既约分数，这时我们需要用到 Farey 数列。读者可以轻松地验证，如果既约分数 $\frac{p}{q},\frac{p'}{q'}$ 满足 $qp'-pq'=\pm1$ 则 $\frac{p+p'}{q+q'}$ 也是既约分数，并且由 $0=\frac{0}{1},1=\frac{1}{1}$ 出发通过这种和内插可以生成 $(0,1)$ 间的所有既约分数。我们把分母不超过 n 的 $[0,1]$ 间的所有分数按从小到大排成的序列称为 **n 级 Farey** 数列，记作 $\mathscr{F}_n$，在这个数列中，任一数都是左右相邻两项的和内插。工具齐全，开始正戏。

对任意的 $\mathscr{F}_n$，我们取满足 $\frac{a}{b}<r<\frac{a'}{b'}$ 的最靠近的两个元素，并计算和内插 $\frac{a}{b}<\frac{a^*}{b^*}=\frac{a+a'}{b+b'}<\frac{a'}{b'}$，于是下面

的三个不等式至少有一个成立

$$|r-\frac{a}{b}|<\frac{1}{\sqrt{5}b^2},|r-\frac{a'}{b'}|<\frac{1}{\sqrt{5}(b')^2},|r-\frac{a^*}{b^*}|<\frac{1}{\sqrt{5}(b^*)^2}$$

若均不成立，则有

$$r-\frac{a}{b}\geq\frac{1}{\sqrt{5}b^2},r-\frac{a^*}{b^*}\geq\frac{1}{\sqrt{5}(b^*)^2},\frac{a'}{b'}-r\geq\frac{1}{\sqrt{5}(b')^2},r>\frac{a^*}{b^*}$$

$$\text{或}r-\frac{a}{b}\geq\frac{1}{\sqrt{5}b^2},\frac{a^*}{b^*}-r\geq\frac{1}{\sqrt{5}(b^*)^2},\frac{a'}{b'}-r\geq\frac{1}{\sqrt{5}(b')^2},r\leq\frac{a^*}{b^*}$$

对于后一种情况，我们只需进行 $x\to 1-x$ 的对称替换即可转化为前一种情形。此时，考虑上面三个不等式的矛盾，通过 1+3 和 2+3 可得

$$\begin{cases}\frac{a'}{b'}-\frac{a}{b}= & \frac{1}{b'b}\geq\frac{1}{\sqrt{5}}(\frac{1}{b^2}+\frac{1}{(b')^2})\\ \frac{a'}{b'}-\frac{a^*}{b^*}= & \frac{1}{b^*b}\geq\frac{1}{\sqrt{5}}(\frac{1}{b^2}+\frac{1}{(b^*)^2})\end{cases}\Rightarrow\begin{cases}\sqrt{5}b'b\geq b^2+(b')^2\\ \sqrt{5}b'b\geq b^2+(b^*)^2\end{cases}$$

进一步相加可得

$$\sqrt{5}b'(b+b^*)\leq b^2+2(b')^2+(b^*)^2\Rightarrow\sqrt{5}b'(2b+b')\leq 2b^2+3(b')^2+2b'b$$

化简配方可得

$$(2b-(\sqrt{5}-1)b')^2\leq 0$$

由于 b,b' 为非零整数，从而矛盾，进而我们可以得到一个不等式

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

(2) 由 Farey 数列的定义可知 $2\leq b\leq n$，从而有

$$n\leq nb-b^2+b=(n-b+1)b\leq b'b,ba'-ab'=1$$

$$|\frac{a}{b}-\frac{a'}{b'}|=|\frac{ba'-ab'}{b'b}|=\frac{1}{b'b}\leq\frac{1}{n}$$

由我们之前的假设 $\frac{a}{b}<\frac{a^*}{b^*}<r<\frac{a'}{b'}$ 可知

$$|r-\frac{p}{q}|<|\frac{a}{b}-\frac{a'}{b'}|\leq\frac{1}{n}$$

接着我们只需从 $n_1=2$ 开始，得到第一个 $\frac{p_1}{q_1}$，然后找到一个使得上述不等式能成立的最大 n_2，再得到一个新的 $\frac{p_2}{q_2}$，如此往复即可得到无穷多个 $\frac{p}{q}$。

(3) 容易验证 $r=\frac{\sqrt{5}-1}{2}$ 就是会使得偏移无法成立的无理数。

目前，最常见的有理逼近有三种方式，一种是**连分数**的渐进分式

$$r=x_0+\frac{1}{x_1+\frac{1}{x_2+\frac{1}{x_3+\frac{1}{x_4+\cdots}}}},\frac{p_n}{q_n}=x_0+\frac{1}{x_1+\frac{1}{\ddots_{\frac{1}{x_{n-1}+\frac{1}{x_n}}}}}$$

我们用些简记的符号是 $\frac{p_n}{q_n}=[x_0;x_1,...,x_n]$，并把它称为 r 的**第 n 个渐进分式**，其有着如下的递推关系和共轭关系

$$p_0=x_0,p_1=x_0x_1+1,p_n=x_np_{n-1}+p_{n-2}\ ;\ q_0=1,q_1=x_1,q_n=x_nq_{n-1}+q_{n-2}$$

$$q_np_{n-1}-p_nq_{n-1}=(-1)^n\ ;\ q_{n+1}p_{n-1}-p_{n+1}q_{n-1}=(-1)^nx_{n+1}$$

当然我们还关系连分数的逼近效果如何，这个只需算出相应的误差

$$r-\frac{p_n}{q_n}=\frac{(-1)^n}{q_n(r_{n+1}q_n+q_{n-1})},r_{n+1}=[x_{n+1};x_{n+2},x_{n+3},...]$$

即可得到

$$|r-\frac{p_n}{q_n}|<\frac{1}{q_n^2}$$

如果我们进一步压缩精度，则有下面的三个定理。

定理 6.11

(1)[Vahlen] 两个相邻渐进分式中至少有一个 $\frac{p}{q}$ 满足 $|r-\frac{p}{q}|<\frac{1}{2q^2}$

(2)[Borel] 三个相邻渐进分式中至少有一个 $\frac{p}{q}$ 满足 $|r-\frac{p}{q}|<\frac{1}{\sqrt{5}q^2}$

(3)[最佳逼近] 对于 $n>0$，如果 $0<q\leq q_n, \frac{p}{q}\neq\frac{p_n}{q_n}$，则 $|r-\frac{p_n}{q_n}|<|r-\frac{p}{q}|$。

♡

如果我们想要用渐进分式来逼近无理数的话，前题是我们必需完全了解它的连分数构造，但除了二次无理数以外的连分数都没有周期规律，因此这种逼近其实用处并不大。而剩下的两种，自然是数学分析中无处不在的**无穷和** (也称为**无穷级数**或**函数项级数**) 和**无穷积**的部分和 (积) 了

$$\sum_{i=1}^{\infty}x_i=\lim_{n\to\infty}S_n, S_n=\sum_{i=1}^{n}x_i$$

$$\prod_{i=1}^{\infty}p_i=\lim_{n\to\infty}P_n, P_n=\prod_{i=1}^{n}p_i$$

最常见的无穷和就是泰勒级数或洛朗级数

$$e^x=\sum_{i=0}^{\infty}\frac{x^n}{n!}$$

$$\ln(1+x)=\sum_{i=1}^{\infty}\frac{(-1)^{n+1}x^n}{n}, |x|<1$$

$$J(z)=\frac{1}{12^3z}+\sum_{n=0}^{\infty}c(n)z^n$$

$$\cot z=\frac{1}{z}-\sum_{n=1}^{\infty}\frac{2^{2n}B_{2n}}{(2n)!}z^{2n-1}$$

其实还有一种工程上常用的无穷和，即有理分式展开

$$\frac{1}{\sin^2 z}=\frac{1}{z^2}+\sum_{n=1}^{\infty}(\frac{1}{(z-n\pi)^2}+\frac{1}{(z+n\pi)^2})$$

$$\frac{1}{\sin\pi x}=\frac{1}{\pi x}+\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}x}{n^2-x^2}$$

实际上我们有一个 Mittag-Leffler 定理来确保亚纯函数的有理分式展开。更深入时，还有些奇奇怪怪的无穷和展开

$$\sin(ax)=\frac{\sin(a\pi)}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n 2n}{a^2-n^2}\sin(nx), \cos(ax)=\frac{\sin(a\pi)}{a\pi}+\frac{\sin(a\pi)}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n 2a}{a^2-n^2}\cos(nx)$$

$$\cot(z)=\frac{1}{z}-\sum_{n=1}^{\infty}\frac{1}{2^n}\tan(\frac{z}{2^n})$$

$$\sum_{n=1}^{\infty}\frac{n^5}{e^{2\pi n}-1}=\frac{1}{504},\sum_{n=1}^{\infty}\frac{n^{13}}{e^{2\pi n}-1}=\frac{1}{24}$$

$$\sum_{n=1}^{\infty}\frac{\zeta(2n)}{(2n+1)2^{2n}}=\frac{1}{2}-\frac{1}{2}\ln 2$$

但由于它们又杂又乱又多，就，就不探究太多了。而无穷乘积常见于以下的几种地方，解析数论中的欧拉积

$$\zeta(s)=\prod_p(1-p^{-s})^{-1},L(s,\chi)=\prod_p(1-\chi(p)p^{-s})^{-1}$$

模形式中的 q 级数

$$\eta(\tau)=q^{\frac{1}{24}}\prod_{n=1}^{\infty}(1-q^n),\Delta(\tau)=q\prod_{r=1}^{\infty}(1-q^n)^{24},q=e^{2\pi i\tau}$$

复分析中的 Weierstrass 分解定理

$$\frac{\sin z}{z}=\prod_{n=1}^{\infty}(1-\frac{z^2}{n^2\pi^2}),\frac{1}{\Gamma(z)}=e^{\gamma z}z\prod_{n=1}^{\infty}(1+\frac{z}{n})e^{-\frac{z}{n}}$$

剩下的就是各种推出的零散公式

$$\frac{\pi}{2}=\prod_{n=1}^{\infty}\frac{4n^2}{4n^2-1}$$

$$\cdots$$

如果想要了解更多的无穷和或无穷积，可以翻阅有关“特殊函数”的书籍，保证你大饱眼福。

有理逼近

接着我们来探究代数数 $\mathbb{A}$ 的有理逼近，选取的原因不言而喻了，因为超越数不好研究，而且代数数自带一个参数 $n\geq 2$(极小多项式的次数) 具有代数性可以少用分析、并且还自带着一个估计，即

$$\forall r\in\mathbb{A},\varepsilon>0,|r-\frac{p}{q}|<\frac{1}{q^{2+\varepsilon}}$$

只有有限个有理数解 $\frac{p}{q}$，这就是 **Roth 定理**，是前面刘维尔代数数性质的推广。但我们这一部分的核心内容是它的推广，即**子空间定理** (Subspace Theorem)

定理 6.12 (Schmidt)

设 $a_1,\ldots,a_n$ 是代数数且满足 $1,a_1,\ldots,a_n$ 在 $\mathbb{Q}$ 上线性无关，则对任意的 $\varepsilon>0$ 只存在有限多个整数 $q,p_1,\ldots,p_n$ 满足

$$q^{1+\varepsilon}\prod_{i=1}^{n}(qa_i-p_i)<1$$

♡

显然我们只要在 Schmidt 子空间定理中取 $n=1$ 就能得到 Roth 定理了。由于这个定理证明太长了，还附带了不少一次性工具，所以我们就单纯地简述一下证明过程，并给出几个核心踏脚点。对于一个向量 $x_i\in\mathbb{R}^l$，我们把它的分量形式记为 $x_i=(x_{i1},x_{i2},\ldots,x_{il})$，并把它的一元一次齐次式 $L_i(x_i)=\sum_{j=1}^{l}a_{ij}x_{ij}$ 称为一个线性型，给定一组 $r_1,\ldots,r_m\in\mathbb{N}$、线性型 $L_1(x_1),\ldots,L_m(x_m)$ 和 ml 元多项式 $P(\{x_{ij}\})\in\mathbb{R}[\{x_{ij}\}]$，我们定义 Roth-Schmidt 指标为

$$P=\sum_j b(j)L_1^{j_{11}}...L_m^{j_{m1}}\prod x, \mathrm{Ind}P(L_1,...,L_m;r_1,...,r_m)=\min\{\frac{j_{11}}{r_1}+...+\frac{j_{m1}}{r_m}\mid b(j)\neq 0\}$$

接着是至关重要的的 Roth 引理，它是在一大堆条件的基础上给出一个指标的估计式

$$m\geq 1,0<C\leq 1,0<\varepsilon<\frac{1}{12},\omega(m,\varepsilon)=24\times 2^{-m}(\frac{\varepsilon}{12})^{2^{m-1}},r_h,p_h,q_h\in\mathbb{N},h=1,...,m$$

$$\omega r_h\geq r_{h+1},q_h>0,(p_h,q_h)=1,q_h^{r_h}\geq q_1^{Cr_1},q^{\omega C}\geq 2^{3m},0\neq P\in\mathbb{Z}[x_{ij}],\deg_{x_h}P\leq r_h,|P|\leq q_1^{\omega r_1 C}$$

$$\text{以上}\Rightarrow R-\mathrm{Ind}P(\frac{p_1}{q_1},...,\frac{p_m}{q_m};r_1,...,r_m)<\varepsilon$$

此时我们需要先证明 Roth 定理，它是证明 Schmidt 定理的必经路径。首先我们可以假设 $r\in\overline{\mathbb{Q}}$ 是代数整数 $r\in\overline{\mathbb{Z}}$，否则我们可以找到一个整数 $a\in\mathbb{Z}$ 使得 $ar\in\overline{\mathbb{Z}}$ 是代数整数，如果有无穷多个有理数 $\frac{p}{q}$ 满足

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

那么就会有无穷多个 $\frac{p'}{q}=\frac{ap}{q}$ 满足

$$|ar-\frac{p'}{q}|=|ar-\frac{ap}{q}|<\frac{a}{q^{2+\varepsilon}}<\frac{1}{q^{2+\frac{\varepsilon}{2}}}$$

换言之我们要证明一个稍强的命题。在 r 是代数整数的条件下，接着我们使用反证法，假设有无穷多个 $\frac{p}{q}$ 满足不等式，并取一个充分小的 δ 再去靠近上面的一大堆条件，如下

$$L_1(x,y)=x-ry,L_2(x,y)=y,0<\delta<\frac{1}{12},0<\varepsilon<\frac{\delta}{20},m\leq\varepsilon^{-2}\ln(4\deg r),\omega=24.2^{-m}(\frac{\delta}{12})^{2^{m-1}}$$

结合解的无穷假定，我们可以取出 $\frac{p_1}{q_1},...,\frac{p_m}{q_m}$ 满足下面条件

$$q_1\leq q_2\leq...\leq q_m,q_h^{\omega}\geq 2^{5m},q_{h+1}^{\frac{\omega}{2}}>q_h,q_h^{\varepsilon}>64(D+1)\max(1,|r|),q_1^{\omega}>D^m$$

再引入 $r_1\geq\frac{\ln q_m}{\varepsilon\ln q_1},r_h=[\frac{r_1\ln q_1}{q_h}]+1$ 即可完成条件从而得到一个不等式

$$R-P(\frac{p_1}{q_1},...,\frac{p_m}{q_m};r_1,...,r_m)<\varepsilon$$

但另一方面，我们又可以得到不等式

$$R-P(\frac{p_1}{q_1},...,\frac{p_m}{q_m};r_1,...,r_m)\geq\frac{\delta m}{8}$$

最后，两者结合可以推出 $\varepsilon>\frac{\delta m}{8}$，这与我们的预设条件 $\varepsilon<\frac{\delta}{20}$ 是矛盾的，从而证明了 Roth 定理。接着我们要推广到 Schmidt 定理，此时的代数数是 n 个 $r_1,...,r_n$，相应构造的线性型为

$$L_i(x)=x_i-r_ix_l,(i=1,...,n),L_l(x)=x_l,l=n+1\geq 2$$

于是我们同样可以得到一个不等式估计，也同样预设了一些条件

$$A_1...A_l=1,0<A_i<1,A_l>1,H_A=\{(x_1,...,x_l)\mid|L_i|\leq A_i,1\leq i\leq l\}$$

$$\lambda_1,...,\lambda_l=\min H_A(x_1,...,x_l)$$

则可以得到对任何 $\delta>0$ 存在常数 Q_1 使得 $Q\leq\max(A_l,Q_1)$ 时有

$$\lambda_1 > Q^{-\delta}$$

回到原来的定理，我们需要使用归纳法来证明。当 $n = 1$ 时，就是 Roth 定理，于是我们可以假设命题对小于 n 的情形，并来考虑 n 自己的情形。以反证法作为开始，假设有无穷多个 q 满足原来的不等式，自然就可以带来一波的 p_i，基础记号如下

$$l = n + 1, \eta = \frac{\varepsilon}{l}, A_i = |qa_i - p_i|q^{\eta}, 1 \leq i \leq n, A_l = (A_1 ... A_n)^{-1}$$

若 $A_1, ..., A_n$ 中有一个大等 1(不失一般性，我们就选它为 A_1)，就可得到 $|qa_i - p_i|q^{\eta} \leq 1$，从而

$$q^{1+\varepsilon-\eta} \prod_{i=2}^{n} (qa_i - p_i) < 1$$

由归纳假设，此时的 q 又是有限的，与反证假设矛盾，从而对于足够大的 q，我们只需考虑 $A_i < 1, 1 \leq i \leq n$ 情形下的无穷多个 q 即可，此时我们可以得到

$$A_l = (A_1 ... A_n)^{-1} = q^{-n\eta} (\prod_{i=1}^{n} (qa_i - p_i))^{-1} > q^{-n\eta} q^{1+\varepsilon} = q^{1+\eta}$$

于是我们发现，在上面假设下，这些 $A_h, 1 \leq h \leq l$ 自然地满足了上一个估计式的条件

$$A_i < 1, 1 \leq i \leq n, A_l > 1 \Rightarrow \forall Q \leq \max(A_l, Q_1), \lambda_1 > Q^{-\frac{\eta}{2n}}$$

我们结合每个局部的 Roth 定理 $\forall i, |qa_i - p_i| < q^{1+\varepsilon}$，并假设 $\varepsilon < 1$ 就可以得到

$$A_l = q^{-n\eta} (\prod_{i=1}^{n} (qa_i - p_i))^{-1} < q^{-n\eta} q^{n(1+\varepsilon)} < q^{2n}$$

我们在前一个不等式中取 $Q = q^{2n}$，就可得到不等式

$$\lambda_1 > q^{-\eta}$$

但另一方面，很容易注意到

$$|L_i| \leq A_i q^{-\eta}$$

从而 H_A 的第一个最小值满足

$$\lambda_1 \leq q^{-\eta}$$

引出不等式矛盾，从而完成了我们的证明。其实还有很多杂七杂八的逼近方法，个人认为这种东西应该去“数值计算方法”里寻找，而我所给的属于通用的数学逼近理论，在很多纯数学证明中有应用，所以才拿出来讨论的，至于其它的东西就真没啥好讲了，浅尝辄止何尝不行呢？

6.3 非初等函数

于我个人而言，除了觉得数学 (实、复) 分析在实数论中有不可取代的地位以外，最多也就加上一个更进一步的实变函数论，前面我们所见过的代数技巧，能很好分析的是代数数，最多可以到达代数数与超越数的边界，但超越数确实是数中的大头，那么我看重分析的原因就不言而喻了。但分析可谓算是每个本科生的基本功了，所

以我基本很少介绍分析的内容，而更多的都放在了应用方面，不过我还是要提醒读者，分析是一个整体的数学学科，不应该被微积分、实变函数、复分析、泛函分析、微分流形、微分方程等具体科目给分开了。读者可能注意到了，我们把泛函分析、微分流形、微分方程都放了进来，它们的核心其实都是函数，或者稍微转化后的图形方程，而且这三个学科具有十分紧密的关系，流形的精密结构需要靠微分方程研究，微分方程的求解又大量用着泛函分析的理论，如果我们计划部分地抛弃分析的细节内容，可否建立函数的代数理论呢？当然可以，这就是我们接下来要讨论的微分伽罗瓦理论 (Differential Galois Theory)，这节的内容主要来自 [25] 和 [6]。

两个定理

所谓**初等函数**是指：由幂函数 (power function)x^n、指数函数 (exponential function)$a^x, a>0$、对数函数 (logarithmic function)$\log_a x, a>0$、三角函数 (trigonometric function)$\sin x, \cos x, \ldots$、反三角函数 (inverse trigonometric function)$\arcsin x, \arccos x, \ldots$ 与常数 $c \in \mathbb{C}$ 经过有限次的有理运算 (加、减、乘、除、有理数次乘方、有理数次开方) 及有限次函数复合所产生，并且能用一个解析式表示的函数。

定理 6.13

(1)[Liouville] *初等函数 $f(x)$ 存在初等原函数当且仅当，存在初等函数 $u_1(x), \ldots, u_n(x), v(x)$ 和常数 $c_1, \ldots, c_n$ 满足*

$$f(x) = v'(x) + \sum_{i=1}^{n} c_i \frac{u_i'(x)}{u_i(x)}$$

(2)[Chebyshev] *函数 $f(x) = x^m(a+bx^n)^p, a, b \in \mathbb{R}, m, n, p \in \mathbb{Q}$ 存在初等原函数当且仅当*

$$p, \frac{m+1}{n}, \frac{m+1}{n} + p$$

中的任意一个是整数。

♡

作为一个新时代的人，我们去使用旧版的符号和证明是不合适的，因此我们将引入微分代数的语言，并重新描述定理后再证明。如果我们在交换幺环 R 上定义了**微分** (derivation)$\delta : R \to R, r \mapsto r'$，并且满足 $\forall r, s \in R, (rs)' = r's + rs', (r+s)' = r' + s'$，就称 (R, δ) 是一个**微分环**，微分环的同态 $\varphi : R \to S$ 指映射与微分是交换的 $\delta_S \varphi = \varphi \delta_R$，下面是一些微分环的实例。

例题 6.1 (1) 任意的环 R 和平凡微分 $\delta : r \mapsto 0$

(2) 多项式环 $\mathbb{R}[x]$ 和求导运算 $\frac{d}{dx} : \sum_i a_i x^i \mapsto \sum_i i a_i x^{i-1}$

(3) 有限域 $\mathbb{F}_q$ 上只有平凡微分。

什么是微分子环 $R \subset S$，什么是微分环扩张 S/R，其实就是在原来子环或扩张的基础上，加了一个保持微分运算 $\delta_S|_R = \delta_R$。更进一步，把环直接换成域就得到了微分域之类的基础概念，对此并没有重复的必要，因为域本身可以视为一个环。我们把 $R_C = \ker \delta = \{r \in R \mid \delta(r) = r' = 0\}$ 里的元素称为环的**常数**，以后我们所说的域均是特征为零的无限域，此时求导与微分有一个重要的联系。

定理 6.14

对于微分域扩张 $K \subset L$，如果 $e \in L - K$ 满足 $e' \in K, e' \neq 0$，则 e 是 K 上的超越元。

♡

证明 使用反证法，假设 e 是 K 上的代数元，设其极小多项式为 $p(t) = t^n + \sum_{i=0}^{m} c_i t^i \in K[t]$(即首一不可约多项式，$n > 1, 0 \leq m < n, c_m \neq 0$)。由于 $e' \in K$，我们可以构造一个新的多项式

$$q(t) = ne't^{n-1} + \sum_{i=0}^{m} c_i' t^i + \sum_{i=1}^{m} i c_i e' t^{i-1} \in K[t]$$

借助 $p(e) = e^n + \sum_{i=0}^{m} c_i e^i = 0$ 的两边求导 (别被多项式的样子给带偏了，这里的 e 其实应该是我们通常所见到的

函数) 可得

$$0=(p(e))'=(e^n)'+\sum_{i=1}^{m}(c_ie^i)'+(c_0e^0)'=ne^{n-1}e'+\sum_{i=1}^{m}(c_i'e^i+c_iie^{i-1}e')+c_0'e^0=q(e)$$

由于 $\deg p(t)=n-1<n=\deg q(t)$，故只能 $q(t)\equiv 0$，但又由于 $e'\neq 0\Rightarrow ne'\neq 0\Rightarrow q(t)\neq 0$，从而矛盾。

我们考虑最基础的分式函数域 $K=\mathbb{R}(x),\mathbb{C}(x)$，由于 $0\neq(\ln x)'=\frac{1}{x}\in K, 0\neq(\arctan x)'=\frac{1}{1+x^2}\in K\ldots$，故除了幂函数以外其实基本所有的初等函数，都是上面函数域上的超越元。更进一步，如果考虑域扩张 K/K_C，则任意 $K-K_C$ 的元素都是 K_C 上的超越元，如果 K 的特征非零且使用非平凡微分 (不有限即可 $\mathbb{F}_q(x)$) 则 $K-K_C$ 的元素都是 K_C 上的代数元。实际上，上面的域扩张还未从我们需要的扩张中提取足够信息，如果微分环扩张 $R\subset S$ 还满足 $R_C=S_C$，就称其是一个**无新常数** (with no new constant) 微分环扩张，它有几个等价命题。

命题 6.1

对于微分环扩张 $R\subset S$，以下三个命题互相等价

(1)$R_C=S_C$

(2) 如果 r 在 R 中有原函数，则在 $S-R$ 中没有原函数

(3) 如果 $s\in S-R, s'\in R$，则 s' 在 R 中没有原函数。

♠

通常我们把 r 称为 r' 的一个**原函数**，这是大家在不定积分中常见的一个事实 $(r+R_C)=r'$，所以上面的命题其实是十分显然的，想要求证的读者可以自己用形式化的方式验证一把。在这个定义下，我们同样也是要获得一个超越元的判定定理。

定理 6.15

对于无新常数微分域扩张 $K\subset L$，如果 $e\in L-K$ 满足 $\frac{e'}{e}\in K$，则

(1)e 是 K 上的代数元当且仅当，存在 $n>1$ 使得 $e^n\in K$

(2) 如果 e 是 K 上的超越元，则任意 $n>1$ 次多项式 $p(e)\in K[e]$ 的导数 $(p(e))'$ 依旧是 n 次。

♡

证明 (1) 必要性是显然的，我们来证充分性。由于 e 是 K 上的代数元，我们设它的极小多项式为 $p(t)=t^n+\sum_{i=0}^{m}c_it^i\in K[t]$(即首一不可约多项式，$n\leq 1, 0\leq m<n, c_m\neq 0$)，对 $p(e)$ 求导后可得

$$q(e)=bne^n+\sum_{i=1}^{m}(c_i'+c_iib)e^i+c_0', b=\frac{e'}{e}\in K$$

显然通过 $bnp(e)-q(e)$ 会生成一个次数小于 n 的以 e 为根的多项式，故它只能恒为零，提取首系数可得 $bnc_m-(c_m'+mbc_m)=0$，简化可得 $\frac{c_m'}{c_m}=(n-m)b$，此时有

$$\frac{(c_me^{m-n})'}{c_me^{m-n}}=\frac{(m-n)bc_me^{m-n}+c_m'e^{m-n}}{c_me^{m-n}}=(m-n)b+\frac{c_m'}{c_m}=0$$

故 $c_me^{m-n}\in L_C=K_C\subset K$，即 $e^{m-n}\in K\Rightarrow e^{n-m}\in K$，故存在 $n-m>1$ 使得 $e^{n-m}\in K$。

(2) 我们设 $p(e)=\sum_{i=0}^{n}c_ie^i$，则有

$$(p(e))'=\sum_{i=0}^{n}(c_i'+c_iib)e^i+c_0'$$

如果它的次数不是 n，则有 $0=(c_n'+c_nnb)e^n=(c_ne^n)'$，即 $c_ne^n\in K_C\subset K\Rightarrow e^n\in K$，但 e 是 K 上的超越元，从而与 (1) 矛盾。

和上面一样地，对任意的 $g\in\mathbb{R}(x)$，我们有 $\frac{(e^g)'}{e^g}=g'\in K$ 并且 $\forall n>1,(e^g)^n=e^{ng}\notin K$，故 e^g 是 K 上的超越元。不失一般性，考虑初等函数时，都可以看称复数域情况 $\mathbb{C}(x)$，注意到几个简单的事实

$$a^x=e^{\ln a}e^x \qquad \sin z=\frac{e^{iz}-e^{-iz}}{2i} \qquad \arcsin z=\frac{\ln(iz+\sqrt{1-z^2})}{i}$$
$$\log_a x=(\ln a)^{-1}\ln x \qquad \cos z=\frac{e^{iz}+e^{-iz}}{2} \qquad \arccos=\frac{\ln(z-i\sqrt{1-z^2})}{i}$$

换言之，所谓的初等函数，其实只有 $e^x, \ln x$ 是急需的超越元，我们试着在微分域 K 中把这类元素给定义出来。如果 $s,t \in K$ 满足

$$\frac{s'}{s} = t'$$

我们就称 s 是 t 的**指数**，t 是 s 的**对数**，并分别记为 $s = e^t, t = \ln s$，此时我们的核心定义就出来了。

定义 6.1

如果微分域扩张 $K \subset L$ 存在一个扩张塔

$$K = K_0 \subset K_1 \subset \ldots \subset K_{n-1} \subset K_n = L, K_i = K_{i-1}(e_i), 1 \leq i \leq n$$

使得 $\forall 1 \leq i \leq n, e_i$ 是下列三种情况中的一种

(1)e_i 是 K_{i-1} 上的代数元

(2) 存在 $k \in K_{i-1}$ 使得 $e_i = e^k$

(3) 存在 $k \in K_{i-1}$ 使得 $e_i = \ln k$

我们就称 L/K 是初等微分域扩张，或称为刘维尔扩张 (Liouville extension)。

♣

读者需要注意一点，根据伽罗瓦理论可知，根式一定可以由代数元得到，但代数元不一定可以由根式表示，换言之如果把 $\mathbb{C}(x)$ 的任一初等微分域扩域中的元素称为初等函数的话，其范围是比我们常规所认识的初等函数的范围更大的，但就应用而言，这并不会有太大影响。此时，我们的核心定理的微分域表述也就出来了。

定理 6.16

设 K 是微分域，则 $f \in K$ 的原函数在 K 的一个初等无新常数微分域扩张当且仅当，存在 $u_1, \ldots, u_m, v \in K$ 和 $c_1, \ldots, c_m \in K_C$ 使得

$$f = v' + \sum_{i=1}^{m} c_i \frac{u_i'}{u_i}$$

♡

证明 由 $f = (v + \sum_{i=1}^{n} c_i \ln u_i)'$ 可知必要性是显然的，故我们只考虑充分性。此时我们有一个初等微分域扩张塔

$$K = K_0 \subset K_1 \subset \ldots \subset K_n, \exists g \in K_n, g' = f$$

我们对扩张塔的长度 n 进行归纳法，当 $n = 0$ 时，即 $\exists g \in K, g' = f$，此时 $v = g, m = 1, c_1 = 0$，接着我们假设对 $n-1$ 长度的扩张塔成立，我们视 $f \in K_0 \subset K_1$，则在扩张 $K_1 \subset K_n$ 存在 $u_1, \ldots, u_m, v \in K_1$ 和 $c_1, \ldots, c_m \in K_C$(无新常数扩张) 使得

$$f = v' + \sum_{i=1}^{m} c_i \frac{u_i'}{u_i}$$

在这个条件下，我们来考虑最开始的扩张 $K_1 = K(e)$，根据定义其分为三种情况

(1)[e 是 K 上的代数元] 此时 $K_1 = K(e) = K[e]$，设 $p(x) = \prod_{i=1}^{s}(x - e_i), e_1 = e$ 是 e 的极小多项式，相应的分裂域为 $K^p \supset K_1$，设 $\sigma_i \in Gal(K^p/K), i = 1, \ldots, s$ 是所有的自同构，并不失一般性地假定 σ_1 是恒等映射。此时 $\forall q(e) \in K[e]$ 有 $\sigma_i(q(e)) = q(e_i)$，此时我们可以选取合适的多项式 $q_1(x), \ldots, q_m(x), r(x) \in K[x]$ 使得

$$u_i = q_i(e), 1 \leq i \leq m, v = r(e)$$

则有

$$f = (r(e))' + \sum_{j=1}^{m} c_j \frac{(q_j(e))'}{q_j(e)}$$

由 $\sigma_i(f)=f$ 可得

$$f=\sigma_i(f)=\sigma_i((r(e))'+\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)})=(r(e_i))'+\sum_{j=1}^{m}c_j\frac{(q_j(e_i))'}{q_j(e_i)}$$

遍历所有的 $1\le i\le s$ 并相加后可得

$$f=\frac{\sum_{i=1}^{s}\sigma_i(f)}{s}=(\frac{\sum_{i=1}^{s}r(e_i)}{s})'+\sum_{j=1}^{n}\frac{c_j}{s}\frac{(\prod_{i=1}^{s}q_j(e_i))'}{\prod_{i=1}^{s}q_j(e_i)}$$

由对称多项式的理论即可知，此时有

$$v^{new}=\frac{\sum_{i=1}^{s}r(e_i)}{s}\in K,u_i^{new}=\prod_{j=1}^{s}q_i(e_j)\in K,c_i^{new}=\frac{c_i}{s}\in K_C$$

(2)[共有部分说明] 在后面两种情况下，e 是 K 上的超越元，此时我们可以同样地找到有理多项式 $q_1(x),...,q_m(x),r(x)\in K(x)\cong K(e)$ 使得

$$u_j=q_j(e),1\le j\le m,v=r(e),f=(r(e))'+\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}$$

对于每个有理多项式 $q_j(e)\in K(e)\cong K(x)$，根据因式分解有

$$q_j(e)=k_j\prod_{i=1}^{n_j}(q_{ji}(e))^{n_{ji}},k_j\in K,n_j\ge 1,n_{ji}\in\mathbb{Z},q_{ji}(e)\in K[e]\text{是首一不可约多项式}$$

再根据对数恒等式

$$\frac{(\prod_{j=1}^{m}k_j^{m_j})'}{\prod_{j=1}^{m}k_j^{m_j}}=(\ln\prod_{j=1}^{m}k_j^{m_j})'=(\sum_{j=1}^{m}m_j\ln(k_j))'=\sum_{j=1}^{m}m_j\frac{(k_j)'}{k_j}$$

故对 $\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}$ 进行一轮操作变换以后形式不变，因此我们可以不失一般性地认为 f 分解式中的 $q_j(e)\in K[e]$，或者是非常数[2]首一不可约多项式，或者是 K 中的元素[3]。

(3)[存在 $k\in K$ 使得 $e'=\frac{k'}{k}$] 此时由于 $e'=\frac{k'}{k}\in K$，故必有 $\deg(q_j(e))'<\deg q_j(e)$。我们注意到部分分式的唯一分解定理为

$$g(e)=k\sum_{i=1}^{l}p_i^{m_i}(e),\frac{f(e)}{g(e)}=\sum_{i=1}^{l}\sum_{k_i=1}^{m_i}\frac{r_{ik_i}(e)}{(p_i(e))^{k_i}},p_i(e)\in K[e]\text{是首一不可约多项式},\deg r_{ik_i}(e)<\deg p_i(e)$$

若存在 $p(e)=q_j(e)\notin K$，则在 $\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}$ 的部分分式中最多只有 $\frac{r(e)}{(p(e))^m},m\le 1$。由于式子左边的 $f\in K$，故为了消去这个分式，$r(e)$ 的部分分式中也应该存在一项 $\frac{f(e)}{(p(e))^d}$，并且这里的 d 是可能形式的最大值，此时

$$(\frac{f(e)}{(p(e))^d})'=-\frac{df(e)(p(e))'}{(p(e))^{d+1}}+\sum_{i=1}^{d}\frac{r_i(e)}{(p(e))^i}$$

中的 $-\frac{df(e)(p(e))'}{p(e)^{d+1}}$ 是 $(r(e))'$ 的最大次数项，没法在 $(r(e))'$ 中消去，又因为 $d+1>1$ 导致也没法在 $\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}$ 消去。从而只有 $\forall j,q_j(e)\in K$。

[2]常数时，$c_j\frac{(q_j(e))'}{q_j(e)}=0$，则没有讨论的必有

[3]此处主要是来自为了保证首一而提出来的系数，如前面的 k_j

更进一步，由于 $(r(e))'=f-\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}\in K$，故只有形式 $r(e)=c_{m+1}e+c_0,c_{m+1},c_0\in K_C$，于是我们有

$$f=(r(e))'+\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}=(c_{m+1}e+c_0)'+\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}=c_0'+\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}+c_{m+1}\frac{k'}{k}$$

即

$$v^{new}=c_0\in K,u_i^{new}=q_i(e)\in K,c_i^{new}=c_i\in K_C,1\le i\le m,u_{m+1}=k\in K,c_{m+1}\in K_C$$

(4)[存在 $k\in K$ 使得 $\frac{e'}{e}=k'$] 此时由于 $\frac{e'}{e}=k'\in K$，故 $p(e)=q_j(e)\notin K$ 当且仅当 $p(e)=e$，合并所有项以后，我们不失一般性地假设 $q_1(e)=e$，而其它项均有 $q_j(e)\in K,2\le j\le m$。

更进一步，为了消去 $\frac{e'}{e}$ 项，则 $r(e)=\sum_{i=-t}^{t}k_ie^i,t>0,k_j\in K$，并且同样有 $(r(e))'=f-\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}\in K$，从而 $r(e)\in K$，于是我们有

$$f=(r(e))'+\sum_{j=1}^{m}c_j\frac{(q_j(e))'}{q_j(e)}=(r(e))'+\sum_{j=2}^{m}c_j\frac{(q_j(e))'}{q_j(e)}+c_1\frac{e'}{e}=(r(e)+c_1k)'+\sum_{j=2}^{m}c_j\frac{(q_j(e))'}{q_j(e)}$$

即

$$v^{new}=r(e)+c_1k\in K,u_i^{new}=q_{i-1}(e)\in K,c_i^{new}=c_{i-1}\in K_C,2\le i\le m$$

通过这个核心的判定定理，就能得到不少的结论了，例如

$$fe^g\text{有初等原函数}\Leftrightarrow\exists a,f=a'+ag'$$

借助这个定理我们可以证明 $f(x)=e^{x^2}$(此处 $f=1,g=x^2$) 没有初等原函数，否则存在一个有理式 $a(x)\in\mathbb{R}(x)$ 使得 $1=a'(x)+2a(x)x$，我们假设 $a(x)=\frac{p(x)}{q(x)},(p,q)=1,p(x),q(x)\in\mathbb{R}[x]$ 则有

$$1=a'(x)+2a(x)x\Rightarrow q(q-2px-p')=-q'p\Rightarrow q\mid q'p\Rightarrow q\mid q'\Rightarrow q'=0$$

因此 q 是常数，从而有多项式 $1=b_1p'+2b_2px$，显然是矛盾的。至于另一个定理，我们可以先看必要性，若 p 是整数，则有

$$(\sum_{i=0}^{p}\frac{C_p^ia^{p-i}b^i}{ni+m+1}x^{ni+m+1})'=\sum_{i=0}^{p}C_p^ia^{p-i}b^ix^{ni+m}=x^m(a+bx^n)^p$$

若 $\frac{m+1}{n}$ 是整数，则有

$$t=a+bx^n,\int x^m(a+bx^n)^pdx=\frac{1}{n}b^{-\frac{m+1}{n}}\int t^p(t-a)^{\frac{m+1}{n}-1}dt$$

从而化为上一种情形。若 $\frac{m+1}{n}+p$ 是整数，则有

$$\int x^m(a+bx^n)^pdx=\int x^{m+np}(ax^{-n}+b)^pdx$$

此时 $\frac{m+np+1}{-n}=-(\frac{m+1}{n}+p)$ 是整数，从而化为上一种情形。至于充分性，则属于椭圆积分 (或称为阿贝尔积分) 的判定

$$f(x)=\int_a^xR(t,\sqrt{y(t)})dt,R(x,y)\in\mathbb{R}[x,y],y(x)\in\mathbb{R}[x],\deg y(x)=3,4$$

若 $f(x)$ 在每点有限 (例如 $R(t,\sqrt{y(t)})=\frac{1}{\sqrt{1+t^4}}$)，则称为第一类椭圆积分，其必定没有初等原函数；若 $f(x)$ 在有限个点处值为无限大 (例如 $R(t,\sqrt{y(t)})=\frac{t}{(1-t^2)^{\frac{3}{2}}},\frac{1}{(1-t)^{\frac{3}{2}}}$)，则称为第二类椭圆积分，若其存在初等原函数，则

必定是 $R(t,\sqrt{y(t)})$ 的形式，更多的内容就留给读者练习了，我们就单纯地拿 Liouville 定理开个头吧。

Picard-Vessiot 理论

我们使用 $\mathbb{E}=\overline{\mathbb{C}(x,e^x,\ln x)}$ 来表示初等函数微分域，则我们上面所讨论的刘维尔定理实际就是说明：在 f 满足某种条件下，线性微分方程 $g'=f$ 存在解 $g\in\mathbb{E}$。更进一步，我们可以考虑像下面这样的 1 阶齐次线性微分方程组

$$\frac{d}{dx}\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}=\begin{pmatrix}a_{11}&\cdots&a_{1n}\\ \vdots&\ddots&\vdots\\ a_{n1}&\cdots&a_{nn}\end{pmatrix}\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}$$

对于上面的方程组，我们可以进行像下面这样的简写

$$y'=Ay,y=(y_1,\ldots,y_n),A\in M_n(K),1\le i\le n$$

这里的 K 是我们所讨论的微分域。另外，如果我们取

$$(y_1,\ldots,y_n)=(y,y',\ldots,y^{(n-1)}),A=\begin{pmatrix}0&1&0&0&\cdots&0\\ 0&0&1&0&\cdots&0\\ \vdots&\vdots&\vdots&\vdots&\ldots&\vdots\\ 0&0&0&0&\cdots&1\\ -a_0&-a_1&\cdots&\cdots&\cdots&-a_{n-1}\end{pmatrix}$$

就可以得到一个 n 阶齐次线性微分方程

$$y^{(n)}+a_{n-1}y^{(n-1)}+\ldots+a_1y'+a_0y=0$$

因此这类 1 阶齐次线性微分方程组 $y'=Ay,A\in M_n(K)$ 具有广泛的应用，有关其解的结构定理，属于线性微分方程组的结论，我们直接给出。

定理 6.17 (解的结构定理)

设 K 是微分域，$y'=Ay,A\in M_n(K)$ 是微分方程，L 是 K 的无新常数 $(L_C=K_C)$ 微分域扩张

(1) 若 $y_1,\ldots,y_n\in L^n$ 是微分方程的解，则 $\forall c_i\in K_C,\sum_{i=1}^{n}c_iy_i$ 也是微分方程的解。

(2) 微分方程解 $y_1,\ldots,y_n$ 在 K_C 上线性独立当且仅当 $W(y_1,\ldots,y_n)\neq 0$，其中

$$W(y_1,\ldots,y_n)=\begin{vmatrix}y_1&y_2&\cdots&y_n\\ y_1'&y_2'&\cdots&y_n'\\ \vdots&\vdots&\vdots&\vdots\\ y_1^{(n-1)}&y_2^{(n-1)}&\cdots&y_n^{(n-1)}\end{vmatrix}$$

(3) 微分方程至多有 n 个 K_C 上线性独立的解，即解空间 $\{v\in L^n\mid v'=Av\}$ 是 K_C 上的至多 n 维线性空间。♡

读者需要注意，这里解空间至多 n 维是因为我们的 L 微分域可能不够大，在通常的微分方程领域内，考虑的是一个足够大的可微函数域，所以就一定会有 n 个线性无关的解，我们这 n 个线性无关的解 $y_1,\ldots,y_n$ 称为微分方程 $y'=Ay$ 的一个**基本解组**。针对这类微分域的扩张，我们引入下面的核心概念

定义 6.2

设 K 是微分域，$y'=Ay, A\in M_n(K)$ 是 1 阶齐次线性微分方程组，L/K 是微分域扩张，如果有

(1)[无新常数]$L_C=K_C$

(2)[由解生成]$L=K\langle y_1,...,y_n\rangle$，其中 $y_1,...,y_n\in L^n$ 是微分方程 $y'=Ay$ 的一个基本解组

我们就称 L/K 是关于微分方程 $y'=Ay$ 的 Picard-Vessiot 扩张。

♣

这里比较不清晰的是 $L=K\langle y_1,...,y_n\rangle$，虽然 $y_1,...,y_n\in L^n$ 都是 n 维向量，但实际上我们有一种正交化的手段使得

$$y_1=\begin{pmatrix}y_1\\0\\\vdots\\0\end{pmatrix},y_2=\begin{pmatrix}0\\y_2\\\vdots\\0\end{pmatrix},...,y_n=\begin{pmatrix}0\\0\\\vdots\\y_n\end{pmatrix}$$

从而，我们可以把 $y_1,...,y_n\in L$ 视为一个基本解组，于是其生成的微分域就像下面这样

$$L=K\langle y_1,...,y_n\rangle=\{\sum_{i=1}^{n}a_iy_i\mid a_i\in K\}$$

这种把解添加到域中的做法，实际和我们在抽象代数里见到的多项式分裂域有些类似，至于微分方程 $y'=Ay, A\in M_n(K)$ 的 Picard-Vessiot 扩张的存在性和唯一性，我们在常微分方程理论中已经见过了，因此下一步自然是对其伽罗瓦理论的讨论了。

定义 6.3

(1) 对于微分域 L 和 K，如果映射 $f:L\to K$ 满足，f 是域同态且 $\forall r\in L, f(r')=(f(r))'$，则称 f 是微分域同态，我们将所有微分域同态构成的集合记为 $End_{diff}(L,K)$

(2) 对于微分域扩张 L/K，它的微分伽罗瓦群指下面的微分自同构组成的群

$$Gal(L/K)=\{\sigma\in End_{diff}(L,L)\mid\forall f\in K,\sigma(f)=f\}$$

♣

通常我们所说的，微分方程 $y'=Ay, A\in M_n(K)$ 的微分伽罗瓦群 $Gal(L/K)$ 指其中的 L/K 是关于这个微分方程的 Picard-Vessiot 扩张。接下来我们不仅会要求 K 的特征为零，还会默认 K_C 是代数闭域，从而得到类似代数方程的伽罗瓦对应定理。

定理 6.18 (嵌入定理)

对任意的微分域 K 和它的 Picard-Vessiot 扩张 $L=K\langle y_1,...,y_n\rangle$，存在一个多项式集合 $S=\{F_{ij}(X),1\le i,j\le n\}\subset K_C[X]$ 满足

(1) 如果 $\sigma\in Gal(L/K)$ 且 $\sigma(y_j)=\sum_{i=1}^{n}c_{ij}y_i$，则 $\forall F(X)\in S, F(c_{ij})=0$

(2) 如果矩阵 $(c_{ij})\in GL_n(K_C)$ 满足 $\forall F(X)\in S, F(c_{ij})=0$，则存在 $\sigma\in Gal(L/K)$ 使得 $\sigma(y_j)=\sum_{i=1}^{n}c_{ij}y_i$

换言之，存在嵌入 $G(L/K)\to GL_n(K_C)$ 使得 $G(L/K)$ 是 $GL_n(K_C)$ 的闭子群。

♡

定理 6.19 (对应定理)

设 L/K 是 Picard-Vessiot 扩张，$G=Gal(L/K)$ 是它的微分伽罗瓦群。定义

$\mathcal{S}$ 由 G 的所有闭子群构成 (这里的“闭”指代数群 $GL_n(K_C)$ 的 Zariski 拓扑中的闭集)

$\mathcal{L}$ 由 L 的所有包含 K 的微分闭子域构成

$\alpha:\mathcal{S}\to\mathcal{L}, H\mapsto L^H=\{f\in L\mid\forall\sigma\in H,\sigma(f)=f\}$

$\beta: \mathcal{L} \to \mathcal{S}, F \mapsto Gal(L/K)$，则有

(1)$\alpha = \beta^{-1}$

(2)$F \in \mathcal{L}$ 是 K 的 Picard-Vessiot 扩张当且仅当，$H = Gal(L/F)$ 是 $G = Gal(L/K)$ 的正规子群。且此时有

$$G/H \cong Gal(F/K)$$

对应定理构成了我们微分伽罗瓦理论的基础，最后我们来把它和上一部分中的 Liouville 扩张联系起来，我们的使命也就结束了。

定理 6.20

设 L/K 是 Picard-Vessiot 扩张，$G = Gal(L/K)$ 是它的微分伽罗瓦群，G^o 是 G 包含单位元的不可约成分[a](或简称单位分支)，则下面的命题等价

(1)G^o 是可解群

(2) 存在 Liouville 扩张 M/K 使得 $L \subset M$。

[a]详细定义可以见“代数群”的相关书籍，或者 [45] 的 2.6 中关于代数群的说明

小结

实际上，是否要将 Riemann-Hilbert 对应纳入微分伽罗瓦理论的讨论，我是比较犹豫的，这本书 [28] 好像想要表达从黎曼-希尔伯特对应引入微分伽罗瓦理论。但我们知道 Riemann-Hilbert 对应实际就是 Hilbert 第 21 问，它是有关 Fuchs 类的线性微分方程的延拓表示问题，虽然也在我们的这个微分代数体系下，但与我们微分伽罗瓦理论的关系我暂时没能搞懂，所以我最后也就止步于了 Picard-Vessiot 理论。而有关线性微分方程，还有一个比较重要的 Siegel-Shidlovskii 定理，大家不用去看比较老的书，可以看这篇文章 [9]，其核心内容并不复杂，就是定义了一类 E 函数域，并研究发现其是运算封闭的，这里的运算除了域自带的加法和乘法，还有求导和求原函数，于是我们就能用带着这个重要的特性的微分域去研究线性微分方程，从而发现一个线性无关的传导性。其实它的重要体现在，我们初等函数域所不具备的求导和求原函数的封闭性，$e^z, \sin z, \cos z, \ln|x| = \int \frac{1}{x}dx$，因此 E 函数微分域实际比我们的初等函数域还要大，算是脱离我们的解析式后，值得研究一番的函数域了。

数的分类：

A 数 (代数数 $\mathbb{A}$)：$\sqrt{2}$。S 数 (主要部分)：e。T 数 (所知甚少)。U 数 (最早的超越数)：$\sum_{n=1}^{\infty} \lambda^{-n!}, \lambda > 1$。

超越数判定：

Lindemann-Weierstrass 定理：$\sum_{i=1}^{n} a_i e^{b_i} \neq 0$。Baker 定理：$a_0 + \sum_{i=1}^{n} a_i \ln b_i \neq 0$

近似的三种方式：

连分数：$r - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(r_{n+1}q_n + q_{n-1})}, r_{n+1} = [x_{n+1}; x_{n+2}, x_{n+3}, \ldots]$。无穷级数：$\sum_{i=1}^{\infty} x_i = \lim_{n\to\infty} S_n, S_n = \sum_{i=1}^{n} x_i$。无穷乘积：$\prod_{i=1}^{\infty} p_i = \lim_{n\to\infty} P_n, P_n = \prod_{i=1}^{n} p_i$

微分伽罗瓦理论类比：

域 $F \leftrightarrow$ 微分域 K，多项式方程 $f(x) = 0, f(x) \in F[x] \leftrightarrow$ 微分方程 $y' = Ay, A \in M_n(K)$，伽罗瓦扩张 $L/F \leftrightarrow$Picard-Vessiot 扩张 L/K，伽罗瓦群 $G = Gal(L/F) \leftrightarrow$ 微分伽罗瓦群 $G = Gal(L/F)$，运用根式可解 $\leftrightarrow$ 可以嵌入到 Liouville 扩张中，G 是可解群 $\leftrightarrow G^o$ 是可解群。

核心执念：

指对数函数 $e^x, \ln x$ 起着十分重要的作用，无论是在超越数判定中，还是在非初等原函数判定中。

$$\ln K = \zeta(2)\log_2 \frac{A^{12}}{2\pi e^{\gamma}} + \prod_{k=1}^{\infty} \frac{1 + 2^{2^{-k}}}{2} \sum_{k=3}^{\infty} \frac{(-1)^k (2 - 2^k)\zeta'(k)}{k}, \text{K=Khinchin,A=Glaisher-Kinkelin}$$

第七章 方程与曲线

在整个数学里，我最看重的学科就两个，数学分析和线性代数，对于前者我们在上一个实数篇章中介绍了它的重要性，而这一章中，我们将看到线性代数的威力。不过到底啥是线性代数呢？是所谓的高等代数吗？还是应该也把抽象代数给带进来，我的建议是通通搞进来。你可能会说，数学不应该是分析、代数、几何的三足鼎立吗？那我把几何扔哪里去了。简单来讲，几何是代数，但代数不是几何，几何和数论有着类似的背景，可以同时从代数和分析两个方向进行研究。那最存粹的几何学“点集拓扑”呢？问题是，它跟数论有联系吗？我以前就说过，点集拓扑实际就是一套名词系统，给予各种对象一种类似于几何的感觉，所以有时也能见到闭集等概念，但对此我们没必要特地地拿到数论这里来讲。接着拓扑无论往哪个方向发展，都离不开代数和分析，所以我们舍弃掉几何是有理有据的，读者不必心存芥蒂。在这一章中，我们的主题是方程。

7.1 不定方程

所谓方程问题，实际就是求映射 $f:A\to B$ 在某点逆解集 $f^{-1}(b)\subset A, b\in B$ 的过程，但越泛的问题就越没有可供研究的内容，也就是基本无有用条件，就根本没有研究的必要了。所以我们就把它放到我们熟知的环 $K=\mathbb{Z}$ 或域 $K=\mathbb{Q},\mathbb{Q}_p,\mathbb{R},\mathbb{C}$ 上，再把映射变成有理式 $f(X)\in K(x_1,...,x_n)$，某点变成零 $0\in K$，就有了我们通常意义下研究的方程

$$f(X)=0, X\in K^n$$

但未知数的个数 $n=1$ 时，由代数基本定理可知，其在 $\mathbb{C}$ 上一定有根，所以我们至少需要讨论 $n>1$，而此时我们可以通过函数反解，得到几乎任意多个 $\mathbb{C}$ 和 $\mathbb{R}$ 上的解，因此我们还需要进一步限制解在 $\mathbb{Q}$ 或 $\mathbb{Z}$。更进一步，由于有理方程可以通过通分变成多项式，于是，所谓的**不定方程**问题或**丢番图方程**问题指：求变量个数大于 1 的多项式方程的整数解或有理解 (齐次情形相当于整数解) 问题。比如下面是一些常见的不定方程

$$x^n+y^n=z^n, n=x^2+y^2, \frac{x}{y+z}+\frac{y}{x+z}+\frac{z}{x+y}=n, x^2-dy^2=n, \frac{4}{n}=\frac{1}{x}+\frac{1}{y}+\frac{1}{z}$$

对于不定方程，我们更倾向于找整数解，不要问为什么，因为事实就是这样。我们有两个核心的问题，一是不定方程是否有解，二是如果有解那么怎么把它们全部表示出来，通常我们都只活在第一个问题中，第二个问题基本都是顺带的。于是问题又来了，能否通过有限步骤来判定不定方程是否存在整数解？这是希尔伯特第十问，这其实属于数理逻辑中的图灵机算法问题，结论是通用的算法并不存在，详细过程可以看这本书 [23]，它不仅比较新，还在前面介绍了所需要的各种泛用性的核心工具，至于我们只要知道这个事实就行了。在通用算法不存在的情况下，意味着在不定方程上，我们有着做不完的工作，也就是接下来我们可以讨论的具体方程类型有好多，这样我们就能水大量内容了。

二次不定方程

多项式的次数是一个很好的讨论指标，那么我们的一次不定方程哪里去了？因为它实在是过于简单了，所以也就懒得讨论了。

定理 7.1 (一次不定方程)

设 $s\geq 2$ 元一次不定方程为

$$\sum_{i=1}^{s}a_ix_i=a_1x_1+...+a_sx_s=n, a_i, n\in\mathbb{Z}, \prod_{i=1}^{s}a_i\neq 0$$

(1) 此方程有整数解 $(x_1,...,x_s)$ 当且仅当，$(a_1,...,a_s)\mid n$

(2) 如果 $(x_1^0,...,x_s^0)$ 是此方程的一个特解，$(x_1^*,...,x_s^*)$ 是此方程 $n=0$ 时的齐次方程的通解，则此方程的通解为

$$(x_1,...,x_s)=(x_1^0+x_1^*,...,x_s^0+x_s^*)$$

(3) 齐次方程 $a_1x_1+a_2x_2=0$ 的通解为

$$x_1=\frac{a_2}{(a_1,a_2)}t,x_2=-\frac{a_1}{(a_1,a_2)}t,t\in\mathbb{Z}$$

齐次方程 $\sum_{i=1}^{s}a_ix_i=0$ 的通解由上一级非齐次方程给出

$$\sum_{i=1}^{s-1}a_ix_i=-a_sx_s$$

其中的 x_s 由二元齐次方程

$$a_sx_s+(a_1,...,a_{s-1})y=0$$

给出。

♡

注 我们实际需要证明的只有 (1) 和 (2)，而 (3) 只是告诉你一个递归求解的过程，即“n 元齐次 =n-1 元非齐次 +2 元齐次”，再结合“n 元非齐次 = 辗转相除法求特解 +n 元齐次”，最后的最后得到的都是 2 元齐次方程的求解，而且由递归的过程可知，n 元齐次方程应该至少有 n-1 个自由变量来确定所有的整数解。由于证明的过程十分简单，就留给读者做练习了。于是，求解一次不定方程的基本方式是，先通过 (1) 证明中的方法得到一个特解，再通过 (3) 的公式求出齐次方程的通解，最后利用 (2) 的理论就可以得到原方程的所有整数解了，至此一次方程就没什么可以说的了，难道不是吗？另外，读者如果不想通过齐次的方式去理解解的构造，我们可以直接把原来的 s 元一次不定方程，化为 s-1 个二元一次不定方程

$$d_{s-1}y_{s-1}+a_sx_s=n,d_iy_i+a_{i+1}x_{i+1}=d_{i+1}y_{i+1}(1\le i\le s-2),y_1=x_1,d_1=a_1,d_i=(d_{i-1},a_i)$$

比起无聊的一次，我们还是来看看有趣的二次不定方程吧。由于整数且行列式为正负 *1* 的线性变换会保持两个不定方程整数解的对应，于是根据二次型的理论，或者就是单纯地一个变量一个变量地进行配方，遇到非整数时，可以同时乘以分母来消去它，最终大部分的二次不定方程都具有下面的形式

$$\sum_{i=1}^{s}a_ix_i^2=a_1x_1^2+...+a_sx_s^2=n,a_i,n\in\mathbb{Z},\prod_{i=1}^{s}a_i\neq0$$

于是这类方程也就变成了我们讨论的重点，而其中 $s=2$ 时，就有着极其丰富的理论，所以我们先来花大量笔墨讨论不定方程 $ax^2+by^2=c$。我们进一步取 $a=1,b=-d,c=\pm1,d$ 是非平方数，就得到了著名的 Pell 方程 $x^2-dy^2=\pm1$。一个 Pell 方程，就能得到一本书 [3]，更别说更大的二次不定方程了 [1]，所以我只会挑我觉得重要的来讲，而省去大量的内容，请别在意这是无法摆脱的篇幅限制。

定理 7.2 (Pell 方程)

设二元二次 Pell 方程为

$$x^2-dy^2=1,d>0,d\in\mathbb{Z}$$

(1) 若 $d=s^2,s\in\mathbb{N}$，则方程仅有整数解 $(x,y)=(1,0)$

(2) 若 d 是非平方数，则方程有无数组正整数解[a]，设 $(x,y)=(x_1,y_1)$ 是使得 $x+y\sqrt{d}$ 最小的正整数解，并记

$$x_n+\sqrt{d}y_n=(x_1+y_1\sqrt{d})^n$$

则 (x_n,y_n) 是方程所有的正整数解。

[a]负整数解可以直接对称过去，零解可以直接验证

证明 对于 (1) 可以直接因式分解，借助代数基本定理即可得到答案，所以我们主要来看 (2) 的证明。

(1) 我们先来证明一个引理，由于 $\sqrt{d}$ 是实数，故存在无穷多组 (x,y) 使得 (狄利克雷定理)

$$|\sqrt{d}-\frac{x}{y}|<\frac{1}{y^2}$$

于是存在无穷多组 (x,y) 使得

$$|x^2-dy^2|=|x+y\sqrt{d}||x-y\sqrt{d}|<|x-y\sqrt{d}+2y\sqrt{d}|\frac{1}{y}<(\frac{1}{y}+2y\sqrt{d})\frac{1}{y}\leq 1+2\sqrt{d}$$

由于满足 $|M|<1+2\sqrt{d}$ 的整数 M 的个数有限。故存在 $0<|M|<1+2\sqrt{d}$ 使得有无穷组 (x,y) 满足

$$x^2-dy^2=M$$

(2) 由 (1) 可知，我们能找到两个不同的正整数解 $(x_1,y_1),(x_2,y_2)$ 使得

$$x_1\equiv x_2(\mathrm{mod}|M|),y_1\equiv y_2(\mathrm{mod}|M|)$$

我们注意到两个事实

$$x_1x_2-dy_1y_2\equiv x_1^2-dy_1^2\equiv M\equiv 0(\mathrm{mod}|M|)$$

$$x_1y_2-x_2y_1\equiv x_2y_2-x_2y_2\equiv 0(\mathrm{mod}|M|)$$

此时我们可以构造整数

$$x_0=\frac{x_1x_2-dy_1y_2}{M},y_0=\frac{x_1y_2-x_2y_1}{M}$$

并有

$$x_0^2-dy_0^2=\frac{(x_1x_2)^2-2Dx_1x_2y_1y_2+D^2(y_1y_2)^2-D((x_1y_2)^2-2x_1y_2x_2y_1+(x_2y_1)^2)}{M^2}=\frac{x_1^2M-Dy_1^2M}{M^2}=1$$

显然 $x_0\neq 0,y_0\neq 0$，故原方程至少有一组正整数解。

(3) 同原题假设给出 $x_1^2-dy_1^2=1$，此时可以验证

$$x_n^2-dy_n^2=(x_1+y_1\sqrt{d})^n(x_n-dy_n)=(x_1+y_1\sqrt{d})^n(x_1-y_1\sqrt{d})^n=(x_1^2-dy_1^2)^n=1$$

确实是原方程的解，从而有无穷多个解。接着就是要说明 $(x_n,y_n),n\geq 1$ 是所有的解，我们简记 $\varepsilon=x_1+y_1\sqrt{d},\overline{\varepsilon}=x_1-y_1\sqrt{d}$，若存在一组 $x_0^2-dy_0^2=1$ 使得 $x_0+y_0\sqrt{d}\neq\varepsilon^n$，则存在一个整数 $n>0$ 使得

$$\varepsilon^n<x_0+y_0\sqrt{d}<\varepsilon^{n+1}$$

从而有

$$1<u+v\sqrt{d}<\varepsilon,u+v\sqrt{d}=(x+y\sqrt{d})\overline{\varepsilon}^n$$

计算可得 $u^2-dv^2=1$，从而 (u,v) 是原方程的一组解，进一步有

$$0<u-v\sqrt{d}<\frac{1}{u+v\sqrt{d}}<1,2v\sqrt{d}=(u+v\sqrt{d})-(u-v\sqrt{d})>1-1=0$$

从而 $u>0,v>0$ 是一组正整数解，从而有 $u+v\sqrt{d}>\varepsilon$，矛盾，定理得证。

值得注意的是，Pell 方程的解是存在性的，并不像线性方程一样可以通过辗转相除法来得到特解，但正因为存在性，导致我们可以逐一验证，最简单的方法就是计算 $1+dy^2,y=1,2,\ldots$，直到出现第一个平方数，即可得到最小解，我们把它称为**基本解**，剩下的像定理一样计算就行了。当然有些情形，比如 $d=s(st^2+2),s>0,t>0$ 时，它的基本解就可以直接写出来

$$1+st^2+t\sqrt{d},x_1=1+st^2,y_1=t$$

相比于正 Pell 方程，负 Pell 方程 $x^2-dy^2=-1$ 就麻烦了很多，当然对于此方程同样也是“存在解”推出“存在

无穷个解”，其在 $d\equiv 0,3(\mathrm{mod}4)$ 时由简单的同余判定可知没有整数解，于是我们主要应该聚焦于 $d\equiv 1,2(\mathrm{mod}4)$ 的情形，但可惜我们对此并不明朗，我也就不讲太多以防混淆视听了，而来讨论一下更通用方程 $x^2-dy^2=n$ 的解的样貌。我们设

$$x_1^2-dy_1^2=n, s^2-dt^2=1$$

很容易发现

$$x_2^2-dy_2^2=n, x_2+y_2\sqrt{d}=(x_1+y_1\sqrt{d})(s+t\sqrt{d})$$

此时我们称 $x_1+y_1\sqrt{d}$ 与 $x_2+y_2\sqrt{d}$ **相结合**，并记为 $x_1+y_1\sqrt{d}\sim x_2+y_2\sqrt{d}$。容易验证，我们的“相结合”构成原方程所有整数解的一个等价关系。而在一个等价类中的所有正整数解的生成方式也是十分简单的

$$x_n+y_n=(x_0+y_0\sqrt{d})(s_0+t_0\sqrt{d})^n$$

其中 (s_0,t_0) 是 $x^2-dy^2=1$ 的基本解，(x_0,y_0) 是在这个等价类中使得 $x_0+y_0\sqrt{d}$ 最小的 $x^2-dy^2=n$ 的正整数解，我们也把它称为**基本解**。于是我们的问题是，原方程有几个等价类，至于等价类中的最小解就是枚举法了，而如果只有零个等价类则说明原方程没有整数解。

定理 7.3

(1) 设 p 是素数，如果方程

$$x^2-dy^2=\pm p$$

有解，则当 $p\mid 2d$ 时有一个等价类，$p\nmid 2d$ 时有两个等价类。

(2) 设 p 是奇素数，如果方程

$$x^2-dy^2=\pm 2p$$

有解，则当 $p\mid d$ 时有一个等价类，$p\nmid d$ 时有两个等价类。

♡

由于相关的研究并不多，所以我们就不证这玩意了，而且我们还注意到它要求有解作为前提，与我们的理念背道而驰，所以我单纯告诉你有这么一个事实就足够了，但我们有一个不等式限制定理来告诉我们枚举到什么程度就能判断方程没有整数解。

定理 7.4

设 $u_0+v_0\sqrt{d}$ 是不定方程 $u^2-dv^2=n$ 某个结合类的基本解，$x_0+y_0\sqrt{d}$ 是不定方程 $x^2-dy^2=1$ 的基本解，则

(1) 当 $n>0$ 时有

$$0\le v_0\le\frac{y_0\sqrt{n}}{\sqrt{2(x_0+1)}}, 0\le|u_0|\le\sqrt{\frac{1}{2}(x_0+1)n}$$

(2) 当 $n<0$ 时有

$$0\le v_0\le\frac{y_0\sqrt{-n}}{\sqrt{2(x_0+1)}}, 0\le|u_0|\le\sqrt{-\frac{1}{2}(x_0+1)n}$$

♡

我们举一个简单的例子 $u^2-2v^2=119$，首先可得 $x^2-2y^2=1$ 的基本解为 $3+2\sqrt{2}$，接着计算可得

$$0\le v_0\le\sqrt{\frac{119}{2}}<8$$

最后验算可得所有的基本解为 $11+\sqrt{2},-11+\sqrt{2},13+5\sqrt{2},-13+5\sqrt{2}$，从而有 4 个等价类。又例如 $u^2-82v^2=23$，$x^2-82y^2=1$ 的基本解为 $163+18\sqrt{82}, 0\le v_0\le\frac{9}{2}\sqrt{\frac{46}{41}}<5$，验算可知 $v_0=0,1,2,3,4$ 时均无解，从而原方程无整

数解。剩下的研究就比较零碎了，都是些没有一般性的结论，我们就稍微列举一下。

定理 7.5

(1)[Gauss] 设不定方程

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

满足 $D=b^2-4ac>0$，D 不是平方数，$\Delta=4acf+bdc-ae^2-cd^2-fb^2\neq 0$。如果方程有一组解，就必定有无数组解。

(2)[勾股方程] 设不定方程

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

满足 $(x,y)=1,x>0,y>0,z>0,2\mid x$，则它的全部整数解为

$$x=2mn,y=m^2-n^2,z=m^2+n^2,m>n>0,(m,n)=1$$

(3) 设不定方程

$$ax^2+by^2=cz^2$$

满足 $a>0,b>0,c>0,(a,b)=(a,c)=(b,c)=1,a,b,c$ 均无平方因子 (否则直接移进变量中)，则方程有一组满足 $(x,y,z)=1$ 的不全为零的整数解当且仅当

$$(\frac{-ab}{c})=1,(\frac{bc}{a})=1,(\frac{ac}{b})=1$$

♡

证明 (1) 对此的证明只需转化为 Pell 方程即可，我们构造变换

$$\begin{cases}X &= 2aDx+bDy+dD\\ Y &= Dy-2ae+bd\end{cases}$$

就能将原方程转化为

$$X^2-DY^2=M,M=4aD\Delta\neq 0$$

由于一组解 (x_0,y_0) 可以传递到一组解 (X_0,Y_0)，由 Pell 方程的特性可知有无数组解 (X_n,Y_n)，此时关键在于如何传递回去。容易发现 $x^2-Dy^2=1$ 有无数组解 $T_n+U_n\sqrt{D}$ 满足

$$T_n\equiv 1(\mathrm{mod}2aD^2),U_n\equiv 0(\mathrm{mod}2aD^2)$$

此时上述 Pell 方程的无穷个解为

$$X_n+Y_n\sqrt{D}=(X_0+Y_0\sqrt{D})(T_n+U_n\sqrt{D})$$

结合两个等式，计算可得

$$\begin{cases}X_0T_n+Y_0U_nD &= 2aDx+bDy+dD\\ X_0U_n+Y_0T_n &= Dy-2ae+bd\end{cases}$$

对 $2aD^2$ 取模可得

$$\begin{cases}2aD(x-x_0)+ & bD(y-y_0)\equiv 0(\mathrm{mod}2aD^2)\\ & D(y-y_0)\equiv 0(\mathrm{mod}2aD^2)\end{cases}$$

从而有

$$x=x_0+Dk_2-bDk_1,y=y_0+2aDk_1,k_1,k_2\in\mathbb{Z}$$

(2) 正向验算是简单的

$$x^2+y^2=(2mn)^2+(m^2-n^2)^2=(m^2+n^2)=z^2$$

所以我们主要要说明解都是上面的形式，设 x,y,z 是一组解，由于 $2\mid x$ 是偶数且 $(x,y)=1$，故 y,z 均是奇数，从

而 $\frac{z-y}{2}, \frac{z+y}{2}$ 是整数，注意到

$$\frac{z-y}{2}\frac{z+y}{2}=\frac{z^2-y^2}{4}=(\frac{x}{2})^2$$

且有 $(\frac{z-y}{2}, \frac{z+y}{2})=1$，故它俩都是平方数，从而存在 $m>n>0$ 使得

$$\frac{z+y}{2}=m^2, \frac{z-y}{2}=n^2, x=2mn$$

化简即可得到我们想要的结果

$$x=2mn, y=m^2-n^2, z=m^2+n^2, m>n>0, (m,n)=1$$

(3) 这是希尔伯特符号 (Hilbert symbol)，详细可以参考这本书 [48] 的 2.3 和 2.6 节，我们就不讨论了。

三次不定方程

显然连二次不定方程都还没有完全搞懂，我们又怎么敢去挑战三次不定方程，当然敢啦，反正我们又不会去讨论所有的情况，而是像 Pell 方程一样，十分特殊的下面这种情况 (可以称其为 **Mordell 方程**)

$$y^2=x^3+k, k\in\mathbb{Z}$$

其实使用瞪眼法可知，当 $k=0$ 时，方程仅有解 $(x,y)=(0,0),(1,1),(1,-1)$；当 $K\neq 0$ 时，其是一个椭圆曲线，根据 Siegel 定理可知整数解的个数是有限的 (有没有其实并不确定)，不过我们还是决定用初等数论的方法处理一些特殊情况。

定理 7.6 (Mordell 方程)

设二元三次 Mordell 方程为

$$y^2=x^3+n, n\in\mathbb{Z}$$

则在 n 满足下面任一条件下，方程没有整数解

(1)$n=(4b-1)^3-4a^2, p\mid a\Rightarrow p\not\equiv 3(\mathrm{mod}4)$

(2)$n=(4b+2)^3-(2a+1)^2, p\mid 2a+1\Rightarrow p\not\equiv 3(\mathrm{mod}4)$

(3)$n=2b^2-a^3, a\equiv 2,4(\mathrm{mod}8), b\equiv 1(\mathrm{mod}2), p\mid b\Rightarrow p\not\equiv\pm 3(\mathrm{mod}8)$

(4)$n=-a^3-2b^2, a\equiv 4(\mathrm{mod}8), b\equiv 1(\mathrm{mod}2), p\mid b\Rightarrow p\not\equiv 5,7(\mathrm{mod}8)$

(5)$n=3b^2-a^3, a\equiv 1(\mathrm{mod}4), b\equiv\pm 2(\mathrm{mod}6), p\mid b\Rightarrow p\not\equiv\pm 5(\mathrm{mod}12)$

♡

我就直说，上面这些都是用同余法推出了的，可以看到越后面约束越多，所以同余法终究是无法到达最后结果的。但我们可以借助代数数论的方法来得到解的可能结构，从而在一定条件下进行枚举，进而搞清 Mordell 方程所有的解，而不仅只是停留于没有整数解的阶段。

定理 7.7

设二元三次 Mordell 方程为

$$y^2=x^3+k, k\in\mathbb{Z}$$

(1) 若 $k<-1$ 无平方因子，$k\equiv 2,3(\mathrm{mod}4)$，$\mathbb{Q}(\sqrt{k})$ 的类数为 h 且 $3\nmid h$，则方程的解满足

$$x=a^2-k, k=\pm 1-3a^2, a\in\mathbb{Z}$$

(2) 若 $k>1$ 无平方因子，$k\equiv 2,3(\mathrm{mod}4)$，$\mathbb{Q}(\sqrt{k})$ 的类数为 h 且 $3\nmid h$，设 $u^2-kv^2=1$ 的基本解为 $x_0+y_0\sqrt{k}$，若原方程有整数解，则有等式

$$x_0(3a^2b+kb^3)\pm y_0(a^2+3kab^2)=1, a, b\in\mathbb{Z}$$

(3) 若 $k<0$ 无平方因子，$k\not\equiv -7, k\equiv 1(\mathrm{mod}8)$，$\mathbb{Q}(\sqrt{k})$ 的类数为 $h=3$，$k=-a^2-b^2, b\equiv 2,3(\mathrm{mod}4), p\mid a\Rightarrow p\not\equiv 3(\mathrm{mod}4)$，则方程没有整数解。

♡

证明 由于上述结论证法都是类似的，即在分式理想环中考虑素理想，就像在整数中考虑素数一样，而且我们在前面证明 Kummer 定理其一时也使用了类似的思想，所以我们就只证明 (1) 来让读者稍微回想一下理想唯一分解的感觉。由条件可知，$2\nmid x$，从而可以得到理想方程为

$$(y+\sqrt{k})(y-\sqrt{k})=(x)^3$$

我们记 $K=\mathbb{Q}(\sqrt{k})$，则 $y+\sqrt{k},y-\sqrt{k}\in O_K$ 是代数整数，设它们的公因数为 $d=(y+\sqrt{k},y-\sqrt{k})$，则有

$$d\mid((y+\sqrt{k})+(y-\sqrt{k}),(y+\sqrt{k})-(y-\sqrt{k}))=(2y,-2\sqrt{k})=2(y,-\sqrt{k})$$

此时必有 $(y,-\sqrt{k})=1$，否则有 $N(a)>1,a\mid(y,-\sqrt{k}),a\in O_K$，即有 $y=ab_1,-\sqrt{k}=ab_2,b_1,b_2\in O_K$，取范以后为 $y^2=N(a)N(b_1),-k=N(a)N(b_2)$。由于 $N(a)>1$，故存在 $p\mid y^2,p\mid k$，于是 $p\mid x$，进而 $p^2\mid k$，矛盾，故 $(y,-\sqrt{k})=1$ 成立，于是我们有

$$d\mid 2(y,-\sqrt{k})\Rightarrow d\mid 2$$

当 $K=\mathbb{Q}(\sqrt{-2})$ 时 2 有分解 $2=\sqrt{-2}(-\sqrt{-2})$，在 $K\neq\mathbb{Q}(\sqrt{-2})$ 时 2 是 K 的不可约元，因此若 $d\neq 1$，则只存在两种情况 $d=2$ 或 $d=\sqrt{-2}$。但无论如何均有 $2\mid y^2-k=x^3$，与 $2\nmid x$ 矛盾，从而只能有

$$d=1\Rightarrow(y+\sqrt{k},y-\sqrt{k})=1$$

又由于单位群为 $U_K=\{-1,1\}$，故由代数整数环 O_K 的素理想唯一分解定理可得

$$y+\sqrt{k}=(a+b\sqrt{k})^3,y-\sqrt{k}=(a-b\sqrt{k})^3,x=a^2-kb^2$$

我们消去一个未知数可得 $b(3a^2+kb^2)=1$，从而 $b=\pm1$，即有

$$x=a^2-k,k=\pm1-3a^2,a\in\mathbb{Z}$$

利用这些定理，我们可以轻松地求出各种不定方程的整数解，例如 $y^2+2=x^3$ 的整数解为

$$-2=1-3\times1^2,x=1^2-(-2)=3,y=\pm\sqrt{3^3-2}=\pm5$$

方程 $y^2+13=x^3$ 的整数解为

$$-13=-1-3\times2^2,x=2^2-(-13)=17,y=\pm\sqrt{17^3-13}=\pm70$$

方程 $y^2=x^3+34,y^2=x^3-31$ 没有整数解等等。就像前面的 Pell 方程一样，我们同样可以给出一个估计的区间来辅助你枚举出解的上界

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

虽然这个上界大得离谱，但至少意味着，我们可以在有限的步骤内枚举出 Mordell 方程的所有整数解。这节的最后，我们同样地来说一些零碎的结论。

定理 7.8

(1) 不定方程

$$ax^3+ay^3+bz^3=bc^3,abc\neq0$$

除了平凡解 $x+y=0,z=c$ 外，还有无穷多组整数解。

(2) 不定方程

$$x^3+y^3+z^3+w^3=n$$

如果有一组解使得 $-(x+y)(z+w)>0$ 不是平方数且 $x\neq y,z\neq w$，则方程有无数组解。

♡

证明 (1) 对此我们做代换 $(x,y,z)\to(u,v,t)$ 如下

$$z=c+t(x+y),x+y=u,x-y=v$$

带入化简可得

$$(a+4bt^3)u^2+12bct^2u+12bc^2t+3av^2=0$$

这里的 t,u,v 都是变元，我们不妨先假设 $t=-abk^2,k\neq 0$ 的形式存在，则方程变成了一个二元二次不定方程，计算有

$$D=0-12a(a-4a^3b^4k^6)=12a^2(4a^2b^4k^6-1)>0,\Delta=144abc^2t(a+bt^3)\neq 0$$

通过限制 $k=3s,s=1,2,...,a+bt^3\neq 0$，可知存在 k，进而存在 t，使得上述不定方程有无数组 u,v 解。设 $a=2^\alpha a_0,2\nmid a_0$，此时同样能找到 $2^{\alpha+1}\mid t$，于是可得

$$au^2\equiv -3av^2(\mathrm{mod}2^{\alpha+1})$$

化简可得

$$u\equiv v(\mathrm{mod}2)$$

因此我们可以找到无数组整数解

$$x=\frac{u+v}{2},y=\frac{u-v}{2},z=c+tu$$

(2) 设存在的一组解为 x_0,y_0,z_0,w_0，并做变换

$$x=x_0+X,y=y_0-X,z=z_0+Y,w=w_0-Y$$

于是可得方程

$$(a+b)X^2+(a^2-b^2)X+(c+d)Y^2+(c^2-d^2)Y=0$$

同样我们视为一个二元二次不定方程，计算有

$$D=-(a+b)(c+d)>0,\Delta=-(a+b)(c+d)((c+d)(c-d)^2+(a+b)(a-b)^2)\neq 0$$

从而我们有无穷多组解 X,Y，进而原方程有无穷多组解。

我们看到了 Pell 方程在证明中所起到的作用是无穷无尽的，这也是为什么我们在前面把 Pell 方程作为主要的探究对象。读者应该还注意到了，所谓无穷解并无需说明它的所有解有无穷个，而是部分解有无穷个，例如 $f(x,y,z)=0$ 有无穷个解，我们同样可以进行某种限制 $g(z)=0$，来说明在这个条件下 $f(x,y,z)=0$ 依旧有无穷个解，或更直接点对某个 $z_0\in\mathbb{Z},f(x,y,z_0)=0$ 有无穷个解，这样做的好处，无非就是我们曾反复提过的，条件更多了，更好操控了。既然，二元二次不定方程这么好用，我们是否也要讨论一下像下面这样的二元三次不定方程

$$ey^2=ax^3+bx^2+cx+d(a\neq 0),ax^3+bx^2y+cxy^2+dy^3=0$$

前者本质属于椭圆曲线的研究，使用初等方法讨论有些吃力不讨好，不过椭圆曲线的化简理论也说明了它有着最简式 $u^2=v^3-g_2v-g_3$，当 $g_2=0$ 时就是 Mordell 方程，不过就算 $g_2\neq 0$ 也没什么有趣的结论，无非就是之前所说的椭圆曲线整点个数有限的结论，又或者是些十分具体的例子。例如，椭圆曲线 $6y^2=x(x+1)(2x+1)$ 的所有整点为

$$(x,y)=(0,0),(-1,0),(1,\pm 1),(24,\pm 70)$$

椭圆曲线 $y(y+1)=x(x+1)(x+2)$ 的所有整点为

$$x=-1,-2,0,1,5$$

椭圆曲线 $6y^2=(x+1)(x^2-x+6)$ 的所有整点为

$$x=-1,0,2,7,15,74,767$$

好了好了，我们不列了。对于后面的齐次方程，有些比较奇怪的结论，就不说出来污染你的眼睛了，对于三次我们就止步于此了，更何况二次都还没完全搞懂，又怎么能高瞻远瞩呢？你说是吧。

四次不定方程

到了四次，我们基本就找不到能叫出名字的方程了，初等方法也快筋疲力竭了，其迫切地要我们找到一个新的方法，或许它是算术代数几何，但我们还不能脱离框架，还是得硬生生地拿出著名的几个结果来。

定理 7.9

(1) 不定方程

$$x^4+y^4=z^2$$

没有满足 $xy\neq 0$ 的整数解。

(2) 不定方程

$$x^4-y^4=z^2,(x,y)=1$$

没有满足 $xyz\neq 0$ 的整数解。

(3) 不定方程

$$x^4-6x^2y^2+y^4=z^2,(x,y)=1$$

没有满足 $xy\neq 0$ 的整数解。

(4) 不定方程

$$y^2=5x^4+1$$

仅有解 $x=0,\pm 2$

$$y^2=5x^4-1$$

仅有解 $x=\pm 1$

$$y^2=5x^4+4$$

仅有解 $x=0,\pm 1,\pm 12$

$$y^2=5x^4-4$$

仅有解 $x=\pm 1$。

♡

证明 (1) 我们假设有一组解 $x>0,y>0,z>0,(x,y)=1$ 且 z 是解中最小的，由 mod4，我们可以假定 $2\nmid x,2\mid y$，由之前讨论的勾股方程可知

$$x^2=a^2-b^2,y^2=2ab,z=a^2+b^2,a,b\in\mathbb{Z}$$

更进一步，我们可以设 $2\nmid a,2\mid b$，并对 $x^2+b^2=a^2$ 应用勾股方程可得

$$x=p^2-q^2,b=2pq,a=p^2+q^2,p,q\in\mathbb{Z}$$

在此基础上，我们看另一个方程

$$y^2=2ab=4pq(p^2+q^2)$$

由于 $(p,p^2+q^2)=(q,p^2+q^2)=1$，故由算术基本定理可得

$$p=r^2,q=s^2,p^2+q^2=z_0^2$$

从而我们有一组新的解

$$r^4+s^4=p^2+q^2=z_0^2$$

但 $z_0<z$ 与最小性矛盾了，从而原方程无满足 $xy\neq 0$ 的整数解。

(2) 方法与 (1) 是一样的，不讲了。

(3) 我们假设 $x > y > 0$，并将原方程化为

$$(x^2 - y^2 + z)(x^2 - y^2 - z) = 4x^2y^2$$

设 $(x^2 - y^2 + z, x^2 - y^2 - z)$，则有

$$d \mid (x^2 - y^2 + z) - (x^2 - y^2 - z) = 2z$$

继续借助奇偶关系可得，$2 \nmid x, 2 \mid y$ 或 $2 \mid x, 2 \nmid y$，不论如何均有 $2 \nmid z$。假设 $p \mid d$ 是一个奇素数，则有 $p \mid z$，由 $p \mid x^2 - y^2 + z$ 可得 $p \mid xy$，于是由 $p \mid x^2 - y^2$ 可得 $p \mid x, p \mid y$，这与 $(x, y) = 1$ 矛盾，从而 $(d, z) = 1$，即 $d = 2$。因此我们由唯一分解定理可得

$$x^2 - y^2 + z = 2a^2, x^2 - y^2 - z = 2b^2, xy = ab, a > 0, b > 0, (a, b) = 1$$

从而得到

$$(x^2 + y^2)^2 = a^4 + 6a^2b^2 + b^4, ab \neq 0$$

对方程 $x^4 + 6x^2y^2 + y^4 = z^2$ 做变换 $x = u + v, y = u - v, (x, y) = 1$，则方程形式不变

$$u^4 + 6u^2v^2 + v^4 = z^2$$

它们形成了解 $(x, y) \leftrightarrow (u, v)$ 的一一对应，故方程仅有解 $(x, y) = (\pm 1, 0)$ 或 $(x, y) = (0, \pm 1)$，没有 $xy \neq 0$ 的解，矛盾。

(4) 我们就看第一个后面都是一样的，Pell 方程 $u^2 - 5v^2 = 1$ 的基本解为 $\varepsilon = 9 + 4\sqrt{5}$，故分解给出

$$y + x^2\sqrt{5} = (9 + 4\sqrt{5})^n = (\frac{1 + \sqrt{5}}{2})^{6n}, y - x^2\sqrt{5} = (\frac{1 - \sqrt{5}}{2})^{6n}$$

于是我们有

$$2x^2 = \frac{(\frac{1+\sqrt{5}}{2})^{6n} - (\frac{1-\sqrt{5}}{2})^{6n}}{\sqrt{5}} = F_{6n}$$

这里的 $F_n = F_{n-1} + F_{n-2}, F_1 = F_2 = 1$ 是斐波那契数列，根据斐波那契数列分布规律

$$F_n = 2x^2 \Rightarrow n = 0, \pm 3, 6, x = 0, \pm 1, \pm 2$$

可知原方程仅有整数解 $x = 0, \pm 2$。

实际上，四次不定方程中最爱讨论的是类似于 Pell 方程的一些例子，比如 $a^2X^4 - dY^2 = M$ 或 $X^2 - da^2Y^4 = M$，它们均是 Pell 方程 $x^2 - dy^2 = M$ 中带入平方因子项 $x = aX^2$ 或 $y = aY^2$ 后的结果。另一个是与我们前面所讲过的 $x^4 + kx^2y^2 + y^4 = z^2, xy \neq 0$ 的类齐次方程和更一般的齐次方程 $ax^4 + by^4 + cz^4 = dw^4, abcd \neq 0$，或许这些玩意我们怎么也研究不完，但正是如此，我们才能在数学里不断地寻找乐趣，来使一生变得充实。

不水次数了

我来稍微说句实话吧，不定方程其实没啥好讲的，对于通论其实本质内容不多，介绍一下初等方法、再介绍一下像代数数论那样稍微进阶一些的方法、最后配上一个研究成果丰富且还有开放问题的 Pell 型方程就基本差不多了，剩下的干什么呢？就是不断地收集各种具体方程的解法，仅此而已 (注：目前并未发现有大量收集不定方程的工具书 handbook 或收录集 collection)。比如我所看到的两本比较有代表性的书，[2] 就是使用和我所说的类似的结构，而 [10] 则是把代数数论的方法发挥到了极致讨论了各种次数的代数域。所以，关键在于，哪些不定方程值得出来说道一番，值得被挑选出来？以我的观点来看，是那些解决方法具有深刻内涵的方程，比如费马大定理和 Pell 方程，又或者是十分流行的有很多人研究的方程，比如由埃及分数引出来的 Erdős-Straus 猜想..... 在这一节中，我们将讨论，我们没有讨论过的且我觉得比较有趣的方程，这纯属个人兴趣。

命题 7.1

(1) 不定方程

$$x^m - y^n = 1, m > 1, n > 1$$

没有 $m = 2, x = 3, y = 2, n = 3$ 以外地正整数解。

(2) 如果

$$a^2 + b^2 = c^2, a^x + b^y = c^z$$

则有 $x = y = z = 2$。

(3) 不定方程

$$x^2 + 7 = 2^n$$

仅有以下五组解

$$(x, n) = (1, 3), (3, 4), (5, 5), (11, 7), (181, 15)$$

(4) 不定方程

$$x^2 = 2y^4 - 1$$

仅有以下两组解

$$(x, y) = (1, 1), (239, 13)$$

(5) 设 $f(x) = \sum_{i=0}^{n} a_i x^i, n > 2$ 是整系数不可约多项式，则不定方程

$$\sum_{i=0}^{n} a_i x^i y^{n-i} = c, c \in \mathbb{Z}$$

仅有有限个整数解 (x, y)。

命题 7.2 (Erdős-Straus 猜想)

不定方程

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

对所有地 $n > 1$ 都有正整数解。

7.2 代数曲线

前面一节只能说是引入，来告诉你有不定方程这么一回事，再告诉你一些常规的初等解法，而不定方程发生实质性改变的，果然还是代数几何中的有理点问题，或者称为算术代数几何。而在算术代数几何中，理论较为完整的就是代数曲线了，所以这也是我们将讨论的重点，所谓代数曲线指一维代数簇，而代数簇又是指某种性质的概形，我想你大概也不可能去看现代代数几何的书籍，又或者去看我上次写的总结性文章 [45]，因此，我们决定采用最传统的定义方式，所谓域 k 上的**代数曲线**指

$$E = \{(x, y) \in k^2 \mid f(x, y) = 0\} \cup \{\infty\}, f(x, y) \in k[x, y]$$

这里包含了两方面的约化，一是代数曲线的嵌入定理，二是射影方程与仿射方程的转化，前者是跳过代数几何的核心故省略，后者则是无穷远点的由来，实际上对任意一个代数曲线，我们都有齐次转化的方法

$$\sum_{i=1}^{n} c_i x^{a_i} y^{b_i} = 0 \to \sum_{i=1}^{n} c_i (\frac{x}{z})^{a_i} (\frac{y}{z})^{b_i} = 0$$

这样做可以使得在射影方程零点中自带上无穷远点 $(x,y,z)=(a,b,0)$，而我们最开始给出的也叫做仿射方程，从研究的便利性来看，我们通常都是使用仿射方程并心里默认它有一个无穷远点。代数曲线的点 $(x_0,y_0)\in E$ 是**奇点**指它满足

$$\frac{\partial f}{\partial x}|_{(x,y)=(x_0,y_0)}=0,\frac{\partial f}{\partial y}|_{(x,y)=(x_0,y_0)}=0$$

没有奇点的代数曲线称为**光滑的**，设曲线的奇点个数为 δ，定义方程 $f(x,y)=0$ 的次数为 d，我们就能计算出代数曲线的**亏格**为

$$g=\frac{(d-1)(d-2)}{2}-\delta$$

在我们当前的讨论语境下，可以直接视其为亏格的定义。对于代数曲线，读者知道了上面的内容，我觉得就差不多了。

有理数解与整数解的界线

在正式内容开始前，我们必需理解有理点和整数点的区别，其实要讲的东西并不多，无非就是几个关键点。第一，**齐次方程的有理数解与整数解等价**，这个其实毫无疑问，无非就是分母的扩大与缩小；第二，**齐次方程的整数解等价于对应仿射方程的有理解**，这个其实就是我们之前射影方程与仿射方程的转化过程；**仿射方程 $f(x,y)=0$ 的整数解与有理解是两个问题**，且后者的难度更大。在这一节中我们先讨论相对简单的整数解问题，其对应着 Siegel 定理，而有理解对应着 Mordell-Weil 定理和 Faltings 定理，这是我们后面两部分讨论的重点。这节的主要内容均来自 [16]，其给出了我们这三个核心定理的证明。亏格 $g\in\mathbb{N}$ 是代数曲线很好的一个分类方式，唯一恼人的点是那个计算式中的奇点 δ，它使得我们的曲线可能有无穷多种情况，好在代数曲线的**奇点消解**已经是一个被解决的问题：每个代数曲线都双有理等价于一个光滑代数曲线。双有理等价并不一定是一个双射，而是可以来回抵消的有理映射，它只能保持有理点，不能保持整点，但这又何妨呢？反正亏格 $g\geq 2$ 时的整点问题可以直接被 Faltings 定理取代，一切的特殊与巧合都发生在 $g=1$ 的**椭圆曲线**上，那 $g=0$ 的代数曲线呢？此时我们可以得到 $d=1$ 或 $d=2$，前者是完全清楚的一次不定方程，后者由于我们假设了光滑的条件，其就是我们之前所证明的定理 7.5(1)，即 **$g=0$ 的代数曲线上要么没有整点要么有无穷多个整点**，并且**将整点替换为有理点也是一样的**，故我们实际只需探讨 $g\geq 1$ 的情况。

$$g=0\Rightarrow ax^2+bxy+cy^2+dx+ey+f=0$$

由于整数环 $\mathbb{Z}$ 不是一个域，所以我们有必要引入一些其它的术语来描述曲线 E 的 K-点问题。设 $K/\mathbb{Q}$ 是数域，M_K 是 K 的所有素点构成的集合，$S\subset M_K$ 是包含阿基米德素点的有限集，K 的 **S-整环**为

$$K_S=\{x\in K\mid |x|_p\leq 1,\forall p\in M_K-S\},s=|S|$$

$C=\{(x,y)\in\overline{K}^2\mid f(x,y)=0,f(x,y)\in K[x,y]\}$ 是光滑代数曲线，$K(C)=\mathrm{Frac}(K[x,y]/(f(x,y)))$ 为坐标环的分式域，即代数曲线上的有理多项式环，$C(K)=\{(x,y)\in K^2\mid (x,y)\in C\}$ 表示代数曲线的 **K-点**，于是要证明的定理为

定理 7.10 (Siegel)

若代数曲线 C 的亏格满足 $g\geq 1$，则

$$\forall f\in K(C)\text{非常数},\{P\in C(K)\mid f(P)\in K_S\}$$

是有限集。♡

对于上面的定理你可能觉得有点懵，所以我们把它简化一下，取 $K=\mathbb{Q},S=\{\infty\}$，$|v|_p\leq 1$ 表示 p-进绝对值

小等一，即 p-进赋值大等零，故每个素因子的指数大等零，即 $K_S=\mathbb{Q}_{\{2,3,5,7,11,\dots\}}=O_{\mathbb{Q}}=\mathbb{Z}$，$C(\mathbb{Q})$ 表示代数曲线的有理点，再取 $f(x,y)=x, f(x,y)=y$，故定理给出 $\{(x,y)\in C\mid x\in\mathbb{Z}\}$ 和 $\{(x,y)\in C\mid y\in\mathbb{Z}\}$ 是有限集，从而代数曲线上的整点是有限的，我们所给的定理其实更强一些。对任意一点 $P\in K^2$，我们定义它的**高度**为

$$H_K(P)=\prod_{p\in M_K}\max\{|x_0|_p,|x|_p,|y|_p\},P=(\frac{x}{x_0},\frac{y}{x_0}),x_0,x,y\in O_K,(x,x_0)=(y,y_0)=1$$

高度理论是一个内容十分丰富的理论，它还能用来解决 Mordell-Weil 定理，一时半伙我们是搞不定的，对于里面的大量性质，我就只能默认读者是知道的了，实在搞不懂的话，就找一些像“An Introduction to Height Function”之类标题的讲义，还不行的话就去看椭圆曲线相关的书籍吧。通过之前在估计中讲过的“Roth 定理”和“高度理论”，我们就能得到证明所需要的一个核心估计定理。

定理 7.11

设 $t\in K(C)$ 是在 f(上一个定理中“$\forall f\in K(C)$非常数”的 f) 的极点和零点处均非分歧的有理函数，则对任意的正数 $\rho>0$ 存在常数 $c=c(f,t,C,\rho,S)>0$ 满足

$$\prod_{p\in S}\min\{|f(P)|_p,1\}\geq\frac{c}{H_K(t(P))^\rho}$$

♡

你说证明？说到底，我只想讨论有理点问题，这部分内容巴不得快点跳过，奈何它在同一个体系下，所以不得不提一嘴，既然要证明的话，我们就用上面这个估计定理来轻松推出 Siegel 好了，这也算是一种证明，难道不是吗？

证明 我们假设 $\{P\in C(K)\mid f(P)\in K_S\}$ 是无限集，使用反证法。我们设 $t\in K(C)$ 是满足上面估计定理条件的有理函数，并且选定常数为

$$\rho=\frac{\deg f}{2\deg t}$$

由于 $\frac{1}{f}$ 与 f 的极零点情况一致，我们可以对其同样运用定理可得

$$\exists c_1>0,\prod_{p\in S}\min\{|\frac{1}{f(P)}|_p,1\}\geq\frac{c_1}{H_K(t(P))^\rho}$$

通过代数化简可得

$$H_K(t(P))^\rho\geq c_1\prod_{p\in S}\max\{|f(P)|_p,1\}\overset{f(P)\in R_S}{=}c_1\prod_{p\in M_K}\max\{|f(P)|_p,1\}=c_1H_K(f(P))$$

定义 $h=\ln H_K$，并对上面不等式取对数可得

$$\rho h(t(P))\geq h(f(P))-c_2,\forall P\in C(K),f(P)\in R_S$$

考虑 ρ 的定义可得

$$\frac{\deg f}{2\deg t}\geq\frac{h(f(P))}{h(t(P))}-\frac{c_2}{h(t(P))}$$

由于 $P\in C(K),f(P)\in R_S$ 有无限个点，我们令 $h(t(P))\to\infty$ 则有

$$\frac{\deg f}{2\deg t}\geq\lim_{\substack{P\in C(K)\\h(t(P))\to\infty}}\frac{h(f(P))}{h(t(P))}-0=\frac{\deg f}{\deg t}$$

矛盾，从而定理得证。

上面证明中，除了极限 $\lim\limits_{\substack{P\in C(K)\\h(t(P))\to\infty}}\frac{h(f(P))}{h(t(P))}=\frac{\deg f}{\deg t}$ 是高度理论的结果外，其余的都是基本的代数和极限内容，如果你为这个证明感到惊讶的话，大概率是因为上面的高度估计定理过于奇妙了。

椭圆曲线与 Mordell-Weil 定理

整点问题就那么点，$g \geq 1$ 时都是有限个，稍微复杂点的是有理点问题，其要根据亏格分 $g = 1$ 和 $g > 1$ 两种情况讨论，而其中 $g = 1$ 最为复杂，这也是为什么椭圆曲线在代数曲线中几乎具有统治性的地位。

定理 7.12 (Mordell-Weil 定理)

设 A 是数域 K 上的阿贝尔簇，$A(K)$ 表示 A 的 K-点，则 $A(K)$ 是有限生成交换群。

♡

这里多了一个阿贝尔簇的概念，它表示同时是代数群的射影簇，它可以自带附上光滑交换的条件，我们不需要这么复杂的结果，考虑最朴实的有理数域 $K = \mathbb{Q}$ 和椭圆曲线 $E : y^2 = x^3 + Ax + B, A, B \in \mathbb{Z}$，即一维阿贝尔簇就足够了。至于椭圆曲线 E 上的交换群结构主要指 $\{(x, y) \in \overline{K}^2 \in f(x, y) = 0\} \cup \{\infty\}$ 上的群结构，其中 $f(x, y)$ 是三次不可约多项式，但这不属于 Mordell-Weil 定理，而是来自参数化

$$E : y^2 = 4x^3 - g_2(\tau)x - g_3(\tau), (x, y) = (\mathscr{C}(z), \mathscr{C}'(z)), \mathscr{C}(z) = z^{-2} + \sum_{m=1}^{\infty}(2m + 1)G_{2m+2}(\tau)z^{2m}, E \leftrightarrow \mathbb{C}/\Lambda(\tau)$$

其将椭圆曲线上的某种加法运算 (即连线的第三交点关于 y 的对称点) 变成了复平面 $\mathbb{C}$ 上某块平行四边形上的复数加法

$$\mathscr{C}(z_1 + z_2) = \frac{1}{4}(\frac{\mathscr{C}'(z_1) - \mathscr{C}'(z_2)}{\mathscr{C}(z_1) - \mathscr{C}(z_2)})^2 - \mathscr{C}(z_1) - \mathscr{C}(z_2)$$

从式子很容易看出，椭圆曲线上的加法是保持有理数的，实际也可以联立方程使用韦达定理来推导，因此 Mordell-Weil 定理主要证明“有限生成”这个结论，而“有限生成”这个结论主要由下面的证明原理来完成。

定理 7.13 (下降定理)

设 G 交换群，$q : G \to \mathbb{R}$ 是一个二次型，满足集合

$$\forall C, \{x \in G \mid q(x) \leq C\}$$

是有限集。如果存在 $m \geq 2$ 使得 G/mG 是有限的，则 G 是有限生成的。

♡

注 我们稍微解释一下什么叫做交换群 G 上的二次型 $q : G \to \mathbb{R}$，前后两个集合的运算并不兼容，因此我们只能通过性质来进行定义。设 A, B 都是交换群，如果映射 $q : A \to B$ 满足

$$\forall x, y, z \in A, q(x + y + z) - q(x + y) - q(y + z) - q(z + x) + q(x) + q(y) + q(z) = 0, q(-x) = q(x)$$

则称 q 是一个**二次型**。如果去掉后面偶函数的条件，则称为**二次函数**，因此此处其实有齐二次的含义。在实际运用的时候我们会使用一个更强的条件，即**平行四边形法则**

$$\forall x, y \in A, q(x + y) + q(x - y) = 2q(x) + 2q(y)$$

它能推出上面二次型的条件，并且此时我们能定义一个双线性函数

$$\langle x, y\rangle = \frac{q(x + y) - q(x) - q(y)}{2}, \langle x, x\rangle = q(x)$$

从而与线性代数中的二次型进行了成功的对接，所以我们以后采用平行四边形法则来作为二次型的定义。二次型有些简单的性质，我们后面要用到

$$q(x + x) + q(x - x) = 2q(x) + 2q(x) \Rightarrow q(2x) = 2^2q(x) \Rightarrow q(mx) = m^2q(x)$$

证明 由已知，我们设 $G/mG = \{[g_1], ..., [g_s]\}, g_i \in G$。若存在 $g \in G$ 使得 $q(g) < 0$，则有 $q(kg) = k^2q(g) < 0, k = 1, 2, ...$。若 g 是无限阶元，则 $\{x \in G \mid q(x) \leq 0\}$ 是无限集，矛盾；若 g 是有限阶元，则存在 k 使得 $kg = 0$ 进而 $q(kg) = q(0) = 0$，矛盾，因此 $\forall g \in G, q(g) \geq 0$，此时我们定义

$$|x| = \sqrt{q(x)}, c_0 = \max_{1 \leq i \leq s} |g_i|, S = \{x \in G \mid |x| \leq c_0\}$$

我们来证明 G 由集合 S 有限生成，即 $G = \langle S\rangle$。

任取 $x_0 \in G$，若均有 $x_0 \in S$，则有 $G = S$ 是有限群，得证。否则，我们设 $x_0 \notin S$，即 $|x_0| > c_0$，考虑它在 G/mG 中的像，则有 $\exists i, x_1 \in G, x_0 = g_i + mx_1$，计算可得

$$m|x_1| = |x_0 - g_i| \leq |x_0| + |g_i| < 2|x_0| \leq m|x_0| \Rightarrow |x_1| < |x_0|$$

若 $x_1 \in S$，则 x_0 可以由 S 中的元素线性表示，得证。否则，同样使用 $\exists j, x_2 \in G, x_1 = g_j + mx_2$，则可以得到一些列的

$$|x_0| > |x_1| > |x_2| > \ldots$$

由于 $\{x \in G \mid |x| \leq |x_0|\}$ 是有限集，故最后一定会停在 S 中，从而定理得证。

于是为了完成Mordell-Weil 定理的证明，我们需要做两件事，一是找一个二次型 q 使得 $\forall C, \{x \in G \mid q(x) \leq C\}$ 是有限集，二是找一个 $m \geq 2$ 使得 G/mG 是有限的。我们先来完成第一件事，这属于高度理论的内容，这个高度函数是

$$q(P) = \lim_{n\to\infty} \frac{1}{4^n} \ln H(2^n P), H(P) = H_K(P)^{\frac{1}{[K:\mathbb{Q}]}}$$

其相关性质的讨论是极其麻烦的，所以我们想着把 $q : G \to \mathbb{R}$ 为二次型的条件弱化成下面的两条

1 $\forall Q \in G, \exists C_1, \forall P, q(P+Q) \leq 2h(P) + C_1$

2 $\exists m \geq 2, C_2, \forall P, q(mP) \geq m^2 q(P) - C_2$

由二次型的性质是可以推出上面的性质的，我们来说明将条件替换以后，也同样足够推出有限生成的性质，不过此时我们的生成元集合 S 就不太清晰了。

证明 [下降定理证明改] 同样地设 $G/mG = \{[g_1], \ldots, [g_r]\}, g_i \in G$，并任取元素 $x_0 \in G$，然后不断地在 G/mG 内进行陪集分解

$$x_0 = g_{i_1} + mx_1, x_1 = g_{i_2} + mx_2, \ldots, x_{n-1} = g_{i_n} + mx_n, 1 \leq i_n \leq r$$

对于每个 $1 \leq j \leq n$ 我们有不等式估计

$$q(x_j) \leq \frac{1}{m^2}(q(mx_j) + C_2) = \frac{1}{m^2}(q(x_{j-1} - g_{i_j}) + C_2) \leq \frac{1}{m^2}(2q(x_{j-1}) + C_1 + C_2)$$

我们把上面的式子不断递归可得

$$q(x_n) \leq (\frac{2}{m^2})^n q(x_0) + \frac{(\sum_{i=1}^{n} 2^{i-1})(C_1 + C_2)}{m^2} < (\frac{2}{m^2})^n q(x_0) + \frac{C_1 + C_2}{m^2 - 2} \leq \frac{q(x_0)}{2^n} + \frac{C_1 + C_2}{2}$$

当 $n \to \infty$ 足够大时，有

$$q(x_n) \leq 1 + \frac{C_1 + C_2}{2}$$

相应的生成表示为

$$x_0 = m^n x_n + \sum_{j=1}^{n} m^{j-1} g_{i_j}$$

因此群 G 的全部生成元为

$$\{g_1, \ldots, g_r\} \cup \{x \in G \mid q(x) \leq 1 + \frac{C_1 + C_2}{2}\}$$

从而 G 是有限生成的。

读者要注意，上面定理中的 $m \geq 2$ 是存在性的且依赖于高度函数 q，而我们接下来构造的高度函数给出了 $m = 2$，于是在第二步中我们也只证明 $G/2G$ 是有限的，不过一般性的**弱 Mordell-Weil 定理**应该指：对任意数域 K 上的阿贝尔簇 A 和整数 $m \geq 2$，$A(K)/mA(K)$ 是有限的。但正如前面所说，我们目前只想考虑最简单的椭圆曲线有理点情形 $E(\mathbb{Q})$，并引入高度理论，而概形相关的内容我们并不想详细讲述。设 $\mathbb{Q}$ 上的椭圆曲线 $E : y^2 = x^3 + Ax + B$，对任意 $t = \frac{p}{q} \in \mathbb{Q}, (p, q) = 1$ 定义 $H(t) = \max\{|p|, |q|\}$，使用 $x(P)$ 和 $y(P)$ 分别表示 $P \in E(\mathbb{Q})$

的 x 坐标和 y 坐标，定义高度函数为

$$h_x : E(\mathbb{Q}) \to \mathbb{R}, x \mapsto \begin{cases} \ln H(x(P)) & P \neq O \\ 0 & P = O \end{cases}$$

其中 O 是 $E(\mathbb{Q})$ 的单位元，即无穷远点。

定理 7.14 (有理域上椭圆曲线的高度函数)

如上面定义，则有

(1)$\forall P_0 \in E(\mathbb{Q}), \exists C_1, \forall P, h_x(P+P_0) \leq 2h_x(P) + C_1$

(2)$\exists C_2, \forall P, h_x(2P) \geq 4h_x(P) - C_2$

(3)$\forall C, \{P \in E(\mathbb{Q}) \mid h_x(P) \leq C\}$ 是有限集。

♡

证明 (1) 对于 $P_0 \in E(\mathbb{Q})$，我们任选一个 $C_1 > \max\{h_x(P_0), h_x(2P_0)\}$，当 $P_0 = O$ 或 $P \in \{O, \pm P_0\}$ 不等式显然成立，故我们只考虑 $P \neq \pm P_0$ 和没有无穷远点 O 时的情形。我们记 (方程样式给出)

$$P = (x, y) = (\frac{a}{d^2}, \frac{b}{d^3}), P_0 = (x_0, y_0) = (\frac{a_0}{d_0^2}, \frac{b_0}{d_0^3})$$

由加法定义和方程联立可得

$$x(P+P_0) = (\frac{y-y_0}{x-x_0})^2 - x - x_0 = \frac{(aa_0 + Ad^2d_0^2)(ad_0^2 + a_0d^2) + 2Bd^4d_0^4 - 2bdb_0d_0}{(ad_0^2 - a_0d^2)^2}$$

由于 $H(t)$ 定义中，约分只会让数值变小，故在上界估计中我们可以直接考虑这个分数，接着我们借助主要部分 $f = O(g)$ 的做法来约去低次项，从而存在常数 C_1' 使得

$$H(x(P+P_0)) \leq C_1' \max\{|a|^2, |d|^4, |bd|\}$$

最后我们只需回到椭圆曲线方程 $b^2 = a^3 + Aad^4 + Bd^6$，从而有

$$|b| \leq C_1'' \max\{|a|^{\frac{3}{2}}, |d|^3\}$$

消去估计中的 b，从而得到

$$H(x(P+P_0)) \leq C_1''' \max\{|a|^2, |d|^4\} = C_1 H(x(P))^2$$

同时取对数，即可得

$$h_x(P+P_0) \leq 2h_x(P) + C_1'''$$

(2) 我们记 $E(\mathbb{Q})[m] = \{P \in E(\mathbb{Q}) \mid mP = O\}$，并任选一个 $C_2 > 4\max\{h_x(P) \mid P \in E(\mathbb{Q})[2]\}$，故可假定 $2P \neq O$，对于 $P = (x, y)$，通过加法公式有

$$x(2P) = x(P+P) = \frac{x^4 - 2Ax^2 - 8Bx + A^2}{4x^3 + 4Ax + 4B}$$

设 $x = \frac{a}{b} \in \mathbb{Q}, (a, b) = 1$，并记 $F(x, z) = x^4 - 2Ax^2z^2 - 8Bxz^3 + A^2z^4, G(x, z) = 4x^3z + 4Axz^3 + 4Bz^4$，则有

$$x(2P) = \frac{a^4 - 2Aa^2b^2 - 8Bab^3 + A^2b^4}{4a^3b + 4Aab^3 + 4Bb^4} = \frac{F(a, b)}{G(a, b)}$$

我们记 $\Delta = 4A^3 + 27B^2$，则它们的公因数至少满足 (代数运算后的结论，自己看着办)

$$|\delta| = |(F(a, b), G(a, b))| \leq |4\Delta|$$

约去以后，我们可以得到相应的估计为

$$H(x(2P)) \geq \frac{\max\{|F(a, b)|, |G(a, b)|\}}{|4\Delta|}$$

考察表达式中的余项即可得

$$\frac{\max\{|F(a, b)|, |G(a, b)|\}}{|4\Delta|} \geq \frac{1}{C_2'} \max\{|a|^4, |b|^4\} = \frac{1}{C_2'} H(x(P))^4$$

联立同时取对数即可得

$$h_x(2P) \geq 4h_x(P) - C_2$$

(3) 实际上，显然

$$\forall C, \{t \in \mathbb{Q} \mid H(t) \leq C\}$$

是有限集，即遍历分子和分母后最多就 $(2[C]+1)^2$ 个元素。更进一步，就有

$$\forall C, \{P \in E(\mathbb{Q}) \mid h_x(P) \leq C\}$$

最后，我们只需证明 $E(\mathbb{Q})/2E(\mathbb{Q})$ 是有限的，证明也就全部结束了。

定理 7.15 (Mordell 定理)

对任意 $\mathbb{Q}$ 上的椭圆曲线 E，$E(\mathbb{Q})$ 是有限生成交换群。

证明 证明“$E(\mathbb{Q})/2E(\mathbb{Q})$ 是有限的”并不简单，如果我们假定 $x^3 + Ax + B = 0$ 的根为有理数，则可以看这本书 [48]1.3 的简单构造法，读者需要知道高度理论并不是证明 Mordell 定理的关键，弱 Mordell 定理才是，为了证明的完整我们会采用稍微啰嗦的步骤。

(1) 我们先假设 $P_0 \in E(\mathbb{Q})[2] - O$ 非空，由对称性可知其坐标必定为 $P = (x_0, 0)$，此时我们做一个简单的平移变换 $x \longrightarrow x + x_0$，则曲线的方程可变成下面的形式

$$E' : y^2 = x(x^2 + ax + b)$$

并且 $(0,0) \in E'(\mathbb{Q})[2] - O$，我们使用这个作为椭圆曲线方程的新形式并不影响后续的证明，并且由光滑性有

$$E : y^2 = x(x^2 + ax + b), b \neq 0, a^2 - 4b \neq 0, (0,0) \in E(\mathbb{Q})[2] - O$$

我们考虑乘 2 群同态

$$[2] : E \longrightarrow E, P \mapsto 2P = P + P$$

则其存在一个同源分解

$$[2] = \hat{\varphi}\varphi, \varphi : E \longrightarrow E', \hat{\varphi} : E' \longrightarrow E, \ker\varphi = \{O, (0,0)\}, \ker\hat{\varphi} = \{O, P'\}, P' \in E'(\mathbb{Q})[2] - O$$

(2) 我们需要把上面的同源分解给具体算出来，对于 E' 我们不会使用原来的坐标，而是在一组坐标变换下，使得它与 E 拥有同样的形式

$$E' : Y^2 = X(X^2 + AX + B), B \neq 0, A^2 - 4B \neq 0$$

记 $P = (0,0)$，我们需要考察由同源 φ 诱导出映射

$$\ker\varphi \longrightarrow Aut(\overline{\mathbb{Q}}(E)/\varphi^*(\overline{\mathbb{Q}}(E'))), P \mapsto \tau_P^*$$

这里的 $\tau_P : E \longrightarrow E, x \mapsto x + P$ 是平移映射，τ_P^* 是相应在分式函数域上的诱导映射，其给出的两个不动点

$$\tau_P^*(x + a + \frac{b}{x}) = x + a + \frac{b}{x}, \tau_P^*(y - \frac{by}{x^2}) = y - \frac{by}{x^2}$$

诱导出了我们所需要的坐标变换

$$\begin{cases} X &= x + a + \frac{b}{x} \\ Y &= y - \frac{by}{x^2} \end{cases} \Leftrightarrow \begin{cases} x &= \frac{X + Y\sqrt{X} - a}{2} \\ y &= x\sqrt{X} \end{cases}$$

此时我们可以计算出相应的参数为

$$A = -2a, B = a^2 - 4b$$

于是在这种对应下 φ 具有对称的送点性质，即

$$\varphi(x,y) = (x + a + \frac{b}{x}, y - \frac{by}{x^2}), \hat{\varphi}(X,Y) = (X + A + \frac{B}{X}, Y - \frac{BY}{X^2})$$

$$\ker\hat{\varphi}\varphi = \ker[2] = E[2], \deg\hat{\varphi}\varphi = \deg\hat{\varphi}\deg\varphi = 4 = |E[2]| = |\ker\hat{\varphi}\varphi|$$

(3) 我们继续定义映射

$$\pi : E'(\mathbb{Q}) \to \mathbb{Q}^\times/\mathbb{Q}^{\times 2}, (X,Y) \mapsto \begin{cases} X & X \neq 0 \\ a^2 - 4b & X = 0 \\ 1 & (X,Y) = O \end{cases}$$

我们注意到当 $(X,Y) \in \mathrm{im}\varphi$ 时，若 $X \neq 0$ 则有 $X = (\frac{y}{x})^2 \in \mathbb{Q}^{\times 2}$，若 $X = 0$ 则 $x(x^2 + ax + b) = 0$，由于 $x \neq 0$，故 $x^2 + ax + b = 0$ 存在有理解，即 $a^2 - 4b \in \mathbb{Q}^{\times 2}$，换言之 $\mathrm{im}\varphi = \ker\pi$。接着我们来说明它是一个群同态，由定义可知 $\pi(O) = 1$，对于互逆元有 $\pi(P) = X \Rightarrow \pi(-P) = [\frac{1}{X}] = [X] = \pi(P)$，更进一步考虑两个不互逆的 $P + Q \neq O$，由于 $P + Q = R \neq O$，故可以设它们的连线为 $Y = eX + m$，带入方程可得

$$(eX + m)^2 = X(X^2 + AX + B) \Rightarrow X^3 + (A - e^2)X^2 + (B - 2em)X - m^2 = 0$$

由韦达定理可知 $x(P)x(Q)x(R) = m^2 \in \mathbb{Q}^{\times 2}$，因此 $\pi(P)\pi(Q)\pi(P + Q) = 1$，从而 π 是一个群同态

$$\pi(P + Q) = \frac{1}{\pi(P + Q)} = \pi(P)\pi(Q)$$

(4) 对任意一点 $(X,Y) \in E'(\mathbb{Q}), X \neq 0, [r] = \pi(x,y)$，由于 $\mathbb{Q}^{\times 2}$ 表示平方有理数，故我们可以令 $r \in \mathbb{Z}$ 无平方因子。此时对相应的方程 $Y^2 = X(X^2 + AX + B)$ 运用分解可得

$$X^2 + AX + B = rs^2, X = rt^2, s, t \in \mathbb{Q}^\times$$

我们继续令 $t = \frac{e}{m}, (e,m) = 1, e, m \in \mathbb{Z}$，并消去 X 可得

$$r^2e^4 + Are^2m^2 + Bm^4 = rm^4s^2$$

对任意一个素因子 $p \mid r$，由上面的等式可知，$p \mid Bm^4$。如果有 $p \mid m$，则有 $p^3 \mid r^2e^4$，从而有 $p \mid e$，进而 $(e,m) = p$ 矛盾，故只能 $p \mid B$，即 $r \mid B$，于是我们有有限集

$$\mathrm{im}\pi = \{1, a^2 - 4b\} \cup \{[r] \in \mathbb{Q}^\times/\mathbb{Q}^{\times 2} : r \mid B\}$$

(5) 由 $\mathrm{im}\pi$ 是有限的，我们可以得到

$$E'(\mathbb{Q})/\varphi(E(\mathbb{Q})) = E'(\mathbb{Q})/\ker\pi \cong \mathrm{im}\pi$$

是有限集。由对称性可知，$E(\mathbb{Q})/\hat{\varphi}(E'(\mathbb{Q}))$ 也是有限集。更进一步，由于 $[2](E(\mathbb{Q})) = 2E(\mathbb{Q})$，故有

$$|E(\mathbb{Q})/2E(\mathbb{Q})| = |E(\mathbb{Q})/\mathrm{im}[2]| = |E(\mathbb{Q})/\mathrm{im}\hat{\varphi}\varphi| = |E(\mathbb{Q})/\hat{\varphi}(E'(\mathbb{Q}))||E'(\mathbb{Q})/\varphi(E(\mathbb{Q}))|$$

是个有限集，证毕。

最后根据有限生成交换群的基本定理可得

$$E(\mathbb{Q}) = \mathbb{Z}^r \oplus E(\mathbb{Q})_{tors}$$

定理 7.16 (Mazur)

对于 $\mathbb{Q}$ 上的椭圆曲线 E，挠子群 $E(\mathbb{Q})_{tors}$ 一定同构于下面的 15 种群之一

$$\mathbb{Z}/N\mathbb{Z}, 1 \leq N \leq 10, N = 12$$

$$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2N\mathbb{Z}, 1 \leq N \leq 4$$

♡

至此 Mordell 定理也就告一段落了，不过我还想稍微讲讲与之相关的 BSD 猜想 (Birch and Swinnerton-Dyer Conjecture)，这里是官方的描述。对于 BSD 猜想的成立有两个大前提，一是阿贝尔簇 A 的 K-点有分解

$$A(K) = \mathbb{Z}^r \oplus A(K)_{tors}$$

二是模性定理，你可能会觉得直接根据曲线定义 L 函数就可以了，这是不对的，因为就算我们之前有 D 级数通论，但你依旧无非通过那杂乱无章的 a_n 来推出此函数的解析性，只有通过模性定理将其等于某个模形式的 L 函数，才能说明它的解析性，也才有相应的级数展开

$$L(A,s)=c(s-1)^r+o(s^r)$$

而对于模性定理，目前的进度并不可观，例如这篇文章说明就算是 $\mathbb{Q}$ 上的阿贝尔簇也得加上 GL_2-type 的条件，就算是椭圆曲线这篇文章也只把 $\mathbb{Q}$ 扩充到了实二次域，因此对于 BSD 猜想最好的环境应该是最初的椭圆曲线有理点群 $E(\mathbb{Q})$。实际上，扩展版的 BSD 猜想也不是不行，只是它需要被解决的不单单是“解析秩和代数秩相等”那么简单了，还有许多前置条件有待解决。我们记 $\mathbb{Q}$ 所有素点的集合为 $S=\{\infty,2,3,5,7,11,...\}$，并引入简单的伽罗瓦群符号

$$G_p=Gal(\overline{\mathbb{Q}_p}/\mathbb{Q}_p)\subset G=Gal(\overline{\mathbb{Q}}/\mathbb{Q})$$

此时我们把

$$Ш(E)=\cap_{p\in S}\ker(H^1(G,E(\overline{\mathbb{Q}}))\to H^1(G_p,E(\overline{\mathbb{Q}}_p)))$$

称为 **Shafarevich-Tate 群**，(记号不一样是因为打出俄文的 X 太麻烦了)，这里的映射是由伽罗瓦群的嵌入映射诱导出的群上同调群映射。设椭圆曲线 E 在模形式中对应的 L 函数为 $L(E,s)=c(s-1)^{r'}+o(s^{r'})$，我们记 $\mathrm{rank}L(E,s)=r',\mathrm{rank}E(\mathbb{Q})=r$，则 BSD 猜想的完整表示为两个等式

命题 7.3 (Birch and Swinnerton-Dyer Conjecture)

$$\mathrm{rank}L(E,s)=\mathrm{rank}E(\mathbb{Q})$$

$$\frac{L^{(r)}(E,1)}{r!}=\frac{\Omega_E\mathrm{Reg}_E|Ш(E)|}{|E(\mathbb{Q})_{tors}|^2}\prod_{p|N}c_p(E)$$

♠

大家所认为的 BSD 猜想其实只是前一个，显然后一个式子相当于给出了 $L(E,s)=c(s-1)^r+o(s^r)$ 展开式中 c 的具体值，我们有理由相信证明前一个结论，后一个也会顺带解决，至于里面的一大堆与椭圆曲线 E 相关的常数，知不知道也无所谓了。就讲这些东西吧，我可不想离我们的主题数论太遥远了。

一般代数曲线与 Faltings 定理

定理 7.17 (Faltings 定理)

若 C 是数域 K 上满足亏格 $g\geq 2$ 的代数曲线，则 $C(K)$ 是有限的。

♡

这节定理的表述还是挺简单的，稍微提醒一下读者，我们只是说 $C(K)$ 是有限的，而没有说 $C(K)$ 是有限群，它并不一定能像阿贝尔簇那样拥有群结构。读者需要知道“不等式就是数论的神”，对于大量的不等式估计结果比如前面的 Roth 定理得十分熟悉才行，如果你的基础令人堪忧的话，就看这本书 [18]，其不仅准备了前置知识，还可以直接就在里面看完整证明，目前对 Faltings 定理的复述都不是 Faltings 原版的，而是 Vojta 和 Bombieri 改进后的证明，望读者悉知。雅可比簇 (Jacobian Variety) 具有多种定义观点，由 Abel-Jacobi 定理可知，我们甚至可以直接把它视为 Picard 群

$$\mathrm{Jac}(C)=\mathrm{Pic}(C)=\mathrm{Div}(C)/\mathrm{Div}^0(C)$$

我们把 $\mathrm{Jac}(C)$ 称为代数曲线 C 的**雅可比簇** (一种特殊的阿贝尔簇)，更进一步，我们存在一个**雅可比嵌入**满足

$$j:C\to J=\mathrm{Jac}(C),j(C)^{\oplus g}=J\cong\mathbb{C}^g/\Lambda_g,\dim J=g$$

我们这样做的原因显而易见，目的就是让代数曲线 C 嵌入到阿贝尔簇 J 中去，从而我们可以使用阿贝尔簇上的高度理论，上一部分中，我们提过一个高度函数

$$\hat{h}(P)=\lim_{n\to\infty}\frac{1}{4^n}h(2^nP),P\in A(K)$$

它也被称为 **Neron-Tate 高度**，它最重要的一个特点是构成二次型

$$\hat{h}(P+Q)+\hat{h}(P-Q)=2\hat{h}(P)+2\hat{h}(Q)$$

从而其给出了一个双线性形式

$$\langle,\rangle:A(K)\times A(K)\to\mathbb{R},(P,Q)\mapsto\frac{\hat{h}(P+Q)-\hat{h}(P)-\hat{h}(Q)}{2}$$

从定义也能看出它和常规高度函数的差值是有界的 $\hat{h}(P)-h(P)=O(1)$，对于 $X\in A(K)$，我们记 $|x|=\sqrt{\langle x,x\rangle}$，则为了证明 Faltings 定理的一个核心不等式是

定理 7.18 (Vojta 不等式)

设 C 是数域 K 上满足亏格 $g\geq 2$ 的代数曲线，则对 $1<\alpha<\frac{\pi}{2},\sqrt{g}\cos\alpha>1$，存在常数 $\kappa_1=\kappa_1(C),\kappa_2=\kappa_2(g)$ 使得对任意的 $|z_1|\geq\kappa_1,|z_2|\geq\kappa_2|z_1|,z_1,z_2\in C(K)$ 有

$$\langle z_1,z_2\rangle<\cos\alpha|z_1||z_2|$$

♡

注意 α 可取的范围 $\cos\alpha>\frac{1}{\sqrt{g}}\leq\frac{1}{\sqrt{2}}$，因此 $\cos\alpha>\frac{1}{\sqrt{2}}$，由于 $\frac{3}{4}>\frac{1}{\sqrt{2}}$，因此大多情况下，你也能看到 Vojta 不等式为

$$\langle z_1,z_2\rangle<\frac{3}{4}|z_1||z_2|$$

不过前面这个系数可不可控并不是什么重要的事情，重要的是我们来展示 Vojta 不等式可以推出 Faltings 定理。

证明 [Vojta 不等式 $\Rightarrow$Faltings 定理] (1) 我们先对雅可比簇 $J=\mathrm{Jac}(C)$ 进行系数提升，即做张量积 $L=J\otimes_{\mathbb{Z}}\mathbb{R}$(简单来讲就是把除子中的整系数扩充为实系数)，由 Mordell-Weil 定理可知 (雅可比簇是阿贝尔簇)，L 是一个 $\mathbb{R}$ 上的有限维线性空间，考虑映射复合链

$$\tilde{j}:C(K)\xrightarrow{j}J\to L$$

这里相当于连续两步的嵌入映射，容易发现 $\ker(J\to L)=J_{tors}$ 且 j 是单射，故我们只需证明 $\mathrm{im}\tilde{j}=\tilde{j}(C(K))$ 是有限的即可。借助在 J 中的高度理论，我们可以在线性空间 L 中定义两点 x,y 间的角度为

$$\cos\theta_{x,y}=\frac{\langle x,y\rangle}{|x||y|},0\leq\theta_{x,y}\leq\pi$$

对任意一点 $x_0\in L$ 和角度 θ_0，我们引入一个子集

$$\Gamma_{x_0,\theta_0}=\{x\in L\mid\theta_{x,x_0}<\theta_0\}$$

我们固定一个较小的 θ_0，则我们可以直接“转一圈”[1]填满整个 L，换句话只要有限个 Γ_{x_i,θ_0} 就能填满 L，于是我们只需要证明下面的有限相交性质即可

$$\exists\theta_0\forall x_0,|\Gamma_{x_i,\theta_0}\cap\tilde{j}(C(K))|<\infty$$

(2) 使用反证法，不妨假设存在一个无穷集

$$|\Gamma_{x_i,\theta_0}\cap\tilde{j}(C(K))|=\infty$$

根据有限高度下的元素有限可知，无限集意味着里面的元素可以任意的高，故我们可以选取两个符合条件的元

[1]不懂的话，就试着用有限开覆盖定理证一下

素

$$|z_1| \geq \kappa_1, |z_2| \geq \kappa_2 |z_1|, z_1, z_2 \in \Gamma_{x_i,\theta_0} \cap \tilde{j}(C(K))$$

由 Vojta 不等式可知

$$\langle z_1, z_2 \rangle < \frac{3}{4}|z_1||z_2|$$

即

$$\theta_{z_1,z_2} > \cos^{-1}\frac{3}{4} > \frac{\pi}{6}$$

此时我们取 $\theta_0 = \frac{\pi}{12}$，则由定义可得

$$\forall z \in \Gamma_{x_0,\frac{\pi}{12}}, \theta_{z,x_0} < 2\frac{\pi}{12} = \frac{\pi}{6}$$

矛盾，从而

$$\forall x_0, |\Gamma_{x_0,\frac{\pi}{12}} \cap \tilde{j}(C(K))| < \infty$$

这样我们的工作重心就是证明 Vojta 不等式的可怕工作了，具体有多可怕呢？可怕到我直接不想证明了。嘛，证明的过程是十分冗长的，因此我们必需进行适当地取舍才行，我们的基本想法是不管计算估计的过程，转而展示计算估计所带来的效应。我个人认为，大概率大部分的人想知道的不是怎么计算，自然也不会去关心各种计算的细节，而是关心这个结果为什么可以被证明出来，也就是转化的想法是什么，对于这一点，我在解析数论的思想中也讨论过，计算对于证明者而言确实是重要的基本功，但对于阅读者而言，这不是什么好的选择。如果 $C \times C$ 上的除子 (与三个参数 d_1, d_2, d 有关) 满足

$$V(d_1,d_2,d) = (d_1-d)p_1^*(\theta) + (d_2-d)p_2^*(\theta) + d\Delta, gd^2 < d_1d_2 < g^2d^2$$

则称 $V(d_1,d_2,d)$ 为 **Vojta 除子**，其中 $p_i : C \times C \to C, i = 1,2$ 为两个分量上的投影映射，p_i^* 为对应回拉映射诱导的除子映射，Δ 为 $C \times C$ 上的对角除子，θ 是任意 C 上使得 $(2g-2)\theta$ 为主除子的除子。

定理 7.19

(1)[上界估计] 存在常数 c_1 使得，对任意正整数 d, d_1, d_2 和所有的点 $z, w \in C(\overline{K})$ 满足

$$h_{C\times C, V(d_1,d_2,d)}(z,w) \leq \frac{d_1}{g}|z|^2 + \frac{d_2}{g}|w|^2 - 2d\langle z,w\rangle + c_1(d_1+d_2+d)$$

(2)[下界估计] 存在常数 c_2, c_3 和有限点集 $\mathcal{F} \subset C(\overline{K})$ 使得，对任意 $z, w \in C(\overline{K}), z, w \notin \mathcal{F}$ 满足

$$h_{C\times C, V(d_1,d_2,d)}(z,w) \geq -h(\mathcal{F}) - c_2(i_1^*|z|^2 + i_2^*|w|^2) - c_3(i_1^* + i_2^* + \delta_1 + \delta_2 + 1)$$

(3)[截面估计] 给定任意 $\gamma > 0$ 并使 d_1, d_2, d 足够满足 $d_1d_2 - gd^2 \geq \gamma d_1 d_2$，则有

$$h(\mathcal{F}) \leq c_1\frac{d_1+d_2}{\gamma} + o(d_1+d_2)$$

(4)[无灭衍生] 存在常数 c_4 使得，如果足够小常数 $0 < \varepsilon, \gamma < 1$ 和足够大常数 d_1, d_2, d 和点 $z, w \in C(\overline{K})$ 满足

$$\varepsilon^2 d_1 \geq d_2, \min\{d_2|w|^2, d_1|z|^2\} \geq \frac{c_4}{\gamma\varepsilon^2}d_1, d_1d_2 - gd^2 \geq \gamma d_1 d_2$$

则有足够大的 N 使得

$$\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2} \leq 12N\varepsilon$$

♡

借助上面的四个估计式，我们就能来完成 Vojta 不等式的证明了。

证明 首先假定一个足够大的 κ_1，由于下界估计给出了下界，故我们可以使得 $\forall z \in \{|z| > \kappa_1\}, z \notin \mathcal{F}$。接着我们选取一个足够大的 $D > |w|^2$ 和 $0 < v, \varepsilon < 1$ 并定义出三个整数

$$d_1 = N[\sqrt{g+v}\frac{D}{|z|^2}], d_2 = N[\sqrt{g+v}\frac{D}{|w|^2}], d = N[\frac{D}{|z||w|}]$$

(1) 根据下界估计我们可以得到

$$h(z,w) \geq -h(\mathcal{F}) - c_2(i_1^*|z|^2 + i_2^*|w|^2) - c_3(i_1^* + i_2^* + \delta_1 + \delta_2 + 1)$$

进一步赋予 $\kappa_1 > \max\{1, \varepsilon^{-\frac{1}{2}}\}$，则有 $|z| > 1, |w| > 1$，从而

$$d_1 + d_2 + d \leq ND(\frac{\sqrt{g+v}}{|z|^2} + \frac{\sqrt{g+v}}{|w|^2} + \frac{1}{|z||w|}) \leq \frac{c_5 D}{\kappa_1^2} \leq c_5 \varepsilon D$$

即有 (消去尾项)

$$h(z,w) \geq -h(\mathcal{F}) - c_2(i_1^*|z|^2 + i_2^*|w|^2) - c_5\varepsilon D$$

(2) 容易观测

$$\frac{d_1 d_2 - g d^2}{d_1 d_2} \geq 1 - \frac{g(\frac{D}{|z||w|})^2}{(\frac{\sqrt{g+v}D}{|z|^2} - 1)(\frac{\sqrt{g+v}D}{|w|^2} - 1)} = 1 - \frac{g}{g+v}\frac{1}{1 - \frac{|z|^2}{\sqrt{g+v}D}}\frac{1}{1 - \frac{|w|^2}{\sqrt{g+v}D}}$$

于是我们只需要选取 $\gamma = \frac{v}{3g}$，就可以满足截面估计的条件，从而有

$$h(\mathcal{F}) \leq c_1 \frac{d_1 + d_2}{\gamma} + o(d_1 + d_2) \leq c_6(d_1 + d_2) \leq c_7 \varepsilon D$$

即有 (消去高度项)

$$h(z,w) \geq -c_2(i_1^*|z|^2 + i_2^*|w|^2) - c_8\varepsilon D$$

(3) 我们先让 $\kappa_2 \geq \sqrt{2}\varepsilon^{-1}$，接着我们来验证无灭衍生的条件

$$\frac{d_2}{d_1} \leq \frac{N\sqrt{g+v}D/|w|^2}{N(\sqrt{g+v}D/|z|^2 - 1)} = \frac{2|z|^2}{|w|^2} \leq \frac{2}{\kappa_2^2} \leq \varepsilon^2 \Rightarrow \varepsilon^2 d_1 \geq d_2$$

我们继续令 $\kappa_1^2 \geq 2c_4/(\gamma\varepsilon^2)$，则可以验证

$$0 \leq \eta_1, \eta_2 \leq 2, \frac{d_2|w|^2}{d_1|z|^2} = \frac{1 - \frac{\eta_2|w|^2}{D\sqrt{g+v}}}{1 - \frac{\eta_1|z|^2}{D\sqrt{g+v}}} \Rightarrow \frac{1}{2} \leq \frac{d_2|w|^2}{d_1|z|^2} \leq 2 \Rightarrow \min\{d_2|w|^2, d_1|z|^2\} \geq \frac{c_4}{\gamma\varepsilon^2} d_1$$

此时我们利用无灭衍生估计可得

$$i_1^*|z|^2 + i_2^*|w|^2 \leq c_9\varepsilon(d_1|z|^2 + d_2|w|^2) \leq c_9\varepsilon(N\frac{\sqrt{g+v}D}{|z|^2}|z|^2 + N\frac{\sqrt{g+v}D}{|w|^2}|w|^2) \leq c_{10}\varepsilon D$$

即有 (消去最后项)

$$h(z,w) \geq -c_{11}\varepsilon D$$

(4) 接着，我们来考察上界估计

$$h(z,w) \leq \frac{d_1}{g}|z|^2 + \frac{d_2}{g}|w|^2 - 2d\langle z,w\rangle + c_1(d_1 + d_2 + d)$$

我们可以十分轻而易举地估计前几项

$$\frac{d_1}{g}|z|^2 + \frac{d_2}{g}|w|^2 - 2d\langle z,w\rangle \geq h(z,w) - c_1(d_1 + d_2 + d) \geq -c_{12}\varepsilon D$$

我们需要的是这个反向估计结果，以便向我们的结论靠近，即我们需要化简上面的不等式，借助 $\lim_{B\to\infty} \frac{[aD]}{D} = a$ 可得

$$\frac{[\frac{\sqrt{g+v}D}{|z|^2}]|z|^2}{g} + \frac{[\frac{\sqrt{g+v}D}{|w|^2}]|w|^2}{g} - 2[\frac{D}{|z||w|}]\langle z,w\rangle \geq -c_{12}\varepsilon D \Rightarrow \frac{2\sqrt{g+v}}{g} - \frac{2\langle z,w\rangle}{|z||w|} \geq -c_{12}\varepsilon$$

于是我们化简可得

$$\langle z,w\rangle \leq (\frac{\sqrt{g+v}}{g} + \frac{c_{12}\varepsilon}{2})|z||w|$$

我们只需令 $\varepsilon \to 0, v \to 0$，即有

$$\langle z,w\rangle \leq (\frac{\sqrt{g}}{g})|z||w| \leq \frac{1}{\sqrt{2}}|z||w| < \frac{3}{4}|z||w|$$

希望读者不要对我们随便让 κ_i 大于某个数而感到顾虑，由于它们本身是存在性的常数，因此我们可以不断

地往它身上加不会导致矛盾的条件，在数论中大多都是充分大，意味着我们让常数往无穷方向发展，实在不能接受的话，读者可以把所有出现的地方取极大值即可。实际上，就算我只是展示了 Faltings 定理证明的一角，对于不经常玩不等式的人来说，一定是满脑子的问号，为什么估计这个？为什么这么估计？为什么要添上这些条件？为什么要这么算？... 这是情有可原的，因为证明是结果不是过程，过分地阅读很容易让我们迷失在计算中，这也是为什么我并不推荐将各种估计证明从头看到尾的原因。以我的个人经验来看，这种从头到尾的验证式阅读，最后会给你一种好像读懂了又好像没读懂的感觉，而结果就是几天后什么也不知道了，甚至可能还不如我给出的概要阅读后的记忆深刻，因为此时你还需要读懂这个关节上的内容，而在缺少信息的情况下是很容易促发思考，从而提升理解的。

7.3 零散内容

你知道吗？对于算术几何，或者称为丢番图几何，我觉得也没有更多可讲的东西了。虽然我们还没完全搞懂代数曲线的具体结构，而且代数曲线之外还有代数曲面，甚至更高维的代数超曲面，但我还是用最经典的一句话“过于零碎，没法总结”来表达我的逃避，一个最简单的例子就是这个，它是一种类似于暑期学校的算术几何总结，看也是能看，但也只能分散着来看，前后内容基本没有联系。

定理 7.20 (阿贝尔子簇问题)

设 A 是 $\mathbb{C}$ 上的阿贝尔簇，X 是 A 的闭子簇，Γ 是 $A(\mathbb{C})$ 有限秩子群。则存在有限个点 $\gamma_1,...,\gamma_r\in\Gamma$ 和 A 的有限个阿贝尔子簇 $B_1,...,B_r$，满足

$$\forall 1\leq i\leq r,\gamma_i+B_i\subset X, X(\mathbb{C})\cap\Gamma=\cup_{1\leq i\leq r}\gamma_i+(B_i(\mathbb{C})\cap\Gamma)$$

特别地，如果 X 不包含任何 A 的非平凡阿贝尔子簇的平移时，$X(\mathbb{C})\cap\Gamma$ 是有限的。

定理 7.21 (扩域问题)

设 X 是亏格 $g\geq 2$ 的代数曲线，记 $W_d(X)=\underbrace{X+...+X}_{d}\subset \mathrm{Jac}(X), d\geq 1$，如果 X 不允许态射 $X\to\mathbb{P}^1$ 的次数超过 d，且 $W_d(X)$ 不包含任何 $\mathrm{Jac}(X)$ 的阿贝尔子簇的平移，则

$$X^{(d)}(k)=\cup_{[K,k]\leq d}X(K)$$

是有限的。

命题 7.4 (ABC 猜想)

定义 $\mathrm{rad}(n)=\prod_{p|n}p, n\in\mathbb{Z}, n\neq 0$。对任意的 $\varepsilon>0$ 存在常数 C_ε 使得，如果 $a,b,c\in\mathbb{Z},(a,b,c)=1,a+b+c=0$，则有

$$\max\{|a|,|b|,|c|\}\leq C_\varepsilon(\mathrm{rad}(abc))^{1+\varepsilon}$$

命题 7.5 (Bombieri-Lang 猜想)

设 X 是数域 k 上的射影簇，Sp_X 是所有非平凡有理映射 $A\to X$(A 为阿贝尔簇) 的像的并集的 Zariski 闭包，记 $U=X-\mathrm{Sp}_X$。则对任意的有限扩张 k'/k 有 $U(k')$ 是有限的。

命题 7.6 (Vojta 猜想)

设 X 是数域 k 上的光滑射影簇，S 是 k 素点的有限集，E 是 X 上的丰富除子，D 是 X 上只有正规交叉的有效约化除子。则对任意的 $\varepsilon>0$ 存在真闭子集 $Z\subset X$ 满足

$$\forall P\in(X-Z)(k), m_S(D,P)+h_{K_X}(P)\leq\varepsilon h_E(P)+O_\varepsilon(1)$$

第八章 计算与验证

不论是数学分析还是线性代数，它们最重要的特点就是**计算**，计算是数学的灵魂，缺少逻辑的计算是无理的，但缺少计算的逻辑是无趣的，数理逻辑固然可贵，但它所能带来的惊喜是有限的，而数字中却蕴含着无穷的奥秘，而支撑起这些数字的正是那些计算，而且就算是数理逻辑的一些命题，比如哥德尔定理也是离不开数字的，计算的重要性就点到为止了。问题是，计算有什么可以讨论的吗？简单来讲，就是**计算机**，从而可以引申出算法和复杂度等问题。计算本身确实没有可谈的东西，但怎么算？能算出什么？就是一个内涵十分丰富的学科了。

8.1 计算

计算复杂度

既然谈到算法，那么自然就有一个计算复杂度的概念，其主要符号就是我们在解析数论中见过的大 O 记号

$$f(n)=O(g(n))\Leftrightarrow \exists C,|f(n)|\leq g(n)$$

而且对于一些常见复杂度的关系，也要足够清楚 (从小到大)

$$\ldots,\ln\ln n,\ln n,n,n\ln n,n^2,n^3,\ldots,2^n,3^n,\ldots,n!,n^n$$

通常它们都是无界增函数。由于 $\log_a n=\ln a^{-1}\ln n$，因此对数复杂度具有底数无关性，故通常我们就直接写成 $\log n$，任意一个整数 n 在 b 进制下的长度为 $[\log_b n]+1=O(\log n)$，因此对于整数而言其是几进制并不影响算法复杂度。此时我们考察按位加法可知，其计算复杂度为

$$T(n+n)=O(n\text{的位数})=O(\log n)$$

同样地，我们考察传统数乘可知，其计算复杂度为

$$T(nn)=O((n\text{的位数})^2)=O(\log^2 n)$$

实际上，通过快速傅里叶变换 (Fast Fourier Transform (FFT)) 可以得到一个复杂度更小的算法

$$T(nn)=O(\log n\log\log n\log\log\log n)$$

除了最基础的加减法，我们还可以看些其它的计算，比如模指数 $x^e \bmod n$，即计算出 x^e 以后除以 n 取余数。显然如果直接计算的话，x^e 会很大，甚至可能超过计算机的容量，因此我们最好通过同余的性质来简化计算，即边算边取余数。如果更进一步，在以二进制为基础的计算机上，计算以 2 的幂为指数是比较简单的，因此我们可以把指数进行二进制分解来简化计算

$$x^e=x^{\sum_{i=0}^{k}\beta_k 2^k}=x^{\cdots(2\beta_k+\beta_{k-1})2+\beta_{k-2})2+\beta_{k-3})2+\cdots},\beta_k=0,1$$

对于每个节点，我们利用指数性质来递归计算 $\ldots((x^{\beta_k})^2x^{\beta_{k-1}})^2\ldots$，从而得到算法复杂度为

$$T(x^e \bmod n)=O(\log e\log^2 n)$$

我们再来讨论一个与之前椭圆曲线相关的计算，即椭圆曲线 $E(\mathbb{F}_q)$ 上的加法

$$kP = \underbrace{P + ... + P}_{k}, P \in E(\mathbb{F}_q)$$

显然计算机是不会联立方程求解的，只能使用我们求出来的加法公式，但如果 k 十分大的话，就没那么方便了，而更快的算法其实也就是去适应计算机的二进制，把 k 的二进制表示出来为 $k = \sum_{i=0}^{k} \beta_k 2^k$，则有

$$kP = \beta_0 P + 2(\beta_1 P + ...2(\beta_{k-2} P + 2(\beta_{k-1} P + 2\beta_k P)))...$$

从而得到算法复杂度为

$$T(kP) = O(\log k \log^3 q)$$

因数分解

$$n = p_1^{\alpha_1} p_2^{\alpha_2} ... p_k^{\alpha_k} = \prod_{i=1}^{k} p_i^{\alpha_i}, p_i < p_{i+1}$$

对于质因数分解，最简单的方法就是不断地从小到到试整数，然后不断缩小 n，直到 1，提醒读者试质数是不靠谱的，因为如果遇到大质数连判定都是一个问题，更别说大整数分解了。相应的优化算法有点小多，我们就来讲一下大家熟悉的“筛法”(Sieve)，或者说平方筛法，其利用了 N 的素因子不会超过 $\sqrt{N}$ 的事实。若要分解 N，我们先试着分解成两个数的积，在 $\sqrt{N}$ 的附近，我们先计算几个数

$$W_k = (k + [\sqrt{N}])^2 - N \equiv (k + [\sqrt{N}])^2 (\mathrm{mod} N), 0 < k < L, 0 < W_k < (2L+1)\sqrt{N} + L^2$$

接着对较小的 W_k 进行因式分解，以凑出平方数

$$\prod_{i=1}^{t} W_{n_i} = y^2$$

从而有

$$x^2 \equiv y^2 (\mathrm{mod} N), x \neq \prod_{i=1}^{t} ([\sqrt{N}] + n_i) (\mathrm{mod} N)$$

此时计算 $\gcd(N, x-y)$ 或 $\gcd(N, x+y)$ 就能得到两个非平凡的因子。我们举个简单的例子，考虑 $N = 2041$，此时

$$W_{-2} = -2^6 \times 3, W_{-1} = -3 \times 5 \times 7, W_0 = -2^4, W_1 = 3 \times 5^4$$

接着我们凑出平方数

$$(43 \times 45 \times 46)^2 \equiv (-1)^2 \times 2^{10} 3^2 5^2 = (2^5 \times 3 \times 5)^2 (\mathrm{mod} N)$$

此时计算 $\gcd(2041, 43 \times 45 \times 46 + 2^5 \times 3 \times 5) = 157, \gcd(2041, 43 \times 45 \times 46 - 2^5 \times 3 \times 5) = 13$，验算可知有 $2041 = 157 \times 13$。读者可知注意到了，在这类算法中出现了其它算法，比如较小数的因式分解和求公因数等操作，但将复杂度更大的事变成复杂度更小的事是符合算法原则的，因此并没有恰不恰当的说法。不过这个算法的复杂度不好估计，不确定性太大，因此是一个猜想

$$O(e^{(1+o(1))\sqrt{\log N \log\log N}}) = O(N^{(1+o(1))\sqrt{\log\log N/\log N}})$$

离散对数

$$a, b, n \in \mathbb{N}, a^x \equiv b(\mathrm{mod}n)$$

所谓离散对数问题就是求上式中的 x，当然必需在存在的前提下找最小的，基本最原始的算法就是对 n 进行枚举，于是它本身就包含了模指数的运算，由于这个问题与密码学密切相关，所有相应的算法也有一大坨。但可能大多人对这个问题不感兴趣，所以我就随便讲一个比较简单的算法，即“大步小步算法”(Baby-Step Giant-Step)，名字有些小美妙，此时假定 $(a, n) = 1$ 是合理的，其过程比较简单，我们直接举个例子 $2^x \equiv 6(\mathrm{mod}19)$，此时各参数为 $a = 2, b = 6, n = 19, s = [\sqrt{19}] = 4$，先计算小步集为

$$S\mathrm{mod}19 = \{(b, 0), (ba, 1), (ba^2, 2), (ba^3, 3)\} = \{(6, 0), (12, 1), (5, 2), (10, 3)\}$$

然后计算大步集为

$$T\mathrm{mod}19 = \{(a^s, s), (a^{2s}, 2s), (a^{3s}, 3s), (a^{4s}, 4s)\} = \{(16, 4), (9, 8), (11, 12), (5, 16)\}$$

他俩按首项重新排序可得

$$S = \{(5, 2), (6, 0), (10, 3), (12, 1)\}, T == \{(5, 16), (9, 8), (11, 12), (16, 4)\}$$

其中 5 是共有的首项，从而 $x = 16 - 2 = 14$，即有 $2^{14} \equiv 6(\mathrm{mod}19)$。这个算法的时间复杂度和空间复杂度分别为

$$O(\sqrt{n}\log n), O(\sqrt{n})$$

数论计算

最后我们再来讨论一下数论相关的计算，例如质数个数函数 $\pi(n)$，由于其涉及素数判别，所以之间遍历判定是个速度不够的算法。算法界有句名言“想优化，就二分”，一个惊奇的现象是不论在哪里拆解总能让算法变得更快，例如我们可以把“计算 $\pi(n)$”变成“计算 $\pi(\sqrt{n})$+ 简单运算”，即有下面的公式

$$\pi(n) = \pi(\sqrt{n}) - 1 + \sum_{p|d \Rightarrow p \le \sqrt{n}} \mu(d)[\frac{n}{d}]$$

对于后面的求和式，我们可以进一步细化，假设 $p_1, ..., p_k$ 是所有不超越 $\sqrt{n}$ 的素数，则有

$$\sum \mu(d)[\frac{n}{d}] = n - \sum_i [\frac{n}{p_i}] + \sum_{i \ne j} [\frac{n}{p_i p_j}] - \sum_{i \ne j \ne t} [\frac{n}{p_i p_j p_t}] + ... + (-1)^k [\frac{n}{p_1 ... p_k}]$$

我们举一个简单的例子

$$\begin{aligned}\pi(129) &= \pi(\sqrt{129}) - 1 + 129 - \sum_i [\frac{129}{p_i}] + \sum_{i \ne j} [\frac{129}{p_i p_j}] - \sum_{i \ne j \ne t} [\frac{129}{p_i p_j p_t}] + ... + (-1)^k [\frac{129}{p_1 ... p_k}] \\ &= 5 - 1 + 129 - (64 + 43 + 25 + 18 + 11) + (21 + 12 + 9 + 5 + 8 + 6 + 3 + 3 + 2 + 1) \\ &\quad - (4 + 4 + 1 + 1 + 1 + 0 + 1 + 0 + 0 + 0) + (0 + 0 + 0 + 0 + 0) - 0 \\ &= 31\end{aligned}$$

实际上，如果我们能得到 $\pi(\sqrt{n})$，就应该能得到所有的素数 $p \le \pi(\sqrt{n})$。如果 $\sigma(n) = m, \sigma(m) = n$，则称 (m,n) 是**亲和数**，寻找亲和数的最基本方法是：遍历 $n \le \sup$，计算 $m = \sigma(n)$(所有因数之和)，验证 $\sigma(m) = n$。另外有一个优化的 Riele 算法，其遍历整数 $s \le \sup$，然后求解方程 $\sigma(x) = s$ 的所有解 $\{x_1, ..., x_k\}$，再寻找满足 $x_i + x_j = s$ 的解，就得到了一对亲和数。例如，我们发现遍历到 $s = 504$ 时有

$$\sigma(x) = 504 \Rightarrow x = \{286, 334, 220, 284, 224\}, 220 + 284 = 504$$

从而 $(220, 284)$ 是一对亲和数。哥德巴赫猜想验证，指验证每一个大于 2 偶数能否写成两个素数之和的形式，最基本的方法是：对每个偶数 n，找到所有素数 $p_i < n$，然后计算每一个 $n - p_i$，并验证其是否为素数。其实还有一个基于筛法的 DSR 算法，但我个人认为把哥德巴赫猜想验证转化为素数判定，再去研究素数判定才是好的路径，所有也就不深入探讨了。由于寻找偶完全数相当于找梅森素数，因此在算法中，我们自然想要找奇完全数的算法，不过对此也没什么特别的，也是遍历法，只不过我们有一个定理说其至少有三个素因子，从而可以减少一些枚举量罢了。

8.2 验证

验证和计算是相伴相生的，计算是为了验证，验证需要计算，拆开来讨论并不是明智之举。例如判定 Pell 型方程 $x^2 - dy^2 = N$ 是否有整数解，就是在某个不等式范围内计算的过程，其结果就是我们验证了 $x^2 - dy^2 = N$ 是否有整数解。但素数在数论中有着举足轻重的地位，单独拿出来讨论素数的验证问题，我个人觉得是值得且有必要的，于是这就是这节的核心内容。从逻辑上来看，它就是一个 Yes/No(True/False) 的问题，却有着巨大的难度。

定理 8.1

设整数 $n > 1$，如果 n 没有素因子超过 $\sqrt{n}$，则 n 是一个素数。

♡

验证素数最简单的方法就是，试除来找非平凡因子，上述的定理可以讲试除的范围由 $2 \to n$ 减小为 $2 \to [\sqrt{n}]$，算是素数判定的起点了。而其它的办法就是去找数论中的定理，需要哪些条件可以推出素数，例如下面的几个例子。

定理 8.2

下述的条件均可以推出 $n > 1$ 是一个素数

(1)[费马小定理的 Lucas 逆] 存在整数 a 满足

$$a^{n-1} \equiv 1 (\mathrm{mod} n); \forall p \mid n-1, a^{\frac{n-1}{p}} \not\equiv 1 (\mathrm{mod} n)$$

(2)[(1) 的等价定理] 存在整数 a 满足

$$\gcd(a, n) = 1, \mathrm{ord}_n a = \varphi(n) = n - 1$$

(3)[(1) 的分散形式] 对每个素因子 $p_i \mid n-1$ 存在整数 a_i 满足

$$a_i^{n-1} \equiv 1 (\mathrm{mod} n); a_i^{\frac{n-1}{p_i}} \not\equiv 1 (\mathrm{mod} n)$$

♡

通过上面的定理，可以容易推出，合数和素数的判定算法的复杂度为

$$O(\log^2 n), O(\log^4 n)$$

素数判定的理论比较稀缺，但伪素数的理论倒是不少，所谓**伪素数** (也可以称为概率素数) 指包含素数但比素数集大的数集 (素数的必要条件)，虽然伪素数判定不能判定素数，但可以排除合数，或者判定大概率是素数，还是举些例子比较好。根据费马小定理，如果 p 是素数，则必有 $b^{p-1} \equiv 1 (\mathrm{mod} p)$，但反过来并不成立，由此就能给出一种伪素数。

定义 8.1

如果 n 满足 $b^{n-1} \equiv 1(\mathrm{mod}n)$，就称 n 是一个 **b-基 (费马) 概率素数** (base-b probable prime)。如果合数 n 是一个 b-基概率素数，就称它是一个 **b-基 (费马) 伪素数** (base-b pseudoprime)。

例如有 $2^{341-1} \equiv 1(\mathrm{mod}341), 341 = 11 \times 31$，因此 341 是一个费马伪素数，考虑 2000 内的所有 2-基伪素数为

$$\{341, 561, 645, 1105, 1387, 1729, 1905\}, \pi(2000) = 303, \frac{7}{303} \approx 0.023$$

可见伪素数的占比算是较低的了。上面的 b 是存在性的，如果我们进行遍历，则有范围更小的一种伪素数。

定义 8.2

如果合数 n 满足 $\forall \gcd(b, n) = 1, b^{n-1} \equiv 1(\mathrm{mod}n)$，则称它是卡迈克尔数 (Carmichael number)。

它的占比就更小了，在 40000 以内所有的卡迈克尔数为

$$\{561, 1105, 1729, 2465, 2821, 6601, 8911, 10585, 15841, 29341\}, \pi(40000) = 4203, \frac{10}{4203} \approx 0.0024$$

当然我们并不需要真的遍历所有的 b 来判定卡迈克尔数，考虑因子即可，显然如果 $(a, n) \neq 1 \Rightarrow (ka, n) \neq 1$，由此可以削减到与 n 的质因子均互素的数，或者说我们有下面的判定定理。

定理 8.3

合数 $n > 2$ 是卡迈克尔数当且仅当，有分解 $n = \prod_{i=1}^{k} p_i, k \geq 3(p_i$ 互不相同) 并且每个奇素数 p_i 满足 $\forall 1 \leq i \leq k, p_i - 1 \mid n - 1$。

简单来说我们将“卡迈克尔数判定”转化成了“因数分解 + 整除判定”。下面是几个判定的实例

$$29341 = 13 \times 37 \times 61, 13 - 1 \mid 29341 - 1, 37 - 1 \mid 29341 - 1, 61 - 1 \mid 29341 - 1, 29341\text{是迈克尔数}$$

$$341 = 11 \times 31, k < 3 \Rightarrow 341\text{不是迈克尔数}$$

$$1905 = 3 \times 5 \times 127, 3 - 1 \mid 1905 - 1, 5 - 1 \mid 1905 - 1, 127 - 1 \nmid 1905 - 1 \Rightarrow 1905\text{不是迈克尔数}$$

由于剩下的篇幅不多了[1]，对于强伪素数检验、欧拉伪素数检验、卢卡斯伪素数简单、ECPP(Elliptic Curve Primality Proving) 等，我就放个关键词，然后自己看着办吧[2]。

8.3 散列函数

可怕的不是自己的无知，而是对自己的无知一无所知。简介的作用是告诉你有这么一个东西的存在，而不是真的要教会你这个东西，如果你仅仅通过简介就能搞懂这些复杂的理论，那原来的教程还有存在的必要吗？另外，如果你只想通过简介或科普就搞懂一个深刻的理论的话，我只能希望你不要成为“民科”。没错，我们最后的主题竟然不是密码学 (Cryptography)，而是密码学里的散列函数 (Hash Function)。对于密码学我只要告诉你三个字，你就可以自己去学习了，但我敢保证散列函数的地位会被你忽略掉。

[1] 老板要求不能超过 100 页，她说这是概括性的文章，尽可能写精简就行了

[2] 自己用各种方式去查找，应该不需要我来教吧

定义 8.3

设 R,S,T 是三个有限非空集，对每个 $r\in R$ 有映射 $h_r:S\to T$，则称 h_r 是 S 到 T 的散列函数。R 的元素称为密钥 (key)。如果 $h_r(s)=t$，则称为 s 在 r 下被散列到 t。

♣

定义其实没啥用，无非就是把一个东西转化为另一个东西，还是得看一些具体得例子，比如计算机里常用的 SHA-1 和 MD5，它们将一个文件散列成一串十六进制数，也被称为文件校验码，不过我个人认为文件摘要是个更好的称呼。这两者算法最大的特点是尽可能保证唯一性的情况下去压缩内容，从而形成几乎独一无二的数字签名。在 MD5 算法中，S 是各种文件的集合，T 是在某个范围内的自然数，至于密钥实际就是 MD5 中选取的几个常数，这些常数本身就是在某个范围内可以变动的，只不过由于计算机规范性的要求，使得在算法中，几个常数变得固定了。散列函数的另一个例子是校验位，比如身份证和 ISBN 的最后一位，不过校验这件事在计算机里基本无处不在，比如下载大量文件后会在本地进行某种摘要计算来与云端进行对比从而有很大的把握判定下载数据是完整的，这其实基于几个显然的事实，传输不会大量丢失内容，小量的数据变换会引起摘要的剧烈变化。

如果我们给出一串规律数 $\{1,2,...,n\}$，那么其经过散列后可能就变得没那么有规律了，从而我们得到了一系列看似随机的数，我们称其为伪随机数。伪随机数的生成方法其实很多，比如最简单的线性同余生成法

$$x_{n+1}\equiv ax_n+c(\mathrm{mod}m)$$

通常伪随机都需要一个种子 x_0 作为开始，一般都是借助现实的随机内容，比如系统时之类的，稍微想想也明白，如果没有种子，那么在计算机固定的算法下，生成的结果也是固定的，自然就谈不上随机了。下面是线性同余生成得到随机数的示例

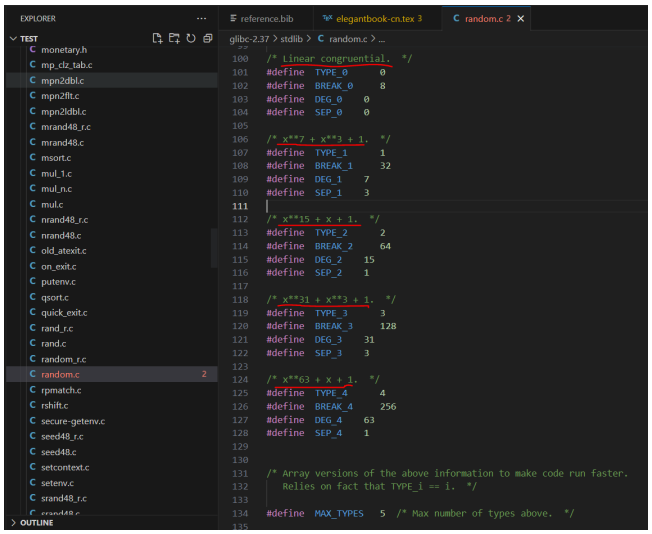
$$m=9,a=7,c=4,x_0=3,\{3,7,8,6,1,2,0,4,5,3,...\}$$

更进一步，还有类似的乘法同余生成

$$x_n\equiv ax_{n-1}^b(\mathrm{mod}m)$$

相应的简单例子为

$$a=1,b=2,m=209,x_0=6,\{6,36,42,92,104,157,196,169,137,...\}$$



可见 C 标准库使用的也是同余生成器 $x_{n+1}\equiv f(x_n)(\mathrm{mod}m),f(x)\in\mathbb{Z}[x]$，其中可以选择 $f(x)=x^7+x^3+1,x^{15}+x+1,x^{31}+x^3+1,x^{63}+x+1$。至此，我们的“数论大观园”结束了。

参考文献

[1] Titu Andreescu and Dorin Andrica. “Quadratic Diophantine Equations”. In: 2015.

[2] Titu Andreescu, Dorin Andrica, and Ion Cucurezeanu. “An Introduction to Diophantine Equations”. In: 2002.

[3] Edward J. Barbeau. “Pell’s Equation”. In: 2003.

[4] Kalyan Chakraborty and Takao Komatsu. “Generalized hypergeometric Bernoulli numbers”. In: *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas* 115 (2021).

[5] J. Coates and R. Sujatha. *Cyclotomic Fields and Zeta Values*. Springer-Verlag, 2006.

[6] Teresa Crespo and Zbigniew Hajto. “Algebraic Groups and Differential Galois Theory”. In: 2011.

[7] Abhijit Das. “Computational Number Theory”. In: 2013.

[8] William Ellison. “Waring’s Problem”. In: *American Mathematical Monthly* 78 (1971), pp. 10–36.

[9] F.Beukers. “A refined version of the Siegel-Shidlovskii theorem”. In: 2004.

[10] István Gaál. “Diophantine Equations and Power Integral Bases”. In: 2019.

[11] Alfred Geroldinger and Imre Z. Ruzsa. “Combinatorial Number Theory and Additive Group Theory”. In: 2009.

[12] D. A. Goldston, Janos Pintz, and C. Y. Yildirim. “Primes in tuples I”. In: *Annals of Mathematics* 170 (2005), pp. 819–862.

[13] D. A. Goldston, Janos Pintz, and Cem Yalçin Yildirim. “Primes in tuples II”. In: *Acta Mathematica* 204 (2007), pp. 1–47.

[14] Richard K. Guy. “Unsolved Problems in Number Theory”. In: 1981.

[15] Harald Helfgott. “The ternary Goldbach conjecture is true”. In: *arXiv: Number Theory* (2013).

[16] Marc Hindry and Joseph H. Silverman. “Diophantine Geometry: An Introduction”. In: 2000.

[17] David Hubbard and Lawrence C. Washington. “Iwasawa invariants of some non-cyclotomic Zp-extensions”. In: *Journal of Number Theory* 188 (2018), pp. 18–47. ISSN: 0022-314X. DOI: https://doi.org/10.1016/j.jnt.2018.01.009. URL: https://www.sciencedirect.com/science/article/pii/S0022314X18300398.

[18] Hideaki Ikoma, Shu Kawaguchi, and Atsushi Moriwaki. *The Mordell Conjecture: A Complete Proof from Diophantine Geometry*. Feb. 2022. ISBN: 9781108845953. DOI: 10.1017/9781108991445.

[19] Graham J. O. Jameson. “The Prime Number Theorem”. In: 2003.

[20] Masanobu Kaneko, Tomoyoshi Ibukiyama, and Tsuneo Arakawa. “Bernoulli Numbers and Zeta Functions”. In: 2014.

[21] Serge. Lang. *Cyclotomic fields I and II*. Springer-Verlag, 1990.

[22] Michel Langevin. “Méthodes élémentaires en vue du théorème de Sylvester”. In: 1976.

[23] M. Murty and Brandon Fodden. *Hilbert’ s Tenth Problem - An Introduction to Logic, Number Theory, and Computability*. American Mathematical Society, 2019.

[24] Mauro Di Nasso, Isaac Goldbring, and Martino Lupini. “Nonstandard Methods in Ramsey Theory and Combinatorial Number Theory”. In: 2017.

[25] Marius van der Put and Michael F. Singer. “Galois Theory of Linear Differential Equations”. In: 2012.

[26] Joseph Rabinoff. “The Theory of Witt Vectors”. In: *arXiv: Number Theory* (2014).

[27] Alain M. Robert. “A Course in p-adic Analysis”. In: 2000.

[28] Jacques Sauloy. “Differential Galois Theory Through Riemann-hilbert Correspondence: An Elementary Introduction”. In: 2016.

[29] Pascal Sebah. “Collection of formulae for Euler’ s constant γ”. In: 2002.

[30] Z.I. Borevich，Igor R. Shafarevich. *Number theory*. Academic Press, 1986.

[31] Elias M. Stein. “Complex Analysis”. In: 2003.

[32] Gérald Tenenbaum. “Introduction to Analytic and Probabilistic Number Theory”. In: 1995.

[33] L. Washington. *Introduction to Cyclotomic Fields*. Springer-Verlag, 1997.

[34] Tomohiro Yamada. “Explicit Chen’s theorem”. In: *arXiv: Number Theory* (2015).

[35] Yitang Zhang. “Bounded gaps between primes”. In: *Annals of Mathematics* 179 (2014), pp. 1121–1174.

[36] 于秀源. 超越数论基础. 哈尔滨工业大学出版社, 2011.

[37] 雷蒙德 M. 斯穆里安，著. 余俊伟，译. 哥德尔不完全性定理. 科学出版社, 2019.

[38] 佩捷，郭梦舒. 从华林到华罗庚：华林问题的历史. 哈尔滨工业大学出版社, 2013.

[39] 冯克勤. 代数数论. 科学出版社, 2000.

[40] 朱尧辰，徐广善. 超越数引论. 科学出版社, 2003.

[41] 潘承彪. 从切比雪夫到爱尔特希（上）：素数定理的初等证明. 哈尔滨工业大学出版社, 2013.

[42] 潘承洞. 模形式导引. 北京大学出版社, 2002.

[43] 潘承洞，潘承彪. 解析数论基础. 科学出版社, 1997.

[44] 陶哲轩，著. 王昆扬，译. 陶哲轩实分析. 人民邮电出版社, 2008.

[45] 逯晓零. 代数分析几何简介. https://lixing48.gitee.io/download/AlgebraicAnalyticGeometry.pdf.

[46] 逯晓零. 朗兰兹纲领简介. https://lixing48.gitee.io/download/LanglandsProgram.pdf.

[47] 黑川信重，栗原将人，斋藤毅，著. 印林生，胥鸣伟，译. 数论 *II*：岩泽理论和自守形式. 高等教育出版社, 2009.

[48] 黑川信重，栗原将人，斋藤毅，著. 印林生，胥鸣伟，译. 数论 *I*：*Fermat* 的梦想和类域论. 高等教育出版社, 2009.