A Garden of Mathematics

riversky

2019-2023

前言：函数的究极定义（数学哲学向）

现代数学大厦建立在**集合论** (set theory) 上. 从集合论的角度看来：

- 原始观念(primitive notion)，即不能被其他更基本的概念定义的概念，只有集合(set)、类(class) 和属于(“$\in$”) 三个，其中“属于”是二元谓词，用来连接数学对象.
- 其他一切数学对象可以解释为集合或类，其他一切谓词都可解释为基本的逻辑连接词（或、且、非）以及“属于”.

例如命题“$1+1=2$”在集合论中改写为（在 ZF 集合论和冯 · 诺依曼序数定义的框架下）

$$\{\emptyset\}\cup\{\{\emptyset\}\}=\{\emptyset,\{\emptyset\}\}$$

上式仍然是“给人看”的表达式，还可以继续形式化展开，用一阶形式语言描述上面这个式子，最终结果为：

$(\exists x(x\in l)\wedge\neg\exists y((y\in x)\wedge(x\in l)))\wedge$

$(\neg\exists x(((x\in m)\wedge(\exists y(((y\in x)\wedge\neg(y\in l))\vee(\neg(y\in x)\wedge(y\in l)))))\vee(\neg(x\in m)\wedge(\exists y((y\in x)\wedge(y\in l))))))\wedge$

$(\neg\exists x(((x\in l)\vee(x\in m))\wedge\neg(x\in n))\vee(\neg((x\in l)\vee(x\in m))\wedge(x\in n)))\wedge$

$(\exists x(\exists y((y\in x)\wedge(x\in p)\wedge(y\in p)))\wedge(\neg\exists z(\exists x(\exists y((z\in y)\wedge(y\in x)\wedge(x\in p)\wedge(y\in p))))\wedge$

$(\neg\exists x(((x\in n)\wedge\neg(x\in p))\vee(\neg(x\in n)\wedge(x\in p))))$

有序对(ordered pair)，形式上写作 $\langle x,y\rangle$，无论怎么定义，都要满足“提出有序对这个概念时希望它满足的性质”：

$$\langle a,b\rangle=\langle c,d\rangle\Leftrightarrow a=c\wedge b=d$$

所以目前采用的定义是

$$\langle x,y\rangle:=\{\{a\},\{a,b\}\}$$

易验证它满足以上性质，其实我们曾经就见过有序对，只不过那时我们把它叫做“坐标”.

在有序对的基础上，定义两个集合之间的直积(Cartesian product) 为有序对组成的集合，通俗地理解，左元来自第一个集合、右元来自第二个集合的所有可能的有序对组成的集合就是两个集合的直积.

$$A\times B:=\{\langle x,y\rangle|x\in A\wedge y\in B\}$$

例如 $A=\{1,2\},B=\{3,4,5\}$，则 $A\times B=\{\langle 1,3\rangle,\langle 2,3\rangle,\langle 1,4\rangle,\langle 2,4\rangle,\langle 1,5\rangle,\langle 2,5\rangle\}$.

二元关系(binary relation) 是直积的子集，我们常说的描述关系的语句都符合这种格式：“A 和 B 具有 R 关系”，以表明不同的对象之间存在联系. 有序对就体现了这样的联系，而同种联系可能存在不止一

对对象之间，把具有同种联系的一对对对象放入一个集合 R，则该集合就完整地刻画了这种联系. 因此 “x 和 y 具有 R 关系” 就形式地写成 $\langle x,y\rangle \in R$，有时也写作 xRy，这种写法暗示了函数的雏形.

例如已知集合 $A=\{1,2,3,4\}$，在上面建立二元关系 “$<$”，其含义同我们日常所理解的 “$<$”，因为 $1<2$、$1<3$、$1<4$、$2<3$、$2<4$ 和 $3<4$，所以 $< = \{\langle 1,2\rangle,\langle 1,3\rangle,\langle 1,4\rangle,\langle 2,3\rangle,\langle 2,4\rangle,\langle 3,4\rangle\}$，充分贯彻了集合论 “万物皆集合” 的思想.

有些特殊的二元关系 $R\subseteq A\times B$，只要左元确定后，右元也能唯一地确定，即

$$\langle x,y\rangle \in R \wedge \langle x,z\rangle \in R \Rightarrow y=z$$

就称二元关系 R 在集合 B 上具有外延性(extensionality). 明显，上例中的 “$<$” 在集合 A 上就不具有外延性. 到现在为止，函数的定义呼之欲出了.

定义　函数

具有外延性的二元关系称为函数(function) 或映射(mapping)，形式化地表达：

$$f\text{是函数} \iff \forall p(p\in f\rightarrow(\exists x\exists y(p=\langle x,y\rangle \wedge(\exists z((\langle x,z\rangle \in f)\rightarrow y=z)))))$$

- 左元组成的集合称为定义域(domain)，记作 $\mathrm{dom}(f)=\{x|\langle x,y\rangle \in f\}$；右元组成的集合称为值域(range)，记作 $\mathrm{ran}(f)=\{y|\langle x,y\rangle \in f\}$；陪域的超集称为陪域(codomain)；将函数写作 $f:A\rightarrow B$ 以表明函数 f 的定义域是 A、陪域是 B.
- 定义域中的元素或定义域的子集都称为原象(preimage)，定义域中的元素 x 所唯一确定的值域中的元素 y 称为 x 的像(image)，记作 $y=f(x)$；定义域的子集 X 所唯一确定的值域的子集 $Y=\{f(x)|x\in X\}$ 称为集合 X 在 f 下的象，也记作 $Y=f(X)$. 这两种原象和象需要注意区分，但以下将不再强调这两种原象和象的区别.
- 若 f 在定义域上也有外延性，即 $f(x)=f(y)\rightarrow x=y$，则称 f 是单射(injection)；若 $X\subseteq \mathrm{dom}(f), Y=f(Y)$，则称 $f:X\rightarrow Y$ 是满射(surjection)，注意这里需要强调定义域和值域；若 $f:X\rightarrow Y$ 既是单射又是满射，则称 f 为双射(bijection).
- 对于定义域的子集 X，$f\restriction X=\{\langle x,y\rangle|x\in X\}$ 称为 f 的限制(restriction)；在满足定义域外延性的前提下，$f^{-1}=\{\langle y,x\rangle|\langle x,y\rangle \in f\}$ 称为 f 的反函数(inversion)；对于两个函数 f,g，称 $g\circ f=\{\langle x,z|\exists y(y=f(x)\wedge z=g(x))\}$ 称为 f 和 g 的复合函数(composition).

上面的定义需要注意几点：

- 任何函数都是定义域到值域上的满射，所以 “满射” 这个概念往往伴随着 “限制” 的概念. 当定义域受到限制而对应法则不变的情况下，f 是否能射满某个集合，这一点值得讨论.
- 以前我们见到的说法都是 “给定一个函数 $f(x)\cdots\cdots$”，而在以上的定义中，f 才是函数本身，而 $f(x)$ 是元素 x 在 f 下的像，是一个值. 尽管我们在平时经常忽略它们两者之间的微小区别，而且在后面的讨论中也会接着忽略，但这个区别正是接下来从 λ 演算和类型论的角度看待函数的引子.

先把集合论放在一边，看看 λ **演算** (λ-calculus) 是怎么看待函数的. 集合论将函数定义为一个集合，常常命名为 f. 但这样的处理方法有一些缺点，我们常说函数是一个对应法则，单从这个 f，或者将 f 的元素逐个列举出来（如果可能的话），很难看出这究竟是一个怎样的法则. 而在 λ 演算中，使用 $\lambda x.M[x]$ 来代替 f 本身，其中 $M[x]$ 是一个和 x 有关的语句. 例如函数 $f(x)=x^2$，从计算的角度来看，就是把

输入的变量取平方然后输出. 在 λ 演算中，就用 $\lambda x.x^2$ 来代替这个 f，此时 $f(5) = 25$ 这个命题就写成 $(\lambda x.x^2)(5) = 25$. 这种写法既可以表现函数的对应法则，又可以免去为函数命名的麻烦.

集合论关心的是某个对象属不属于某个集合的问题，**类型论** (type theory) 则关心某个对象具有什么样的类型的问题. 例如上文的 $f(5) = 25$ 中，这个 "5" 究竟是以自然数还是实数的身份放入函数参与运算的呢? 毕竟函数是一个集合，定义在自然数和定义在实数上的函数是不同的. 而且在之后 "实数集的构造" 一章中就可以了解到，从集合论的角度看来，实数的 "5" 和自然数的 "5" 是两个完全不同的集合，前者要比后者复杂得多. 其实这种事情其实对不关心数学大厦基础的人来说完全就是自作多情，希尔伯特之前的数学家也是这么想的，但这个基础不牢固会动摇上层（例如集合论中的 Mediate 基数可以引出和古典分析学完全对立的理论）.

在类型论中，说明某个对象是某种类型的语句称为判断(judgement)，是类型论中最基本的语句，谓词是冒号 ":"，对应集合论中的 "属于" 符号. 例如 $a : \mathsf{nat}$ 和 $a : \mathsf{real}$ 分别表示对象 a 为自然数和实数类型. 而对于函数，$\lambda x^{\mathsf{nat}}.(x^2)^{\mathsf{real}}$ 意味着这个函数输入为自然数类型的对象，输出为实数类型的对象，而该函数的类型为 $\mathsf{nat} \to \mathsf{real}$. 不过大多数时候，用 $(\lambda x^{\mathsf{nat}}.x^2) : \mathsf{nat} \to \mathsf{real}$ 来表示输出为实数类型.

前言：超好用的逻辑函数

- Kronecker δ 函数：$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$
- Iverson 括号：$\{P\} = \begin{cases} 1, & P\text{是真命题} \\ 0, & P\text{是假命题} \end{cases}$
- 示性函数：$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$

Contents

第一部分 数理逻辑和集合论

在高等数学中，场论和多元函数微积分建立在一元函数微积分之上，后者又建立在极限理论上. 一层层搭起来，你也许会好奇数学大厦最底层架构究竟是怎样的. 其实数学大厦并没有什么“最底层架构”的说法，因为这已经涉及到了哲学的层面. 但目前的主流是将数理逻辑和公理化的集合论作为底层架构，从而推演出数学大厦其他部分，而集合论公理的选取又以策梅洛-弗兰克尔集合论(ZF) 为主流，在大部分时候会加上选择公理 (AC). 此外，冯 · 诺伊曼 - 博内伊 - 哥德尔集合论 (NBG) 中“类”的概念在某些时候又特别有用，所以经常也一并讨论，并将选择公理的适用范围推广到了“类”上. 公理化集合论的前身是朴素集合论，曾触发了罗素悖论而被很多数学家，后来引入了公理化方法规避了罗素悖论才得以重新启用. 然而公理化方法是形式逻辑的内容，其在定义各种概念时引入了朴素集合论的相关概念. 所以接下来首先介绍朴素集合论，然后讨论数理逻辑的内容，最后到公理集合论.

1 数学第一步，朴素集合论

朴素集合论于 19 世纪 70 年代由康托尔 (G. Cantor) 开始研究，它对“什么是集合”没有限制，把任意对象放在一起都可以形成集合(set)，而这些对象就称为元素(element)，记作“元素 $\in$ 集合”. 具体来说，关于“什么是集合”的问题，有如下概括原则(comprehensive scheme)：

$$\text{对于任意命题}\varphi\text{，}\{x|\varphi\}\text{是一个集合.}$$

这样构造集合的方法我们称为描述法，当然还有列举法，欲说明集合 A 中只含有 $1,2,3,4$ 这 4 个元素，写作 $A=\{1,2,3,4\}$ 即可. 使用列举法来描述集合时，元素的书写顺序是无序的，也就是说 $\{1,2,3\}$ 和 $\{1,3,2\}$ 是相同的集合，同时我们也要求列举时元素不能重复，例如 $\{1,2,2,3\}$ 就是错误的列举，不过我们不必画地为牢，可以认为它等于 $\{1,2,3\}$，但不是最简的表达而已，剧透：公理集合论中的外延公理肯定了这种想法.

除了“属于”以外，还有“包含”的概念. 如果集合 A 的元素在 B 中都能找到，就称 A 是 B 的子集(subset)、B 是 A 的超集(superset)，或称 B 包含 A，记作 $A\subseteq B$，欲进一步表明 $A\neq B$，则写作 $A\subset B$.

有两个特殊的集合，第一个什么元素都没有，称为空集(empty set)，记作 $\emptyset$；第二个集合是所有集合的集合，可以用描述法表示为 $\{x|x=x\}$，记作 $\mathbb{V}$，称为大全集或者集合宇宙(universe).

此外，集合有几种基本运算，集合 A 和 B 的并集(union) 由 A 的全部元素和 B 的全部元素组成，没有其他元素，记作 $A\cup B$；集合 A 和 B 的交集(intersection) 由 A 和 B 共有的元素组成，记作 $A\cap B$；A 和 B 的差集(difference) 由属于 A 但不属于 B 的元素组成，记作 $A-B$，有一个类似的概念称为补集(complement)，B 相对于 A 的补集就是 $A-B$，接下来不使用这个概念；A 和 B 的对称差集(symmetric difference) 由只属于 A 和 B 其中一个的元素组成，等价于 $(A-B)\cup(B-A)$，记作 $A\triangle B$，这个概念很少用到；集合 A 的幂集(power set) 由 A 的子集构成，记作 $\mathcal{P}(A)$. 集合的并、交、补和命题逻辑的或、且、非表达的意思相近，但根据定义可以看出来，后者更为基本.

虽然朴素集合论提出的概念非常少，但是我们已经可以借助它完成很多事情了.

2 关系理论

2.1 一般的二元关系

集合中的元素书写顺序是无序的，因此我们希望发明出一种符号，也像集合由多个对象组成，但将其元素列举出来时的书写是有序的. 坐标就是这样的记号，例如 $(1,2,3)\neq(1,3,2)$. 考虑到集合论自诞生以

来便有一统数学江湖之势，我们希望能将坐标看作一种特殊的集合. 目前大多采用 Kuratowski 的定义：

$$(a,b):=\{\{a\},\{a,b\}\}$$

可以证明这种定义时有效的：当且仅当 $a=c,b=d$ 时，才有 $(a,b)=(c,d)$. 对于含有更多元素的有序组，按照以下这种方式定义：

$$(a,b,c):=(a,(b,c))$$

$$(a,b,c,d):=(a,(b,c,d))$$

定义好了有序对，我们便可以讨论一些高级话题了. 定义两个集合的直积(cartesian product)：

$$A\times B:=\{(a,b)|a\in A且b\in B\}$$

通俗地理解，左元来自第一个集合、右元来自第二个集合的所有可能的有序对组成的集合就是两个集合的直积. 例如 $A=\{1,2\},B=\{3,4,5\}$，则 $A\times B=\{(1,3),(2,3),(1,4),(2,4),(1,5),(2,5)\}$.

二元关系(binary relation) 是直积的子集，我们常说的描述关系的语句都符合这种格式："A 和 B 具有 R 关系"，以表明不同的对象之间存在联系. 有序对就体现了这样的联系，而联系是普遍的，同种联系可能存在不止一对元素之间，把具有同种联系的多对元素放入一个集合 R，则该集合就完整地刻画了这种联系. 因此"x 和 y 具有 R 关系"就形式地写成 $(x,y)\in R$，有时也习惯性地写作 $R(x,y)$ 或 xRy.

例如已知集合 $A=\{1,2,3,4\}$，在上面建立二元关系"$<$"，其含义同我们日常所理解的"$<$"，因为 $1<2$、$1<3$、$1<4$、$2<3$、$2<4$ 和 $3<4$，所以 $<\ =\ \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$，充分贯彻了集合论"万物皆集合"的思想. 接下来给出关系的正式定义.

定义　二元关系

对于集合 R，如果存在 X,Y，使得 $R\subseteq X\times Y$，则称 R 为二元关系(binary relation). 如果 $X=Y$，即 $R\subseteq X^2$，则可以说该关系装备(equipped) 在集合 X 上，此时可将该集合及其关系一并记作 $\langle X;R\rangle$.

- 左元组成的集合称为定义域(domain)，记作 $\mathrm{dom}(R)=\{x|(x,y)\in R\}$；右元组成的集合称为值域(range)，记作 $\mathrm{ran}(R)=\{y|(x,y)\in R\}$；值域的超集称为陪域(codomain).
- 定义域的子集 X 中的元素作为左元放入 R 得到的右元集合 $Y=\{y|xRy\}$ 称为 X 在 R 下的像(image)，记作 $Y=R[X]$；反过来，集合 Y 的元素作为右元放入关系 R 得到的左元集合 $X=\{x|xRy\}$ 称为 Y 在 R 下的逆像(preimage)，记作 $X=R^{-1}(Y)$.
- 对于两个关系 R,S，称 $S\circ R=\{(x,z)|xRy\text{ 且}yRz\}$ 为 R 和 S 的复合(composition).
- 对于关系 R，若 xRx 恒成立，则称 R 是自反的(reflexive)；若 xRx 恒不成立，那就是反自反的(anti-reflexive).
- 对于关系 R，若 xRy 成立能推出 yRx 成立，则称 R 是对称的(symmetric)；若 xRy 与 yRx 只有一个成立，或两者都成立时能推出 $x=y$，则称 R 是反对称的(antisymmetric).
- 对于关系 R，若 xRy 和 yRz 成立能推出 xRz 成立，则称 R 是传递的(transitive).
- 对于关系 R，若存在唯一的元素 x_0，使得 xRx_0 能推出 $x=x_0$，则称 R 是良基的(well-founded)，而该元素称为关于 R 的最小元（如果觉得难以理解可以试试把 R 看成 $\leq$）.
- 对于关系 $R\subseteq X\times X$，若 X 任意两个元素 x,y 满足 $xRy,yRx,x=y$ 三者至少其一，则称 R 在 X 上是连通的(connective)（X 中的任意两个元素都会被 R 触及到）.

需要注意，自反性与反自反性可以两者都不具有，即 xRx 有时成立有时不成立；对称性与反对称性也是同样的道理.

2.2 序关系

序关系(ordering) 作为一个大类没有统一的定义，但共同点是**装备在某个集合上**，且**至少拥有传递性**，而在满足传递性的基础上，如果还具有更多的性质，那么就有不同的名称，给个定义：

定义 各种各样的序关系

具有反自反性的序关系也称为严格的(strict) 序关系，否则就是不严格的序关系（例如 $<$ 与 $\leq$），**接下来如果没有特别强调，默认为非严格的序关系，对自反性不做要求**.
具有传递性的二元关系称为预序关系(quasi-ordering)，是要求最宽的序关系；
具有传递性、反对称性的序关系称为偏序关系(partial ordering)；
具有传递性、反对称性、连通性的二元关系称为全序关系(total ordering) 或线序关系(linear ordering)；
具有传递性、反对称性、连通性、良基性的二元关系称为良序关系(well-ordering)；
具有传递性、自反性、对称性的二元关系称为等价关系(equivalence relation)；

对于装备在集合 X 上的序关系 R，将组合 $\langle X;R\rangle$ 统称为有序集(ordered set)，当然有时候会省略掉符号 R，就简单将 X 称为有序集，然后用 “$<$” 之类的符号来表示 R 上的序关系. 如果关系 R 满足更多的性质，$\langle X;R\rangle$ 也可能会被称为偏序集、全序集、良序集等. 但凡 R 能够达到偏序的要求，就可以定义出以下这些概念.

定义 偏序关系的 8 个特殊元

如果 $\langle X;R\rangle$ 是偏序集，$S\subseteq X$，那么对于 $a\in X$：

- 如果 $a\in S$，且任意 $x\in S$ 均不满足 aRx，那么 a 称为 S 的极大元(maximal)，如果进一步地均满足 xRa，那么 a 称为 S 的最大元(greatest element).
- 如果 $a\in S$，且任意 $x\in S$ 均不满足 xRa，那么 a 称为 S 的极小元(minimal)，如果进一步地均满足 aRx，那么 a 称为 S 的最小元(least element).
- 如果任意 $x\in S$ 均满足 xRa，那么 a 称为 S 的上界(upper bound)；S 的所有上界中的最小元（如果存在的话）称为 S 的上确界(supremum)，记作 $\sup S$.
- 如果任意 $x\in S$ 均满足 aRx，那么 a 称为 S 的下界(lower bound)；S 的所有下界中的最大元（如果存在的话）称为 S 的下确界(infimum)，记作 $\inf S$.

如果难以理解，把 R 改写成 $<$ 就好理解多了. 需要注意，S 的极大元意味着 S 中的元素要么小于 a，要么和 a 不可比较（注意 R 不一定是全序，不是所有元素都可以比较），最大元则意味着 S 中的元素都可以和 a 相比较（这也不需要 R 是全序关系），且都小于 a. 极大元可以拥有多个，而最大元唯一. 最大元一定是极大元，反之不然. 如果 R 是全序，那么极大元与最大元等价，是唯一的.

定义 集合的切片

设 $\langle X;\leq\rangle$ 是偏序集，若 $S\subseteq X$，且不存在 $x\in X$ 对全体 $s\in S$ 都满足 $x\geq s$，那么称 S 在 X 中是无界的(unbounded).
设 $\langle X;\leq\rangle$ 是偏序集，若 $S\subseteq X$，且对于任意 $x\in X$ 都存在 $s\in S$ 使得 $s\geq x$，那么称 S 在 X 中是占优的(dominating).
设 $\langle X;\leq\rangle$ 是全序集，A,B 是其两个非空子集并满足 $A\cup B=X$ 且 $A\cap B=\emptyset$，若对于任意 $a\in A,b\in B$ 都有 $a\leq b$，那么称该划分是一个分割(cut).
设 $\langle X;\leq\rangle$ 是良序集，$x\in X$，将小于 x 的元素的集合 $\{t|t\leq x\}$ 称为 x 的前段(segment)，记作 $\mathrm{seg}(x)$.

稍微学过一点数学分析的话，你可能会敏锐地察觉到“分割”这个概念就是戴德金分割在全序集上的推广版本.

2.3 函数关系和指标系统

有些特殊的二元关系 $R\subseteq A\times B$，只要左元确定后，右元也能唯一地确定，就称二元关系 R 在集合 B 上具有外延性(extensionality). 初等数学中的函数恰好具有这样的特性，以下给出函数在集合论上的定义.

定义 函数

左元能唯一确定右元的二元关系称为函数(function) 或映射(mapping)，形式化地表达:

$$f\text{是函数}\iff \forall p(p\in f\to(\exists x\exists y(p=(x,y)\wedge(\exists z(((x,z)\in f)\to y=z)))))$$

- 将函数写作 $f:A\to B$ 以表明函数 f 的定义域是 A、陪域是 B；将函数写作 $f:x\mapsto y$ 表示 $(x,y)\in f$，大多数情况下也写作 $y=f(x)$.
- 定义域中的元素称为原象(preimage)，定义域中的元素 x 所唯一确定的值域中的元素 y 称为 x 的像(image).
- 若 f 在定义域上也有外延性，即 $f(x)=f(y)\to x=y$，则称 f 是单射(injection)；若 $X\subseteq\mathrm{dom}(f),Y=f(Y)$，则称 $f:X\to Y$ 是满射(surjection)，注意这里需要强调定义域和值域；若 $f:X\to Y$ 既是单射又是满射，则称 f 为双射(bijection).
- 对于定义域的子集 X，$f\restriction X=\{\langle x,y\rangle|x\in X\}$ 称为 f 的限制(restriction)；在满足定义域外延性的前提下，$f^{-1}=\{\langle y,x\rangle|\langle x,y\rangle\in f\}$ 称为 f 的反函数(inversion)；对于两个函数 f,g，称 $g\circ f=\{\langle x,z|\exists y(y=f(x)\wedge z=g(x))\}$ 称为 f 和 g 的复合函数(composition).
- 将所有函数 $f:A\to B$ 的集合记作 B^A.

对于多元函数 $y=f(x_1,x_2,\cdots,x_n)$，我们可以构造 $X=(x_1,x_2,\cdots,x_n)$ 这个有序组，此时该函数就可以写成 $y=f(X)$，化成了一元函数. 此外，在介绍直积的时候我们说 n 个集合 A 的直积 $A\times A\times\cdots\times A$ 可以记作 A^n，而在这里如果我们将数字 n 看成一个集合，全体从 n 到 A 的函数的集合也记作 A^n，看起来两者不同，到后面介绍序数的时候你就会发现两者其实是一致的.

借助函数的概念，接下来系统地讨论一种在数学的任何领域都非常有用的记号——指标系统. 在初等数学中，我们知道数列实际上也是函数，以有限长的数列 $\{a_1,a_2,\cdots,a_n\}$ 为例，它暗含了一个函数关系 $i\mapsto a_i$，在这里 i 是整数且 $1\leq i\leq n$，将数列的每一项写出来相当于列举了它的值域. 在更一般的场合，

我们不必限制 i 是整数，而可以取自任意一个集合 I，此时某个函数 $f: i \mapsto x_i$ 的值域就可以写作 $\{x_i\}_{i\in I}$. 这一整套记号称为指标系统，而函数的定义域 I 称为指标集(indexed set).

有了指标系统，我们可以将任意集合 X 写成 $\{x_i\}_{i\in I}$ 的形式，这算是描述法和列举法以外又一种构造集合的方式. 指标系统还体现在集合的运算上. $X_1 \cup X_2$ 可以写成 $\bigcup\limits_{i\in\{1,2\}} X_i$ 的形式，交集、直积等运算类似. 你可能会觉得这种写法在无病呻吟，但我们可以参考初等数学中将 $x_1 + x_2 + \cdots + x_n$ 写成 $\sum\limits_{i=1}^{n} x_i$ 的例子，这种写法的优越性就在于多个对象参与运算的时候表达仍然是简洁且精确的，推广到无穷时还可以衍生出级数的概念. 类似地，无穷多个集合参与运算的时候，利用指标系统写起来仍然是简洁且精确的，例如

$$\bigcup_{i\in\mathbb{N}^+} X_i = \{x|(\exists i \in \mathbb{N}^+)(x \in X_i)\}$$

但是如果写成 $X_1 \cup X_2 \cup X_3 \cup \cdots$ 的形式，表述就会不精确，谁也不知道省略号到底省略了多少个对象，尤其是当接下来引入序数以后，“无穷”这个概念也需要讨论，进一步增加了这种写法的不确定程度.

由于任意集合 X 都可以写成 $\{x_i\}_{i\in I}$ 的形式，所以我们定义 $\bigcup X := \bigcup\limits_{i\in I} x_i$，表示 X 的全体元素取并集，这种运算称为 X 的“大并集”(big union)，或直接称作并集，其他运算类似. 从现在开始并集、交集、直积等运算不再仅仅是两个集合之间的运算，下文如果你看到了“X 的交集”，要立即反应过来其意思是 X 中的所有元素取交集.

有了指标系统，我们可以发现直积和函数之间的关系. 设指标集是 $I = \{a, b, c\}$，有三个集合 $X_a = \{a_1, a_2\}, X_b = \{b_1, b_2\}, X_c = \{c_1, c_2\},$，则

$$\begin{aligned}
X_a \times X_b \times X_c &= \{(m, n, p)|m \in X_a, n \in X_b, p \in X_c\} \\
&= \{ \\
&(a_1, b_1, c_1), (a_1, b_1, c_2), (a_1, b_2, c_1), (a_1, b_2, c_2) \\
&(a_2, b_1, c_1), (a_2, b_1, c_2), (a_2, b_2, c_1), (a_2, b_2, c_2) \\
&(a_3, b_1, c_1), (a_3, b_1, c_2), (a_3, b_2, c_1), (a_3, b_2, c_2) \\
&\}
\end{aligned}$$

将数列推广到指标系统后，失去了一个好的性质：数列的下标是整数，可以比较大小，而指标不一定. 这个问题将在后文得到解决，引入序数、序型和序列等概念之后便可初见端倪，最终在讨论选择公理时证明良序定理以彻底解决.

2.4 等价关系

等价关系也是一种重要的二元关系. 等价关系具有自反性、对称性和传递性，是一种特殊的序关系，但是它具有一些其他序关系都不具有的特殊性质. 值得拿出来单独讨论.

对于一般的序关系 R，若 $(a, b) \in R$ 或 $(b, a) \in R$，那么就说 a 和 b 是可比的(comparable)，全序关系意味着所有元素两两之间都是可比的. 对于等价关系 R，自反性意味着自己和自己是可比的，传递性意味着如果 a_1 和 a_2 可比、a_2 和 a_3 可比、a_3 和 a_4 可比等等，不管这个“可比链条”有多长，链条上的任意两个元素（比如 a_1 和 a_5，a_{114514} 和 $a_{1919810}$）都是可比的，而只要某个元素 b 和某一个 a_i 不可比，那么立马就可以说 b 就和所有的 a_i 不可比. 可以说等价关系划出了一个个小圈子，同一个圈子内的元素两两可比，不同的圈子完全没有关系. 更重要的是对称性，使得圈内所有元素都是“平等”的，可以互换位置. 现在我们先正式定义这样的“小圈子”.

定义 等价与划分

对于任意集合 A，若存在一系列集合 $\{S_i\}_{i\in I}$，使得 $i \neq j \Leftrightarrow S_i \cap S_j = \emptyset$（即两两不交），并且 $\bigcup_{i\in I} S_i = A$（即能互相补充成完整的 A），那么称 $\{S_i\}_{i\in I}$ 是 A 的划分(partition).

对于建立了等价关系的集合 $\langle A;\equiv\rangle$，对于任意 $x \in A$，所有与 x 等价的元素的集合 $\{y \in A|y \equiv x\}$ 称为关于 x 的等价类(equivalence class)，记作 $[x]_{\equiv}$. A 中所有等价类的集合 $\{[x]_{\equiv}|x \in A\}$ 称为 A 对 $\equiv$ 的商集(quotient set)，记作 $A/\equiv$.

3 序型与序数

3.1 序型

假设集合 $A = \{1,2,3,4\}$，把里面的每个元素都乘 2，得到 $A' = \{2,4,6,8\}$. 集合 A 中的元素满足 $1 < 2 < 3 < 4$，集合 A' 的元素满足 $2 < 4 < 6 < 8$. 在乘以 2 的过程中，我们发现：

- “$<$”是 A 和 A' 上的全序，用该全序把两个集合的元素串成一条链，链长都是 4.
- A 中的任意两个元素 x, y，如果 $x < y$，乘以 2 之后有 $2x < 2y$，不等号方向不变. 我们可以称这种性质为“乘以 2 这个操作的保序性”.

你可能感觉莫名其妙，为什么要举这个例子呢？上面两条“发现”非常显然，把它写出来有什么用呢？你肯定见过类似的情况，我们以前学数学第一次遇到负数时，。。。

定义 序同构

设有两个带关系的集合 $\langle A;R\rangle, \langle B;S\rangle$，如果存在双射 $f: A \to B$，如果 $x, y \in A$ 满足 xRy 就能推出 $f(x)Sf(y)$，那么称这两个集合是同构的(isomorphic)，该双射称为同构映射(isomorphism)；

进一步地，如果 R 和 S 是偏序关系，则称它们拥有相同的序型(order type)，记作 $\mathrm{OT}(R) = \mathrm{OT}(S)$，该同构映射又称为序嵌入(order-embedding).

因为同构这个概念涉及到的两个集合上关系的关系，是集合结构的问题，将来还会在讨论代数结构的时候深入探讨.

定义序数的动机是想构造出一个良序集 $\langle X,<\rangle$，使得 $x < y \Leftrightarrow x \in y$. 但是这样的集合不是唯一的，因为该良序的最小元尚未指定，比如 $\{x_0, \{x_0\}, \{\{x_0\}\}, \cdots\}$ 就满足条件，此时 x_0 是最小元，可以任意指定. 于是冯 · 诺伊曼加了一个条件：$x \in y \Leftrightarrow x \subset y$（满足这个条件的集合 y 称为传递集(transitive set)）. 这样立马就得到了唯一确定的最小元：空集.

定义 序数

如果 α 是传递集，且 $\langle \alpha, \in\rangle$ 是良序集，那么 α 称为序数(ordinal number). 全体序数组成的类记作 $\mathbb{ON}$.

这里为什么用“类”而不是“集合”呢？答案将在后面揭晓.

最小的序数是 $\emptyset$，除去 $\emptyset$ 后，最小的序数是 $\{\emptyset\}$，其次到 $\{\emptyset, \{\emptyset\}\}$，冯 · 诺伊曼给这些序数都做了记

号:

$$
\begin{aligned}
0 &:= \emptyset \\
1 &:= \{\emptyset\} && = \{0\} \\
2 &:= \{\emptyset, \{\emptyset\}\} && = \{0, 1\} \\
3 &:= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} && = \{0, 1, 2\} \\
4 &:= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\} && = \{0, 1, 2, 3\} \\
&\cdots
\end{aligned}
$$

可见任何一个序数 α 和比该序数大的最小序数 α^+ 都满足 $\alpha^+ = \alpha \cup \{\alpha\}$,所以将 α^+ 称为 α 的后继序数(successor ordinal).

这就导出了自然数:

定义 自然数

如果 $\alpha = 0$,或存在自然数 β 使得 $\alpha = \beta^+$,那么 α 是自然数(natural number). 全体自然数的集合称为 $\mathbb{N}$.

注意,自然数的定义是递归的,α 是自然数的前提是它是某个自然数的后继序数,因此 0 是自然数,那么 1 跟着就是自然数,2 也是自然数⋯⋯这样就导出了整个自然数集.

现在我们来讨论一下集合论中最重要的概念,也是数千年来的数学都不敢触碰的一个概念:无穷. 一些集合论的书直言:集合论就是研究无穷的数学分支. 你也许从刚才序数和自然数的定义看出了一些端倪,将序数从 0 开始列举,得到 $1, 2, 3, 4, \cdots$,似乎就是序数就是自然数. 现在来问一个关键的问题:$\mathbb{N}$ **是序数吗?**

首先 $\mathbb{N}$ 是传递集,因为任意自然数 $n = \{0, 1, \cdots, n-1\} \subset \mathbb{N}$,$\in$ 和 $\subset$ 是等价的;其次 $\langle \mathbb{N}, \in \rangle$ 是良序集,因为任意两个自然数都可以比较大小(即两个不相等的自然数必然一个属于另一个),而且 0 是最小元. 所以 $\mathbb{N}$ 是序数.

但 $\mathbb{N}$ 却不是自然数,因为对于任意自然数均有 $n \in \mathbb{N}$ 且 $n \neq \mathbb{N}$. 所以 $\mathbb{N}$ 并不是任何自然数的后继序数.

这里开始体现出了自然数与序数的差别,自然数若不是 0,就是另一个自然数的后继序数,即总可以从最小的自然数取**有限次**后继序数达到. 但序数可不一定,除了 0 和后继序数之外,还存在一类序数,不能从比它小的序数取有限次后继序数达到,但它等于比它小的所有序数的并集,这类序数称为极限序数(limit ordinal). $\mathbb{N}$ 是大于 0 的最小极限序数,为了符合序数运算法则,当我们把 $\mathbb{N}$ 当作序数来看的时候,会将它记作 ω.

3.2 归纳和递归

我们以前学过数学归纳法,它的具体内容是这样:

> 如果一个命题对 0 成立,而且只要它对自然数 n 成立就对 $n+1$ 也成立,那么这个命题对任意自然数都成立.

这种方法适用于所有关于自然数的命题，甚至许多命题在证明的时候只能用数学归纳法. 但其实这只是一系列数学归纳命题之一.

定理 数学归纳 (mathematical induction)

- 数学归纳定理，第一形式
  如果 $0 \in X$，并且任意自然数 n 都满足 "如果 $n \in X$，那么 $n+1 \in X$"，那么 $\mathbb{N} \subseteq X$.
- 数学归纳定理，第二形式
  如果任意自然数 n 都满足 "如果所有比 n 小的自然数属于集合 X，那么 $n \in X$"，那么 $\mathbb{N} \subseteq X$.
- 数学归纳原理，第一形式
  设有命题 $\phi(n)$，如果 $\phi(0)$ 成立，且任意自然数 n 都满足 "如果 $\phi(n)$ 成立，那么 $\phi(n+1)$ 成立"，那么 $\phi(n)$ 对任意自然数 n 成立.
- 数学归纳原理，第二形式
  设有命题 $\phi(n)$，如果任意自然数 n 都满足 "如果所有比 n 小的自然数 k 都满足 $\phi(k)$，那么 $\phi(n)$ 成立"，那么 $\phi(n)$ 对任意自然数 n 成立.

数学归纳 "定理" 和数学归纳 "原理" 有什么区别呢? 我们尚未讨论到数理逻辑，这里先剧透一下，数学归纳 "定理" 只断言某个元素属于某个集合，是一个定理 (theorem)，能用一阶逻辑语言描述；数学归纳 "原理" 涉及到了尚不明确的命题 ϕ，而在一阶逻辑下，"任意" 这个量词不能修饰命题、函数等对象 (二阶逻辑才可以)，即没有 "任意命题 ϕ" 这种说法，只能说 "对于不同的命题 ϕ，该句子都是一个命题"，所以这里很谨慎地写作 "设有命题 ϕ" 的格式. 对于这种本身不是定理但提供了模板、代入命题会成为定理的句子，我们把它称为 "定理模式" (theorem schema).

证明 归纳定理第一形式. 反证法，如果有哪个自然数满足 "$0 \in X$ 且 '如果 $n \in X$ 则 $n+1 \in X$'" 但不属于 X，那么所有这样的自然数中必有最小者 (良序集的子集是良序集，有最小元)，该最小者必然违背双引号框住的命题.

归纳定理第二形式. 若存在某个自然数满足 "如果所有比 n 小的自然数属于集合 X，那么 $n \in X$" 但不属于 X，那么就一定存在一个比它小的自然数不属于 X，所以可以构造集合 $\{n|n \notin X\}$，因为良序集的任意子集都是良序集，所以该集合一定有最小元，该最小元立即违背红色部分的命题.

归纳原理. 在对应的归纳定理中让集合 X 形如 $\{n|\phi(n)\}$，那么某个自然数 n 要想属于这个集合就必须满足命题 $\phi(n)$，这就证明了归纳原理. □

你也许觉得这些证明过程很水，有两点原因，一是这些命题实际上是皮亚诺算术系统的公理，是一个很基础的命题；二是该证明过程在朴素集合论语境下进行，如果拿到公理集合论当中就要考虑使用了哪些公理，到时候就不显得水了.

递归是比归纳更抽象的概念，我们可以从以前学过的数列来理解. 你一定遇到过这样的题目：给出数列的递推公式，求出它的通项公式. 这种题目根据通项公式的形式有不同的解题方法，把你学过的全列出来，那叫一个五花八门. 但实际上递推公式只要稍微复杂一些，就不一定能求出它的通项公式了，例如 $\frac{1}{n}a_{n+1} = a_n^2 + 2n$，它的通项公式没有初等形式. "没有初等形式" 和 "没有通项公式" 的区别很大，尽管没有初等形式，但递归定理保证了通项公式 (加上初值) 可以唯一确定通项公式，即完整地刻画这个数列.

上面这个例子中的递推公式整理一下，得到 $a_{n+1} = na_n^2 + 2n^2$，其中涉及的变量只有 a_n 和 n[1]，记 $g(x,y) = xy^2 + 2x^2$，那么断言 "数列 $\{a_n\}_{n\in\mathbb{N}}$ 有通项公式" 等价为 "已知 a_0 和函数 $g(a_n, n)$，a_n 可被唯

[1] 当然你认为变量只有 n 也没问题，因为我们要证明的命题就是 "映射 $n \mapsto a_n$ 是唯一确定的"，即 a_n 是 n 的因变量，真正的自变量只有 n，但是现在尚未证明，所以可将 a_n 也认为是自变量.

一确定”.

斐波那契数列则是更一般的情况，它的递推公式为 $a_n = a_{n-1} + a_{n-2}$，涉及到前两项，在这种情况下仅根据 a_0 不能唯一确定通项公式，还需要 a_1 才行；同理，如果递推公式涉及到更多之前的项，则可能需要更多的初值. 在最极端的情况下，a_n 需要 a_0 到 a_{n-1} 所有这 n 项才能求出来，这好办，我们让函数 g 的自变量为 $\{a_1, a_2, \cdots, a_n\} = \{a_k\}_{k<n}$ 这个集合即可.

定理　递归定理 (recursion theorem)

设有一个集合 X，函数 $f : X^{<\omega} \to X$，那么有且仅有一个序列 $\{a_n\}_{n\in\mathbb{N}}$，使得 $a_n = f(\{a_k\}_{k<n})$.

证明　存在性很好证明，首先 $a_0 = f(\emptyset)$ 存在，$\{a_k\}_{k\leq 0}$ 存在. 假设 $\{a_k\}_{k\leq n}$ 存在，由于 $a_{n+1} = f(\{a_k\}_{k\leq n})$ 存在，所以 $\{a_k\}_{k\leq n+1}$，根据数学归纳原理的第一形式，任意 $n \in \mathbb{N}$，$\{a_k\}_{k\leq n}$ 都存在，把这些序列一起取个并集就是 $\{a_n\}_{n\in\mathbb{N}}$.

唯一性更好证明了，假设 $\{a_n\}_{n\in\mathbb{N}}, \{b_n\}_{n\in\mathbb{N}}$ 两个序列都符合条件，首先 $\{a_k\}_{k<0} = \{b_k\}_{k<0} = \emptyset$，假设 $\{a_k\}_{k<n} = \{b_k\}_{k<n}$，此时有 $a_n = f(\{a_k\}_{k<n}) = f(\{b_k\}_{k<n}) = b_n$，所以 $\{a_k\}_{k<n+1} = \{b_k\}_{k<n+1}$，根据数学归纳原理第一形式，对任意 n 均有 $\{a_k\}_{k<n} = \{b_k\}_{k<n}$，即对于任意 $n, a_n = b_n$，所以 $\{a_n\}_{n\in\mathbb{N}} = \{b_n\}_{n\in\mathbb{N}}$.
□

接下来我们把归纳和递归推广到涉及到超穷序数的情况. 从数学归纳原理的第一形式切入，如果我们想让所有序数都满足某个性质 ϕ，**仅仅声明“0 满足 ϕ”和“如果 α 满足则 α^+ 满足”足够吗？**最简单的反例：我们设命题 ϕ 为“小于 ω”，这时“$0 < \omega$”和“如果 $\alpha < \omega$ 那么 $\alpha^+ < \omega$”都是成立的，但是“$\omega < \omega$”就不对了，归纳假设还需要针对极限序数增加一条.

定理　超穷归纳 (transfinite induction)

- 超穷归纳定理，第一形式
  设有一个类 X，如果 $0 \in X$，并且任意序数 α 都满足“若 $\alpha \in X$ 则 $\alpha + 1 \in X$”，并且任意极限序数 α 都满足“若 $\beta < \alpha$ 则 $\alpha \in X$”，那么 $\mathbb{ON} \subseteq X$.

- 超穷归纳定理，第二形式
  设有一个类 X，如果任意序数 α 都满足“若 $\beta < \alpha$ 则 $\alpha \in X$”，那么 $\mathbb{N} \subseteq X$.

- 超穷归纳原理，第一形式
  设有命题 $\phi(\alpha)$，如果 $\phi(0)$ 成立，且序数 α 都满足“如果 $\phi(\alpha)$ 成立，那么 $\phi(n+1)$ 成立”，并且任意极限序数 α 都满足“若 $\phi(\beta)$ 对所有小于 α 的序数 β 成立则 $\phi(\alpha)$ 成立”，那么 $\phi(\alpha)$ 对任意序数 α 成立.

- 超穷归纳原理，第二形式
  设有命题 $\phi(\alpha)$，如果任意序数 α 都满足“若 $\phi(\beta)$ 对所有小于 α 的序数 β 成立则 $\phi(\alpha)$ 成立”，那么 $\phi(\alpha)$ 对任意序数 α 成立.

将数学归纳推广到超穷归纳以后，第一形式变长了，第二形式还像原来那样简洁，所以将来我们会经常使用第二形式.

定理　超穷递归 (transfinite recursion)

对于任意函数 $f : \mathbb{V} \to \mathbb{V}$，存在唯一的序列 $\{a_\xi\}_{\xi\in\mathbb{ON}}$ 满足 $a_\alpha = f(\{a_\xi\}_{\xi<\alpha})$.

3.3 序数运算法则

定义 序数运算法则

加法

$$\begin{aligned} \alpha + 0 &:= \alpha \\ \alpha + \beta^+ &:= (\alpha + \beta)^+ \\ \alpha + \beta &:= \sup_{\gamma < \beta}(\alpha + \gamma) && \text{其中}\beta\text{是极限序数} \end{aligned}$$

乘法

$$\begin{aligned} \alpha \cdot 0 &:= 0 \\ \alpha \cdot \beta^+ &:= \alpha \cdot \beta + \alpha \\ \alpha \cdot \beta &:= \sup_{\gamma < \beta}(\alpha \cdot \gamma) && \text{其中}\beta\text{是极限序数} \end{aligned}$$

乘方

$$\begin{aligned} \alpha^0 &:= \alpha \\ \alpha^{\beta^+} &:= \alpha^\beta \cdot \alpha \\ \alpha^\beta &:= \sup_{\gamma < \beta}(\alpha^\gamma) && \text{其中}\beta\text{是极限序数} \end{aligned}$$

有了序数，你就知道“如何数到无穷之后”了：

$$0, 1, 2, 3, \cdots, \omega, \omega + 1, \omega + 2, \cdots, \omega \cdot 2, \omega \cdot 2 + 1, \cdots, \omega^2, \cdots, \omega^\omega, \cdots$$

ω 及以后的序数都不是自然数，不能从 0 取有限次后继得到，所以将它们称为超穷序数(transfinite ordinal). 你也许会问有没有最大的序数，答案是没有，因为任何序数加一都会更大.

4 势与基数

4.1 等势

在朴素集合论中，集合的势通俗理解起来比较容易，但严格定义则非常抽象. 通俗理解，集合的势是集合的性质，表征了集合中元素的多少，是一个可以比较的对象. 问题来了，如何比较两个集合的元素谁多谁少?

一种方法是数个数，但这种方法有缺陷. 首先，“个数”是一个自然数，需要借助另一个集合来描述该集合，而且当集合的元素不是有限个的时候，这种方法就失效了. 也许你会提议使用序数，但如果使用序数来描述无穷集合的个数，就可能会出现因元素排序不同而有不同结果的现象，例如 $\mathbb{N}$ 有两种表示方法：

$$\mathbb{N} = \{n | n \in \mathbb{N}\} = \{2n, 2n + 1 | n \in \mathbb{N}\}$$

按照第一种表示方法，直觉上就会觉得 $\mathbb{N}$ 包含 ω 个元素，而第二种表示则将自然数分了奇偶，各自都有 ω 个，暗示 $\mathbb{N}$ 含有 $\omega \cdot 2$ 个元素. 集合是同一个集合，但因描述不同而出现两种结果，这是我们不能接受的.

因此应该使用另一种方法. 我们可以从幼儿园的游戏获得启发：有 5 棵树和 3 朵花，现在要比较数和花哪个更多. 然而小朋友不一定认识 3 和 5 这么“大”的数字，于是让一棵树与一朵花连线，连了线之后

的不能再连，最后所有的花都连上了线，但有的树没有连线，就这样得出树比较多的结论，这个过程没有引入自然数，却也比较了多少. 把“连线”翻译成“一一对应”（即双射），我们就得出了比较集合元素个数的方法：如果两个集合之间能建立双射，那么这两个集合的元素就是一样多的；如果一个集合能建立到另一个集合的单射，但无论怎么修改该单射的对应法则，都不能射满，那么后者的元素明显更多. 集合的势正是通过这种方法定义的.

定义　集合的势

集合 A 的势(cardinality) 记作 $\mathrm{card}(A)$，对于集合 A, B：
如果能建立单射 $f: A \to B$，则记 $\mathrm{card}(A) \le \mathrm{card}(B)$；
如果能建立双射 $f: A \to B$，则记 $\mathrm{card}(A) = \mathrm{card}(B)$；
如果能建立单射 $f: A \to B$，但该单射不是满射，则记 $\mathrm{card}(A) < \mathrm{card}(B)$.
如果存在自然数 n，使得 $\mathrm{card}(A) = \mathrm{card}(n)$，就可以简写为 $\mathrm{card}(A) = n$.

在朴素集合论语境下，集合的势是个很奇怪的概念，你可以比较两个集合的势谁大谁小，但没法讨论单个集合的“势”究竟是什么东西，只能通过符号的约定暂时规避. 这个问题将在公理集合论中引入选择公理之后得到解决. 一些教材把单个集合的势定义为与该集合等势所有集合的类，即

$$\mathrm{card}(A) = \{x | \mathrm{card}(x) = \mathrm{card}(A)\}$$

当然，这个定义非常不尽人意，因为没有道清“势”究竟是什么，而是把集合的势定义为一大捆相互等价的集合.

关于势的定义，有以下几个命题符合我们心目中关于“集合元素的个数”的直觉：

(1) 若 $\mathrm{card}(A) = \mathrm{card}(B), \mathrm{card}(B) = \mathrm{card}(C)$，则 $\mathrm{card}(A) = \mathrm{card}(C)$.

(2) 若 $A \subseteq B$，则 $\mathrm{card}(A) \le \mathrm{card}(B)$.

(3) 势的夹挤定理：若 $A \subseteq B \subseteq C$ 且 $\mathrm{card}(A) = \mathrm{card}(C)$，则 $\mathrm{card}(B) = \mathrm{card}(C)$.

证明　关于势的夹挤定理. 既然 $\mathrm{card}(A) = \mathrm{card}(C)$，就一定存在双射 $f: C \to A$. 这时构造两个序列 $\{A_n\}_{n\in\mathbb{N}}$ 和 $\{B_n\}_{n\in\mathbb{N}}$：

$$\begin{aligned} A_0 &= A, & A_{n+1} &= f[A_n] \\ B_0 &= B, & B_{n+1} &= f[B_n] \end{aligned}$$

这时令函数 $g: A \to B$

$$g(x) = \begin{cases} f(x), & \text{存在}n\text{使得}x \in A_n - B_n \\ x, & \text{其他} \end{cases}$$

容易验证 g 就是 $A \to B$ 的双射. □

(4) 康托尔-伯恩斯坦-施罗德定理(Cantor-Bernstein-Schröder theorem)：
若 $\mathrm{card}(X) \le \mathrm{card}(Y)$ 且 $\mathrm{card}(X) \ge \mathrm{card}(Y)$，则 $\mathrm{card}(X) = \mathrm{card}(Y)$，即势的大小关系具有反自反性.

证明 关于康托尔-伯恩斯坦-施罗德定理. 因为 $\mathrm{card}(X) \leq \mathrm{card}(Y)$，所以存在单射 $f: X \to Y$，即 $f[X] \subseteq Y$；因为 $\mathrm{card}(X) \geq \mathrm{card}(Y)$，所以存在单射 $f: Y \to X$，即 $g[Y] \subseteq X$，所以 $g[f[X]] \subseteq g[Y] \subseteq X$. 别忘了任何一个单射都是从定义域到值域的双射，又因为双射的复合也是双射，所以 $\mathrm{cardran}(g \circ f) = \mathrm{card}(X)$. 根据势的夹挤定理就得到了 $\mathrm{card}(Y) = \mathrm{card}(g[Y]) = \mathrm{card}(X)$. □

康托尔-伯恩斯坦-施罗德定理的证明是数学史上的典型，在大部分教材中证明过程包含了夹挤的部分，只不过这里拆开了. 这三位数学家以及戴德金都曾独立证明过该命题. 1887 年康托尔就发表了该定理，但是没有证明，1895 年又发表了一次，并提出将集合的势做全序排列之后便可立即得到该命题，但是他没能证明“基数可以全序排列”——因为这是选择公理的等价命题. 戴德金 1887 年就证明了该命题，但没有发表. 直到 1908 年，策梅洛才找到戴德金的证明，并于同一年提出了策梅洛公理集合论 (Z). 比戴德金晚一些，施罗德 1896 年证明出了该命题，而伯恩斯坦则是 1897 年，不过都是独立证明的. 该定理中夹挤部分的证明实在是太巧妙了，以至于恩德滕 (H.B.Enderton) 的集合论直接把该定理的可视化证明当作封面.

接下来证明几个有趣的命题，在有限集合上，“数个数”和“配对”的方法得出的结果都是符合直觉的，然而涉及到了无穷，情况就有些违反直觉.

(1) 对于无穷集合 A，有 $\mathrm{card}(A \times A) = \mathrm{card}(A)$.

证明 完整的证明需要用到超限归纳法，所以这里来证个弱化的版本：若 $\mathrm{card}(A) = \mathrm{card}(\mathbb{N})$，则 $\mathrm{card}(A \times A) = \mathrm{card}(A)$.

因为 $f(x) = (x, x): A \to (A \times A)$ 是单射，所以 $\mathrm{card}(A) \leq \mathrm{card}(A \times A)$，这是显然的. 接下来证明 $\mathrm{card}(A) \geq \mathrm{card}(A \times A)$. 由于 A 可数，所以它的元素可以被序列 $\{a_n\}_{n \in \mathbb{N}}$ 枚举，所以 $A \times A = \{(a_i, a_j) | i, j \in \mathbb{N}\}$. 这时候定义映射 $f(a_i, a_j) = a_{2^i 3^j}$，根据算数基本定理[2]，$2^i 3^j$ 对于不同的 i, j 值得到不同的自然数，因此 f 是一个单射，即 $\mathrm{card}(A) \geq \mathrm{card}(A \times A)$，最终得到 $\mathrm{card}(A) = \mathrm{card}(A \times A)$. □

根据这个结果，$\mathrm{card}(\mathbb{N}) = \mathrm{card}(\mathbb{Q})$，而 $\mathbb{Z} \subset \mathbb{Q}$，所以 $\mathrm{card}(\mathbb{N}) = \mathrm{card}(\mathbb{Z}) = \mathrm{card}(\mathbb{Q})$. 此外，$n$ 维空间中的有理点、奇数集、偶数集、素数集都可数；

(2) <u>康托尔定理</u>(Cantor's theorem)：$\mathrm{card}(\mathcal{P}(A)) > \mathrm{card}(A)$.

证明 $\mathrm{card}(\mathcal{P}(A)) \geq \mathrm{card}(A)$ 是显然的，因为 $f(x) = \{x\}$ 是 $A \to \mathcal{P}(A)$ 的单射，重点在于证明 $\mathrm{card}(\mathcal{P}(A)) \neq \mathrm{card}(A)$. 反证法，假设存在双射 $g: A \to \mathcal{P}(A)$，构造 A 的子集 $B = \{x \in A | x \notin g(x)\}$，$B \in \mathcal{P}(A)$，所以存在 $x \in A$ 使得 $f(x) = B$，然后矛盾就出来了：

$$x \in B \Longleftrightarrow x \notin g(x) \Longleftrightarrow x \notin B$$

因此该双射是不存在的，$\mathrm{card}(\mathcal{P}(A)) > \mathrm{card}(A)$. □

(3) $\mathrm{card}\left(2^A\right) = \mathrm{card}(\mathcal{P}(A))$.

证明 利用示性函数. 定义 $f(S) = \chi_S: \mathcal{P}(A) \to \{f | f: A \to \{0, 1\}\}$. 易验证 f 是双射. 因为从 $\mathcal{P}(A)$ 中挑出任意子集，得到的示性函数 χ_S 也不同. □

以上第 3 条的证明没有限定集合是有限集还是无限集，这告诉我们，即使是无限集合，其元素个数仍可以比较多少. $\mathrm{card}(\mathbb{N})$ 是我们遇到的第一个无穷集合，事实上在承认选择公理的条件下[3]它也是“包含元素最少的”无穷集合，也就是说如果 X 为无穷集合，就必有 $\mathrm{card}(X) \geq \mathrm{card}(\mathbb{N})$. 如果无穷集合元素的个

[2]算数基本定理 (fundamental theorem of arithmetic)：每个大于 1 的自然数可以唯一地分解为有限个素数的乘积.

[3]不承认选择公理的话就会出现一些非常混乱的集合，比如六种戴德金无穷，测度论将会遭受毁灭性的打击，数学分析也将难以进行.

数与自然数“一样多”，就会存在一个序列 $\{a_n\}_{n\in\mathbb{N}}$ 能把 X 中的元素一一枚举出来，这便是所谓“可数”或者“可列”.

定义 可数/可列与不可数/不可列

若存在自然数 n，使得 $\mathrm{card}(X)=n$，则称 X 是有穷的 (finite).
若 $\mathrm{card}(X)=\mathrm{card}(\mathbb{N})$，则称 X 是可数的(countable) 或可列的(enumerable).
若 $\mathrm{card}(X)>\mathrm{card}(\mathbb{N})$，那么 X 就是不可数的(uncountable) 或不可列的(denumerable).
有限集和可数无穷集合称至多可数集.

这里补充一个和字典-序列有关的命题，后面无论是数理逻辑还是数系的构造，我们都常常会用到这里的结论. 一个句子是若干个字或词按一定顺序拼接而成的，而这些字或词也不是凭空捏造的，而是有一部字典或词典收录所有可能出现的字或词. 从集合的角度来看，句子就是字典里的元素组成的序列，即记字典为 Σ，有限长的句子即序列 $\{a_i\}_{i<n}$，无限长（可数无穷长）的句子即序列 $\{a_i\}_{i\in\mathbb{N}}$，其中所有的 $a_i\in\Sigma$，全体序列的集合记为 Σ^*. 现在问题来了，$\mathrm{card}(\Sigma)$、序列的长度和 $\mathrm{card}(\Sigma^*)$ 有什么关系呢？直接上答案.

- 若 Σ 和序列长度都有限，那么 Σ^* 也是有限的.

证明 这就是个排列组合的问题，假设 $\mathrm{card}(\Sigma)=m$ 为序列长度为 n，这就相当于有 n 个位置、每个位置有 m 种可能的放法，一共有几种可能的情况？m^n 种. 这都没涉及到无穷. □

- 若 Σ 可数无穷、序列长度有限，那么 Σ^* 可数无穷.

证明 该证明需要用到算数基本定理和“素数的个数是无穷的”这两个结论. 既然 Σ 可数无穷，那么就存在双射 $f:\Sigma\to\mathbb{N}$ 将字典中的元素用自然数来编号. 假设序列长度是 n，那么我们构造一个映射

$$g:\{a_0,a_2,\cdots,a_{n-1}\}\mapsto 2^{f(a_0)+1}3^{f(a_1)+1}5^{f(a_2)+1}\cdots p_n^{f(a_{n-1})+1}$$

其中 p_n 表示第 n 各素数. 这个映射就是 $\Sigma\to\mathbb{N}$ 的单射，为什么？首先 g 是一个映射，因为各个 $f(a_i)$ 是唯一确定的自然数。$2^{f(a_0)+1}3^{f(a_1)+1}5^{f(a_2)+1}\cdots p_n^{f(a_{n-1})+1}$ 是有限项相乘的形式，所以也是唯一确定的自然数. 其次 g 是单射，因为根据算数基本定理，每个自然数 c 都能唯一分解成 $2^{c_1}3^{c_2}5^{c_3}\cdots p_n^{c_n}$ 的形式，这就隐含了一个 $c\mapsto\{c_1,c_2,\cdots,c_n\}$ 的双射，现在我们反过来，就得到了 g 是单射. 至于为什么要给每个指数都加 1，是避免 f 将字典元素映射到 0 时会造成的某些问题，无伤大雅.

既然 g 是 $\Sigma^*\to\mathbb{N}$ 的单射，那么 $\mathrm{card}(\Sigma^*)\le\mathrm{card}(\mathbb{N})$. 要证明 $\mathrm{card}(\Sigma^*)\ge\mathrm{card}(\mathbb{N})$ 是很平凡的事情，因为 f 的逆映射把自然数映射到了 Σ 中，而字典中的字可以看作长度为 1 的序列，因此 $n\mapsto\{f^{-1}(n)\}$ 就是单射. 根据势的夹挤定理，$\mathrm{card}(\Sigma^*)=\mathrm{card}(\mathbb{N})$，可数无穷. □

- 若 Σ 有限且元素多于 1 个、序列长度可数无穷，那么 Σ^* 不可数无穷.

证明 反证法. 假设 Σ^* 可数，那么就会有一个序列 $\{S_i\}_{i\in\mathbb{N}}$ 把它列出来，现在构造一个序列 $S^*=\{b_0,b_1,b_2,\cdots\}$，然后随便选取两个字典元素 c_1,c_2，当 S_i 的第 i 项为 c_1 时 $b_i=c_2$，其他时候 $b_i=c_1$，那么 S^* 的第 i 项就与 S_i 的第 i 项不同了，所以 S^* 不等于任意一个 S_i，但很明显 $S^*\in\Sigma^*$，矛盾了，所以 Σ^* 不可数无穷. □

这个证明有点抽象，举个具体的例子. 设 $\Sigma=\{c_1,c_2,c_3,\cdots,c_{114514}\}$，如果 Σ^* 可数，那么就会有 $\{S_i\}_{i\in\mathbb{N}}$ 把它列出来：

$$
\begin{array}{lllllll}
S_0=\{ & c_3, & c_{43590}, & c_{24}, & c_{2346}, & c_{100000}, & \cdots\} \\
S_1=\{ & c_{78}, & c_1, & c_{34}, & c_{786}, & c_{2313}, & \cdots\} \\
S_2=\{ & c_{475}, & c_{81}, & c_{14347}, & c_{468}, & c_{7186}, & \cdots\} \\
S_3=\{ & c_{349}, & c_{4709}, & c_{54136}, & c_{2346}, & c_{1348}, & \cdots\} \\
S_4=\{ & c_{98478}, & c_{8890}, & c_{92515}, & c_{14847}, & c_1, & \cdots\} \\
\cdots \\
S^*=\{ & c_1, & c_2, & c_1, & c_1, & c_2, & \cdots\}
\end{array}
$$

很明显 S^* 的第 i 项就与 S_i 的第 i 项不同，这就说明 S^* 无法被枚举出来，从而导出矛盾. 这种证明方法称为对角线证法(diagonal argument)，最早被康托尔用来证明实数集不可数，他的证法是这样的：已知每个无限小数都是实数，假设无限小数可列，那么数列 $\{a_n\}_{n\in\mathbb{N}}$ 可以把区间 $(0,1)$ 内的小数列出来，然后构造第 i 位和 a_i 的第 i 位不同的小数 a^*（比如选取 4 和 7 来判定）.

$$
\begin{aligned}
a_0 &= 0.21845\cdots \\
a_1 &= 0.98724\cdots \\
a_2 &= 0.83166\cdots \\
a_3 &= 0.39574\cdots \\
a_4 &= 0.58363\cdots \\
\cdots \\
a^* &= 0.77747\cdots
\end{aligned}
$$

这就充分说明了 a^* 没有被枚举出来，因此实数区间 $(0,1)$ 上的实数不可数. 康托尔 1891 年发表了这个证明. 当然，他在 1879 年的文章 *Ueber eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen* 中就已经用区间套的方法证明了连续统（实数区间）是不可数的，这篇文章尽管不到四页半，却被视为集合论的开山之作，它的发表标志着集合论的诞生.

- 若 Σ 和序列长度都可数无穷，那么 Σ 不可数无穷.
  相比于上一种情况，加大字典只会让序列只增不减，所以当然不可数无穷啦.

根据上面的结果，我们自然会觉得如果 Σ 和序列长度都不可数无穷时 Σ^* 也不可数无穷. 但是不可数无穷之间也是相对大小的，尽管 $\mathbb{R}$ 和 $\mathcal{P}(\mathbb{R})$ 都是不可数无穷，但是康托尔定理告诉我们后者比前者势大. 所以 $\mathrm{card}(\Sigma^*)$ 比 $\mathrm{card}(\Sigma^*)$ 大吗？如果大的话，大多少？我们现在无法回答，因为知识储备还太少——连 card 是什么都不知道. 这将在后面引入选择公理的时候详细讨论.

4.2 基数运算法则

以上我们定义了集合的势 $\mathrm{card}(A)$，规定了集合的势什么情况下相等、大于、小于另一个集合的势，但是势的这种定义有些令人不满，因为我们希望将势定义为一个简洁并且单独的对象，而不是而不是一大捆相互等价的集合. 因此我们应该提出自然数往无穷方向扩展的方式，使之能适应集合的势的定义.

定义　基数（依赖选择公理）

对于序数 α，若任意小于 α 的序数 β，都满足 $\mathrm{card}(\beta) < \mathrm{card}(\alpha)$，则称序数 α 是一个基数(cardinal number).

集合 A 的势 $\mathrm{card}(A)$ 定义为与其等势的基数.

这下我们给基数找到了简洁的定义，容易发现，对于两个基数 κ 和 λ，$\mathrm{card}(\kappa) < \mathrm{card}(\lambda), \kappa < \lambda, \kappa \in \lambda, \kappa \subset \lambda$ 这四个命题是等价的. 然而，经过本章开头的论证，我们不应该仅仅用序数来定义基数，还要解决“重新排序后元素个数不等”的问题. 因此应该给基数制定一套运算法则.

定义　基数运算法则

后继

$$\kappa^+ := \min\{\lambda | \lambda > \kappa\}$$

加法

$$\kappa + \lambda := \mathrm{card}(A \cup B), \quad 其中\mathrm{card}(A) = \kappa, \mathrm{card}(B) = \lambda, A \cap B = \emptyset$$

乘法

$$\kappa \cdot \lambda := \mathrm{card}(A \times B), \quad 其中\mathrm{card}(A) = \kappa, \mathrm{card}(B) = \lambda$$

乘方

$$\kappa^\lambda := \mathrm{card}(A^B), \quad 其中\mathrm{card}(A) = \kappa, \mathrm{card}(B) = \lambda$$

基数是一类特殊的序数，它们既可以参与序数运算，也可以参与基数运算，具体参加哪一套需要在记号上加以区分，以上在定义基数运算法则时，把基数用 $\kappa, \lambda, \mu, \nu, \cdots$ 等靠后的希腊字母表示，而强调序数时一般使用 $\alpha, \beta, \gamma, \delta, \cdots$ 等靠前的希腊字母. 很容易验证，当限定在有限集合的时候，序数、基数的运算结果是相同的，都非常符合直觉，比如 2×3，无论套用序数运算法则还是基数运算法则，甚至是将来使用皮亚诺公理定义的自然数运算法则，抑或是更往后的整数、有理数、实数、复数运算法则，它都等于 6. 而序数与基数的差异出现在无穷之后，因此我们使用希伯来字母来强调它的基数身份.

定义　阿列夫数和贝斯数

序列 $\aleph_\alpha : \mathbb{ON} \to \mathbb{ON}$ 和 $\beth_\alpha : \mathbb{ON} \to \mathbb{ON}$ 定义如下：

$$\begin{array}{lllllll}
\aleph_0 & = & \mathrm{card}(\mathbb{N}) & \beth_0 & = & \mathrm{card}(\mathbb{N}) & \\
\aleph_{\alpha+1} & = & \aleph_\alpha^+ & \beth_{\alpha+1} & = & \mathrm{card}(\mathcal{P}(\beth_\alpha)) & \\
\aleph_\alpha & = & \bigcup\limits_{\beta<\alpha} \aleph_\beta & \beth_\alpha & = & \bigcup\limits_{\beta<\alpha} \beth_\beta & \alpha是极限序数
\end{array}$$

根据康托尔定理，$\beth_\alpha \geq \aleph_\alpha$，但是等号是否成立呢？取幂集得到的基数是否正好是下一个基数？康托尔穷尽一生都没能得出答案，但他认为是相等的，这个断言被称为广义连续统假设(generalized continuum hypothesis, GCH)，而连续统假设(continuum hypothesis, CH) 仅断言 $\beth_1 = \aleph_1$. 这个问题到目前为止都没有答案，接下来在公理集合论中还会详细讨论.

5 朴素集合论的局限性

朴素集合论的概括原则断言了符合 $\{x|\varphi(x)\}$ 形式的对象都是集合，如果 x 满足命题 $\varphi(x)$，那么 x 就属于这个集合，否则就不属于. 这个原则在很长一段时间内都被认为是正确的，直到 1901 年罗素构造出了如下集合：

$$A = \{x|x \notin x\}$$

试问，$A \in A$ 吗？如果 $A \in A$，那么 A 就需要满足命题 $x \notin x$，即 $A \notin A$；如果 $A \notin A$，那么 A 就满足了命题，因此 $A \in A$. 无论怎么选择，都会得到矛盾的结果，这个矛盾称为罗素悖论(Russel's paradox).

布劳威尔 (L. E. J Brouwer) 和外尔 (H. Weyl) 认为这是对集合论的致命一击，但哥德尔没有把它当成严重的问题，他认为问题出在数学的边界上，而且是靠近哲学的边界上，并且通过公理化的方法，可以将这个问题完美解决. 而罗素本人则通过赋予集合以“类型”来否定 $x \in x$ 这个命题的合法性，由此创立了类型论(type theory). 现在一般认为，公理化的集合论规避了罗素悖论，至于这种规避是否解决了问题，则依据哲学立场的不同有不同的看法.

6 数学第二步，数理逻辑

~~数理逻辑“以数学的方式研究数学”(by 以色列数学家 Shalah)，这句话有两个层面的意思，一是“使用数学”，即整个学科的研究目标就是指出哪些命题是真的、哪些不是真的、哪些无法判断，这种真有两个层面，第一个层面的“真比如“$1+1=2$”是真的，不依赖于现实世界中“一个苹果和另一个苹果放在一起就得到两个苹果”这样的事实，而任作为反例，网上关于 $0.999\cdots$ 是否等于 1 的争论，甚至用现实世界的例子来解释等于或不等于都不是数理逻辑的研究作为整个数理逻辑部分的开头，首先需要明确以下两对概念.~~

7 命题逻辑

朴素集合论的概括原则断言形如 $\{x|\phi\}$ 的东西是一个集合，没有规定额外的条件，由此就导致了罗素悖论. 为了克服朴素集合论的这个缺陷，数学家对“什么是集合”的问题做了深入的研究，第一步就是规定 ϕ 应该是什么形式，这就是形式逻辑所研究的内容.

自然语言中的命题(proposition) 我们以前学过，就是能写成“如果⋯⋯那么⋯⋯”形式的陈述句. 然而命题最重要的地方不在于它的句子形式，而是它能够判断真假，形式地说，即存在一个函数

$$v : \text{命题的集合} \to \{\text{真}, \text{假}\}.$$

接下来我们不必去理会一个命题本身是什么样的，认为它是良好的命题（即能判断真假），然后像字母代替数一样，也用一个字母（比如 A）表示命题，此时对其判断真假的过程就可以记为 $v(A)$，得到的结果只有两种：真 (T) 和假 (F). 给定了命题的真假，把它用连接词连接成长命题，这个长命题和给定的命题有什么样的真假关系，这就是命题逻辑语义的研究内容.

7.1 语义和真假赋值

定义 命题逻辑的语言要素

最简单的命题逻辑的语言有两种符号.

- 命题符号(proposition)，用 $A_0, A_1, A_2, \cdots$ 等符号来表示，命题符号组成的集合记作 $\mathcal{L}_0^0$.
- 逻辑连接词(logic connective)，有以下两个：
  否定连接词(negation)，写作“$\neg$”，读作“非”；
  条件连接词(condition)，又称实质蕴含连接词(material implication)，写作“$\rightarrow$”，读作“蕴含”.
- 括号，让连接词的作用范围表达得更清楚，这里只使用小括号.

你也许会感到奇怪，以前学过的逻辑连接词是“$\vee$”、“$\wedge$”和“$\neg$”，为什么现在只有两个，还出现了一个“$\rightarrow$”呢？当然使用这三个也是可以的，但是毕竟有三个连接词啊，比两个要复杂一些，并且我们将会证明，定义中所使用的两个连接词和我们学过的三个连接词，表达能力是一样的，也就是说，我们学过的三个连接词所构造出来的复杂句子，使用这两个连接词一样能构造出来，反之亦然. 好了，有了命题符号作为原材料、连接词作为工具，接下来的事情自然就是构建句子了. 就像起房子一样，我们需要正确地安排好两种符号的书写顺序，比如“$((\neg(\neg A_1)) \rightarrow A_2) \rightarrow ((\neg(\neg A_1)) \rightarrow (\neg(\neg A_2)))$”是正确的写法，而“$(A_1)) \rightarrow) \rightarrow ((\neg A_2) \rightarrow (\neg A_1)) \rightarrow (\neg($狗$\neg\neg A_2((\neg) \rightarrow)$”就是错误的写法. 正确的写法写出来的句子称为合式公式(well-formed formula, WFF)，现在我们就来规定什么是合式公式.

定义 命题逻辑的合式公式

命题逻辑合式公式一般使用小写希腊字母 φ, ψ 等表示，有以下三种组成方式[a]：

- 命题符号是合式公式.
- 若 φ 是合式公式，则 $(\neg\varphi)$ 是合式公式.
- 若 φ, ψ 都是是合式公式，则 $(\varphi \rightarrow \psi)$ 是合式公式.

由 $\mathcal{L}_0^0$ 中的命题符号组成的所有可能的合式公式的集合记为 $\mathcal{L}_0$，称为命题逻辑语言.

[a] 严格来说我们需要首先用序列定义字符串及其连接运算，再定义什么是合式公式，但是那样会显得太晦涩，这里姑且略去那些繁文缛节.

这个定义采用递归的方式规定什么是合式公式，我们在朴素集合论里已经证明了递归定义的合法性. 合法是合法，但你有没有觉得太麻烦了呢，就像上面那个蓝色的合式公式一样，括号太多了. 之前说过，括号是为了避免歧义，比如“$\neg A_1 \rightarrow A_2$”的意思是“$(\neg A_1) \rightarrow A_2$”还是“$\neg(A_1 \rightarrow A_2)$”呢？但我们反过来考虑，为什么算式 $3 \times 5 + 6$ 没有歧义呢？是因为我们规定了乘法的优先级要比加法高. 因此我们也给两个连接词规定优先级：**“$\neg$”比“$\rightarrow$”更优先，也比接下来介绍的所有二元连接词更优先.** 这样“$\neg A_1 \rightarrow A_2$”就表示“$(\neg A_1) \rightarrow A_2$”，上面那个蓝色的公式就可以写作 $(\neg\neg A_1 \rightarrow A_2) \rightarrow (\neg\neg A_1 \rightarrow \neg\neg A_2)$，姑且不论基于真假的化简，至少这样看起来顺眼多了.

以上定义出的合式公式，有以下几个性质：

- 对于一个合式公式 φ，在其中任意一个命题符号之前的符号序列都不是合式公式.
- 唯一可读性定理：对于一个合式公式 φ，要么 φ 是逻辑符号，要么 φ 形如 $\neg\psi$ 且 ψ 唯一，要么 φ 形如 $\psi_1 \rightarrow \psi_2$ 且 ψ_1 和 ψ_2 唯一，三者必满足其一且只满足其一.

- 合式公式归纳定理：令 $P(\varphi)$ 是关于合式公式的性质，若对于每个命题符号 A_i，$P(A_i)$ 都成立；且如果 $P(\varphi)$ 和 $P(\psi)$ 都成立 $P(\neg\varphi)$ 和 $P(\varphi \to \psi)$ 就一定成立，那么 $P(\varphi)$ 对全体合式公式 φ 都成立.

命题最重要的一点是可以判断真假. 现在有了函数 $v : \Sigma \to \{T, F\}$，对每一个命题符号判定真假，却不能给符号连接而成的合式公式判断真假，所以现在就来补全它.

定义　合式公式的真假赋值/真值指派

设 $v : \Sigma \to \{T, F\}$，定义 $\bar{v} : \Sigma^* \to \{T, F\}$ 为扩张的真假赋值函数：

- 若 φ 是逻辑符号，则 $\bar{v}(\varphi) = v(\varphi)$.
- 若 φ 形如 $\neg\psi$，且 $\bar{v}(\psi)$ 有定义，则 $\bar{v}(\varphi) = \begin{cases} T, & \bar{v}(\psi) = F \\ F, & \bar{v}(\psi) = T \end{cases}$.
- 若 φ 形如 $\psi_1 \to \psi_2$，则 $\bar{v}(\varphi) = \begin{cases} F, & \bar{v}(\psi_1) = T, \bar{v}(\psi_2) = F \\ T, & \text{其他} \end{cases}$

利用数学归纳可以证明，$\bar{v}$ 是合法的函数，即扩张的真假赋值是唯一的.

难以理解的地方来了，根据定义，只要 A_1 是假命题，无论 A_2 是真是假，"$A_1 \to A_2$" 都是真命题. 由于我们定义蕴含连接词的时候直观上将其理解为"推出"，这就意味着假命题可以推出任意命题，并且你将会看到，后面要讨论的语义后承和句法后承都有"假推任意"的现象. 这种现象称为爆炸原理(ex falso quodlibet)，对很多人来说难以置信. 我们可以这么理解，"$A_1 \to A_2$" 声明了一个承诺：当 A_1 成立的时候 A_2 也要成立，而句子的真值取决于这个承诺有没有被打破. 如果 A_1 为真 A_2 为假，承诺被打破，"$A_1 \to A_2$" 为假，但如果 A_1 为假，连承诺的前提都没达成，那这个承诺就不适用，而没有被打破，句子仍然为真，但这是"无意义的真"，即"虚真"(vacuously true).

现在来说说我们熟悉的另外三个连接词.

- 合取连接词(conjunction)，写作"$\wedge$"，读作"且"，$A_1 \vee A_2$ 等价于 $\neg(A_1 \to \neg A_2)$.
- 析取连接词(disjunction)，写作"$\vee$"，读作"或"，$A_1 \vee A_2$ 等价于 $\neg A_1 \to A_2$.
- 双条件连接词(bicondition)，写作"$\leftrightarrow$"，读作"当且仅当"，$A_1 \vee A_2$ 等价于 $(A_1 \to A_2) \wedge (A_2 \to A_1)$.

这个时候就可以祭出真值表了. 需要注意，只有当 $v(A_i)$ 确定以后，$\bar{v}(\varphi)$ 才有讨论的意义，命题逻辑并不关注某个 A_i 为什么为真，另一个 A_j 为什么为假，而是关注将多个命题连接成的复杂句子的真假. "不关注原子事实的真假"这一思想今后还会在其他地方体现.

φ	ψ	$\neg\varphi$	$\varphi \to \psi$	$\varphi \vee \psi$	$\varphi \wedge \psi$	$\varphi \leftrightarrow \psi$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	T	T	F	F
F	F	T	T	F	F	T

Table 1: 五个连接词的真值表

7.2 命题的语义

定义 语义后承

设真假赋值函数 $\bar{v}$ 是 v 的扩张，若 $\bar{v}(\varphi)=T$，称 v满足(satisfies) 合式公式 φ.
若 $\Gamma\subseteq\mathcal{L}_0$ 且 v 满足 Γ 中的每一个合式公式，称 v 满足集合 Γ，也称 Γ 可被 v 满足.
如果 v 满足集合 Γ，就一定满足合式公式 φ，则称集合 Γ重言蕴含(tautologically implies) 合式公式 φ，也称 φ 为公式集 Γ 的语义后承(sematics consequence)，记作 $\Gamma\vDash\varphi$.
如果 Γ 只有一个公式 ψ，那么 $\{\psi\}\vDash\varphi$ 可以简写为 $\psi\vDash\varphi$. 如果 $\varphi\vDash\psi$ 且 $\psi\vDash\varphi$，那么称 φ 和 ψ重言等价(tautologically equivalent).

比如，在侦探小说里，有三个嫌疑人 X,Y,Z，而现在我们得到了三条线索：1. X,Y 之一是犯人；2.X,Z 是犯人；3. 不可能 Y 或 Z 是犯人. 现在我们将嫌疑人的编号作为其犯罪的命题符号，即 X 表示“X 是犯人”这个命题，三条线索就可以写成 $(X\vee Y),(X\vee Z),\neg(Y\vee Z)$，令 Γ 为这三条公式的集合，然后可以列出一张表：此时只有表中红色那一行所代表的

X	Y	Z	$(X\vee Y)$	$(X\vee Z)$	$\neg(Y\vee Z)$
T	T	T	T	T	F
T	T	F	T	T	F
T	F	T	T	T	F
T	F	F	T	T	T
F	T	T	T	T	F
F	T	F	T	F	F
F	F	T	F	T	T
F	F	F	F	F	T

赋值函数才能满足 Γ，而此时 X 被满足，所以 $\Gamma\vDash X$.

现在来思考一个特殊情况：$\emptyset\vDash\varphi$. 用自然语言翻译：“如果对于每个赋值函数 v，都满足 $\emptyset$ 中的公式，那么 v 满足 φ”. 这又是一个“如果……那么……”的命题，首先 v 是否满足空集中的“每一个公式”呢？弄不清楚就把这个句子展开：“如果某个公式属于空集，那么该公式被 v 满足”. “某个公式属于空集”本身就是假命题了，所以这个句子为真，即 v 满足空集中的“每一个公式”. 所以只要“任意一个 v 都满足 φ”即可. 这意味着，无论命题符号是真是假，φ 总是一个真命题.

定义 重言式和谬论

如果 $\emptyset\vDash\varphi$，则称 φ 是重言式(tautology)，可以简写为 $\vDash\varphi$.
如果任何真假赋值函数都不能满足命题 φ，就称 φ 是一个谬论(contradiction).

接下来给出几个常用的重言式，这些重言式作为非常经典的逻辑规律，在将来还会经常出现.

- 或和且的交换律、结合律、分配律

$$(A\vee B)\leftrightarrow(B\vee A)\qquad (A\wedge B)\leftrightarrow(B\wedge A)$$
$$((A\vee B)\vee C)\leftrightarrow(A\vee(B\vee C))\qquad ((A\wedge B)\wedge C)\leftrightarrow(A\wedge(B\wedge C))$$
$$(A\vee(B\wedge C))\leftrightarrow((A\vee B)\wedge(A\vee C))\qquad (A\wedge(B\vee C))\leftrightarrow((A\wedge B)\vee(A\wedge C))$$

- 德摩根律 (De Morgan law)

$$\neg(A\wedge B)\leftrightarrow(\neg A\vee\neg B)\qquad \neg(A\vee B)\leftrightarrow(\neg A\wedge\neg B)$$

- 双重否定 (double negation)：$\neg\neg A \leftrightarrow A$
- 排中律 (exclude middle)：$A \wedge \neg A$
- 矛盾律 (contradiction)：$\neg(A \wedge \neg A)$
- 导出律 (exportation)：$((A \wedge B) \to C) \leftrightarrow (A \to (B \to C))$
- 逆否命题等价于原命题/换质位法 (contraposition)：$(A \to B) \leftrightarrow (\neg B \to \neg A)$

7.3 逻辑连接词的高级玩法

这一节讲的还是命题的语义. 命题符号可以被真假赋值函数映射为某个真值，用逻辑连接词把命题符号连接成合式公式以后可以被扩展的真假赋值函数映射为某个真值，例如 $v(A) = T, \bar{v}(\neg A) = F$，$\neg A$ 的真值取决于 A 的真值. 如果我们从函数的角度来看问题，就可以把逻辑连接词看作一个把真值映射为真值的函数，例如 “$\neg$” 可以看成一个 $F \mapsto T, T \mapsto F$ 的函数，而 “$\to$” 可以看作一个二元函数: $(T,T) \mapsto T, (T,F) \mapsto F, (F,T) \mapsto T, (F,F) \mapsto T$.

推广这种想法，逻辑函数是一个 n 元函数 $B(X_1, X_2, \cdots, X_n)$，接受 n 个真值，并映射为一个真值. 如果两个逻辑函数映射的规则是相同的，那就算做同一个函数. 如果真值只有 T, F 两个，运用一点点组合学的知识，我们可以很快算出一元逻辑函数有 4 个，二元逻辑函数有 16 个，三元逻辑函数有 256 个，n 元逻辑函数有 2^{2^n} 个（n 个位置，每个位置有两种可能的真值，一共有 2^n 种排列情况，而对于每种排列情况，都有两个可能的输出，因此有 2^{2^n} 个可能的输出）.

对于某个涉及了 n 个命题符号的合式公式 α，其真值由这些命题符号的真值唯一决定，这就确定了一个逻辑函数 B_α. 现在问题就来了，反过来是否成立，即一个逻辑函数 $B_\alpha(X_1, X_2, \cdots, X_n)$ 是否能被某个涉及 n 个命题符号 $A_1, A_2, \cdots, A_n$ 的合式公式 α 表达: 即当 $v(A_i) = X_i$ 时，$\bar{v}(\alpha) = B_\alpha(X_1, X_2, \cdots, X_n)$?

例如给定了一个逻辑函数 $B_\alpha(X_1, X_2, X_3)$ 如下面左边的表格，那么说 $\alpha : A_1 \to \neg(A_2 \to \neg A_3)$ 表达了这个逻辑函数，因为当 A_1, A_2, A_3 的真值分别为 X_1, X_2, X_3 时，α 的真值为 $B_\alpha(X_1, X_2, X_3)$，如下面右边的表格所示.

X_1	X_2	X_3	$B_\alpha(X_1,X_2,X_3)$
F	F	F	T
F	F	T	T
F	T	F	T
F	T	T	T
T	F	F	F
T	F	T	F
T	T	F	F
T	T	T	T

$v(A_1)$	$v(A_2)$	$v(A_3)$	$\bar{v}(\alpha)$
F	F	F	T
F	F	T	T
F	T	F	T
F	T	T	T
T	F	F	F
T	F	T	F
T	T	F	F
T	T	T	T

你一定马上意识到了（也许没意识到）逻辑函数是否能确定一个合式公式和允许选取的逻辑连接词有关. 在上面的例子中如果我们不允许使用 “$\neg$”，那就不存在满足条件的合式公式. 对于逻辑连接词的集合 C，如果任意 n 元的逻辑函数都能用只涉及 n 的命题符号的、由 C 中的连接词连接而成的合式公式表达，就称 C 为功能完全的(complete). 接下来我们证明 $\{\neg, \vee, \wedge\}$ 是功能完全的.

证明 如果 $B_\alpha(X_1, X_2, \cdots, X_n)$ 恒等于 F，那么取 $\alpha : A_1 \wedge \neg A_1$ 就好了.
如果 $B_\alpha(X_1, X_2, \cdots, \cdots, X_n)$ 不恒等于 F，那就把映射至 T 的真值组合全列举出来，假设有 m 个：

$$\begin{aligned}
\bar{X}_1 &= (X_{11}, X_{12}, \cdots, X_{1n}) \\
\bar{X}_2 &= (X_{21}, X_{22}, \cdots, X_{2n}) \\
&\cdots \\
\bar{X}_m &= (X_{m1}, X_{m2}, \cdots, X_{mn})
\end{aligned}$$

此时设

$$B_{ij} = \begin{cases} A_j, & X_{ij} = T \\ \neg A_j, & X_{ij} = F \end{cases}$$

这样 B_{ij} 就恒真了，这时再设

$$C_i = B_{i1} \wedge B_{i2} \wedge \cdots \wedge B_{in}$$

因为该真假赋值满足每个 B_{ij}，所以也满足 C_i. 同时也只有唯一的真假赋值满足 C_i，不信你试试看，哪怕只改变一个命题符号的真值，某个 B_{ij} 就变假了，就不满足 C_i 了. 此时令

$$\alpha = C_1 \vee C_2 \vee \cdots \vee C_m$$

当然 v 也会满足 α. 此时求出的 α 就表达了函数 $B_\alpha(X_1, X_2, \cdots, X_n)$. 另一方面，因为每个 C_i 只有一个真假赋值能满足，所以仅有 m 个真假赋值能满足 α，也就是说，任意调整 α 涉及到的每个命题符号的真假，能满足 α 的情况只有 m 种，而这 m 种就是 X_i 所对应的真假赋值. 于是公式 α 就表达出了逻辑函数 B_α. 证毕.

该证明可能有些抽象了，这里用上面的例子来验证. $B_\alpha(X_1, X_2, X_3)$ 一种有 8 种输入，其中有 5 种能够映射至 T，按照证明中说的去做，写出映射至 T 的真值组合 $\bar{X}_i$ 以及组合好的命题命题 C_i.

$$\begin{aligned}
\bar{X}_1 &= (F, F, F) & C_1 &= \neg A_1 \wedge \neg A_2 \wedge \neg A_3 \\
\bar{X}_2 &= (F, F, T) & C_2 &= \neg A_1 \wedge \neg A_2 \wedge A_3 \\
\bar{X}_3 &= (F, T, F) & C_3 &= \neg A_1 \wedge A_2 \wedge \neg A_3 \\
\bar{X}_4 &= (F, T, T) & C_4 &= \neg A_1 \wedge A_2 \wedge A_3 \\
\bar{X}_5 &= (T, T, T) & C_5 &= A_1 \wedge A_2 \wedge A_3
\end{aligned}$$

这时构造出来的命题 α 就是

$$(\neg A_1 \wedge \neg A_2 \wedge \neg A_3) \vee (\neg A_1 \wedge \neg A_2 \wedge A_3) \vee (\neg A_1 \wedge A_2 \wedge \neg A_3) \vee (\neg A_1 \wedge A_2 \wedge A_3) \vee (A_1 \wedge A_2 \wedge A_3)$$

想让真假赋值满足公式 C_1，那就必须有 $v(A_1) = v(A_2) = v(A_3) = F$；想满足公式 C_2，那就必须有 $v(A_1) = F, v(A_2) = F, v(A_3) = T$；同理，想满足 C_3, C_4, C_5，都有不同且唯一的真假赋值. 来看最终的 $\alpha = C_1 \vee C_2 \vee C_3 \vee C_4 \vee C_5$，由于某个真假赋值至多只能满足一个 C_i，因此共有 5 个真假赋值可以满足 α，而这 5 个真假赋值给 (A_1, A_2, A_3) 指派的真值恰好分别是 5 个 $\bar{X}_i$ 列出的真值，所以 α 表达出了 $B_\alpha(X_1, X_2, X_3)$.

$\square$

好，现在证明了 $\{\neg, \vee, \wedge\}$ 是功能完全的，以它为桥梁，我们可以证明其他某些连接词的组合也是功能完全的，当且仅当那些连接词和 $\{\neg, \vee, \wedge\}$ 能互相表示.

- $\{\neg, \to\}$ 是功能完全的，因为

$$A_1 \to A_2 \text{重言等价于} \neg A_1 \vee A_2 \qquad A_1 \vee A_2 \text{重言等价于} \neg A_1 \to A_2$$
$$A_1 \wedge A_2 \text{重言等价于} \neg(A_1 \to \neg A_2)$$

- $\{\neg, \wedge\}$ 和 $\{\neg, \vee\}$ 都是功能完全的.

- 在数字电路设计中常用到“与非”(NAND) 和“或非”(NOR) 两种逻辑连接词，A_1 NAND A_2 与 $A_1 \wedge A_2$ 的真假相反，A_1 NOR A_2 与 $A_1 \vee A_2$ 的真假相反，{NAND} 和 {NOR} 都是功能完全的，因为

$$\neg A \text{重言等价于} A \text{ NAND } A \qquad A_1 \text{ NAND } A_2 \text{重言等价于} \neg A_1 \wedge A_2$$
$$A_1 \wedge A_2 \text{重言等价于} (A_1 \text{ NAND } A_2) \text{ NAND } (A_1 \text{ NAND } A_2)$$
$$A_1 \vee A_2 \text{重言等价于} (A_1 \text{ NAND } A_1) \text{ NAND } (A_2 \text{ NAND } A_2)$$

$$\neg A \text{重言等价于} A \text{ NOR } A \qquad A_1 \text{ NOR } A_2 \text{重言等价于} \neg A_1 \vee A_2$$
$$A_1 \wedge A_2 \text{重言等价于} (A_1 \text{ NOR } A_2) \text{ NOR } (A_1 \text{ NOR } A_2)$$
$$A_1 \vee A_2 \text{重言等价于} (A_1 \text{ NOR } A_1) \text{ NOR } (A_2 \text{ NOR } A_2)$$

- $\{\wedge, \to\}$ 不是功能完全的.

当我们充分刻画了逻辑连接词的功能完全性和冗余性之后，我们不妨放开束缚着的手脚，接下来大胆地使用冗余的连接词简化公式的书写，比如 $\vee, \wedge, \leftrightarrow$ 等.

7.4 命题的语法

本节主要讨论“证明”这个概念. 你可能觉得，前面朴素集合论已经证明了许多命题，意味着即使不讨论这个概念，我们一样能证明许多命题. 确实，即使不用严格规定“证明”是什么，我们一样能使用它，数学界几千年来也一样能发展得好好的. 然而，在朴素集合论中，当我们用概括原则规定“什么是集合”以后，罗素悖论就逼出了朴素集合论的局限性. 同样地，我们希望通过严格规定“什么是证明”，从而逼出“证明”这个工具的局限性，从而发现某些“不可证”的东西，这就是数理逻辑这门学科存在的意义之一——研究手段的局限性.

从一个例子出发. 我们小学就学过三角形内角之和为 180°，后来使用内错角相等的结论证明. 把证明的过程复现一下，首先我们设三角形三个角分别为 α, β, γ，过 α 的顶点作其对边的平行线，因为内错角相等，所以 $\beta = \beta_1, \gamma = \gamma_1$，而此时 $\alpha + \beta_1 + \gamma_1 = 180°$，所以三个角之和为 180°.

从这个证明的过程中能发现什么？这个证明过程出现了一个又一个命题，用“因为……所以……”这类关联词连接，可以形象地说，证明过程就是从假设到结论的逻辑链条. 不过这个链条也不是任意的，首先它必须要有限长. 在这个证明中，我们用到了“内错角相等”这个命题，在欧氏几何中，这个命题又根据“同位角相等”推出，而同位角相等又由另外一些命题推出，不断往上追溯简直没完没了. 真的能没完没了吗？如果追不到头，就违背了“逻辑链条有限长”的原则了. 因此必须有一些命题不由其他命题证出，这样的命题我们称为公理(axiom). 请注意，这里并没有要求公理必须被赋值为真，现在讨论的是命题逻辑的语法，和语义没有关系. 在上例中，我们所用到的公理既可以是经典的欧几里得五大基本公设，也可以是现代解析几何中的一些假设，但是这些公理都只与几何概念有关，我们还需要另外附加一些的公理，更确切地说称为公理模式(axiom scheme)，只与命题逻辑有关，并且可以把具体的命题套进去，使得证明有

出发点，这些公理称为逻辑公理(logical axiom). 逻辑公理的集合习惯使用大写希腊字母 Λ 表示，而那些于逻辑无关的公理的集合习惯上使用大写希腊字母 Γ 表示. $\Gamma\cup\Lambda$ 就是所有公理，但逻辑公理可能有点太浅显所以常常忽略.

其次，逻辑链条的每一环到下一环都必须要有依据. 问一个看起来很蠢的问题：为什么内错角相等，而 β 和 β_1 互为内错角，就能推出 β 和 β_1 相等？当然，我们可以用真假赋值函数来验证这是一个重言式. 然而，真假赋值只能用来验证某个命题是不是某些命题的语义后承，但并没有明确指出这些命题有哪些语义后承——即只能 "验旧"，没告诉你怎么 "推新"．因此我们需要明确规定怎样才能 "推新"，这种规定称为推理规则(rule of inference). 利用名为 XXX 的推理规则从旧的命题 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 推出新命题 β，记作

$$\{\alpha_1,\alpha_2,\cdots,\alpha_n\}\vdash\beta \qquad (\text{XXX})$$

或者用竖式记作

$$\frac{\alpha_1,\alpha_2,\cdots,\alpha_n}{\beta}(\text{XXX})$$

推理规则直接反映了其制定者的哲学立场，这里算是一处数学与哲学的交汇之地.

规定了公理和推理规则，我们就搭好了一个公理化的形式推演系统.

定义 推演系统 (deduction system)

公理化的形式推演系统建立在形式语言 $\mathcal{L}$ 上，包含公理 Γ、逻辑公理 Λ 和推理规则三部分. 如果一个公式 α 满足 $\alpha\in\Gamma\cup\Lambda$ 或能被 $\Gamma\cup\Lambda$ 使用有限次推理规则得到，就称 α 是 Γ 的一个内定理(inner theorem) 或语法后承(syntactic consequence)，记作 $\Gamma\vdash\alpha$，而该推演过程就称为 α 的证明(proof). 特别地，如果 $\Gamma=\emptyset$，那么 $\emptyset\vdash\alpha$ 可以简写为 $\vdash\alpha$；如果 Γ 只有一个元素 φ，那么 $\{\varphi\}\vdash\alpha$ 可以简写为 $\varphi\vdash\alpha$.

如果 $\Gamma\vdash\varphi$，且 $\Gamma\vdash\neg\varphi$，就说公理 Γ 是不一致的(inconsistent)，否则就是一致的(consistent).

这里需要再三强调，命题逻辑的语言是一个形式语言，而我们正在用自然语言叙述它的方方面面. 在这种情况下，自然语言就是元语言(metalanguage)，命题逻辑的语言就是对象语言. 接下来我们会讨论 "命题逻辑语言本身就具有的性质"，或者 "关于命题逻辑语言本身的定理"，而这些性质定理只能用自然语言描述，所以它们称为元定理(metatheorem). 上面提到的 "内定理" 则是在搭好了推演系统后，完全使用对象语言导出的合式公式，一定要分清这两者的区别. 将来熟悉形式推演之后，叫做定理就好了. 后面当我们讨论到一阶逻辑语言和二阶逻辑语言的时候，还会再遇到元定理/内定理的区别.

另外这里笼统地把证明定义为推演的过程，是因为推演最终会形成有命题组成的树，这棵树就是证明的严格定义. 然而树是组合学的概念，我们将来讨论到组合集合论的时候再详细展开，实际上，在接下来要介绍的希尔伯特推演系统中，推演过程形成的命题树只有一支，即退化为了序列.

在直觉上我们会认为，只要保持一致性，往 Γ 中塞入的公理越多，推演系统能导出的公式就越多（或者不变，但至少不会减少)，这种直觉是对的 (, 因此可以说 $\Gamma=\emptyset$ 能推出来的内定理的 "普适性" 是最广的.

7.5 希尔伯特推演系统

现在我们来介绍一个具体的形式推演系统：希尔伯特系统. 接下来如果没有特别说明，就默认使用这里介绍的希尔伯特系统.

其实严格的措辞应该是 "希尔伯特风格的命题逻辑形式推演系统"，首先不存在所谓 "希尔伯特系统"，只有希尔伯特风格(Hilbert style)，指的是有很多条逻辑公理、但只有极少推理规则的特点；其次这是命题逻辑的推演系统，因为我们接下来还会介绍希尔伯特一阶逻辑的推演系统.

我们要介绍的推演系统建立在命题逻辑语言 $\mathcal{L}_0$ 上，只用 $\neg, \rightarrow$ 两个连接词，逻辑公理 Λ 包含三条公理：

(A1) 宽容律. $\alpha \rightarrow \beta \rightarrow \alpha$

如果 α 成立，那么无论 β 是什么，α 都成立. 例如 α : 明天下雨；β : 我做某事；(A1):"如果明天下雨，那么不管我做什么，明天都会下雨". 需要再三强调，这里的"成立"是语法层面的成立，和真假赋值无关.

(A2) 蕴含分配律. $(\alpha \rightarrow \beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma)$

例如 α : 某天是周二；β : 某天不下雨；γ : 某天上体育课；(A2):"如果每个周二不下雨就上体育课，那么如果每个周二都不下雨，就每个周二都上体育课".

(A3) 逆否命题法则. $(\neg\beta \rightarrow \neg\alpha) \rightarrow (\neg\beta \rightarrow \alpha) \rightarrow \beta$

如果 β 不成立能推出 α 不成立也能推出 α 成立，那么 β 就是成立的.

历史上有不少数学家提出了各自的希尔伯特风格的推演系统，包括一些大数学家例如弗雷格 (Gottlob Frege)、塔斯基 (Alfred Tarski)、罗素，甚至是希尔伯特本人，当然它们的"推演能力"都是相当的（能演绎出一样多的命题），我们选取的这三条公理由美国数学家门德尔松 (Elliot Mendelson) 提出.

你可以把任何简单或复杂的公式当成 α, β, γ 等套入逻辑公理，得到的公式仍然可以称为逻辑公理，例如 (A1):$\alpha \rightarrow (\beta \rightarrow \alpha)$，把 $\beta \rightarrow \gamma$ 套入 α，把公理 (A2):$(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$ 套入 β，就得到了

$$((\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))) \rightarrow ((\beta \rightarrow \gamma) \rightarrow ((\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))))$$

这么一个又臭又长的公式，但它仍然能称为"公理". 所以逻辑公理看似只有三条，实际上是无数条，往里套用不同的公式都能得到不同的公理. 这三条公理只是一个框架，允许把具体的公式套进去. 剧透一下，在一阶逻辑中，我们把这种"框架型"的公理称为公理模式(axiom schema)，而套用了具体公式产生的新公式称为该公理模式的实例(instance)（如果你学过有"类"这个概念的编程语言，相信你一定能很快理解公理模式和实例的概念）. 进一步剧透：在二阶逻辑中这两个概念又将有新的诠释，这种新的诠释正是一阶二阶逻辑最大的不同点.

希尔伯特推演系统的推理规则只有一条，称为肯定前件或分离规则(modus ponens, 'method of replacing by putting' 的拉丁文简写，简称 MP)，

$$\frac{\alpha, \alpha \rightarrow \beta}{\beta}(\text{MP})$$

因为只有一条推理规则，所以**证明过程能简化为一条公式序列**，所以以后的证明将不使用竖式写法，而使用自然语言附上每一步的依据.

接下来我们做三件很无趣的事情：推几个 $\Gamma = \emptyset$ 时的所谓"普适性最广"的内定理、推几个元定理、列举我们认为理所当然的内定理.

首先我们推几个常用的内定理. 推出的内定理就像逻辑公理一样，也是"定理模式"，可以把具体的公式往里套. （我猜你不会想看以下这些又臭又长的演绎过程）

- 自蕴含律-内定理. $\vdash \alpha \rightarrow \alpha$

证明

1. $\vdash (\alpha \rightarrow (\alpha \rightarrow \alpha) \rightarrow \alpha) \rightarrow (\alpha \rightarrow (\alpha \rightarrow \alpha)) \rightarrow (\alpha \rightarrow \alpha)$（公理 A2 的实例）
2. $\vdash \alpha \rightarrow ((\alpha \rightarrow \alpha) \rightarrow \alpha)$（公理 A1 的实例）

3. $\vdash (\alpha \to (\alpha \to \alpha)) \to (\alpha \to \alpha)$（根据 1 和 2 使用肯定前件推出）
4. $\vdash \alpha \to (\alpha \to \alpha)$（公理 A1 的实例）
5. $\vdash \alpha \to \alpha$（根据 3 和 4 使用肯定前件推出）

□

- 假言三段论(hypothetical syllogism)-内定理. $(\beta \to \gamma) \to (\alpha \to \beta) \to (\alpha \to \gamma)$

设 α:“正在讨论的某人是我”，β:“某人上了富豪榜”，γ:“某人很有钱”. 假言三段论:“某人上了富豪榜就说明他很有钱，我上了富豪榜，所以我很有钱”.

证明 有些公式写出来太长了，令 φ 为公理 A2 的实例: $(\alpha \to \beta \to \gamma) \to (\alpha \to \beta) \to (\alpha \to \gamma)$.

1. $\vdash \varphi \to ((\beta \to \gamma) \to \varphi)$（公理 A1 的实例）
2. $\vdash (\beta \to \gamma) \to \varphi$（根据 1 和公理 A2 使用肯定前件推出）
3. $\vdash ((\beta \to \gamma) \to \varphi) \to (((\beta \to \gamma) \to (\alpha \to (\beta \to \gamma))) \to ((\beta \to \gamma) \to ((\alpha \to \beta) \to (\alpha \to \gamma))))$（公理 A2 的实例）
4. $\vdash ((\beta \to \gamma) \to (\alpha \to (\beta \to \gamma))) \to ((\beta \to \gamma) \to ((\alpha \to \beta) \to (\alpha \to \gamma)))$（根据 2 和 3 使用肯定前件推出）
5. $\vdash (\beta \to \gamma) \to (\alpha \to (\beta \to \gamma))$（公理 A1 的实例）
6. $\vdash (\beta \to \gamma) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$（根据 4 和 5 使用肯定前件推出）

□

- 假言三段论-内定理（调换大小前提）.

证明

1. $\vdash \varphi_1 : (\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$（假言三段论的实例）
2. $\vdash \varphi_2 : ((\alpha \to \beta) \to (\alpha \to \gamma)) \to (\beta \to (\alpha \to \beta)) \to (\beta \to (\alpha \to \gamma))$（假言三段论的实例）
3. $\vdash \varphi_2 \to \varphi_1 \to (\alpha \to (\beta \to \gamma)) \to (\beta \to (\alpha \to \beta)) \to (\beta \to (\alpha \to \gamma))$（假言三段论的实例）
4. $\vdash \varphi_1 \to (\alpha \to (\beta \to \gamma)) \to (\beta \to (\alpha \to \beta)) \to (\beta \to (\alpha \to \gamma))$（根据 2 和 3 使用肯定前件推出））
5. $\vdash \varphi_3 : (\alpha \to (\beta \to \gamma)) \to (\beta \to (\alpha \to \beta)) \to (\beta \to (\alpha \to \gamma))$（根据 3 和 4 使用肯定前件推出）
6. $\vdash \varphi_3 \to ((\alpha \to \beta \to \gamma) \to (\beta \to \alpha \to \beta)) \to ((\alpha \to \beta \to \gamma) \to (\beta \to \alpha \to \gamma))$（公理 A2 的实例）
7. $\vdash ((\alpha \to \beta \to \gamma) \to (\beta \to \alpha \to \beta)) \to ((\alpha \to \beta \to \gamma) \to (\beta \to \alpha \to \gamma))$（根据 5 和 6 用肯定前件推出）
8. $\vdash \beta \to \alpha \to \alpha$（公理 A1 的实例）
9. $\vdash (\beta \to \alpha \to \beta) \to (\alpha \to \beta \to \gamma) \to (\beta \to \alpha \to \beta)$（公理 A1 的实例）
10. $\vdash (\alpha \to \beta \to \gamma) \to (\beta \to \alpha \to \beta)$（根据 8 和 9 用肯定前件推出）
11. $\vdash (\alpha \to \beta \to \gamma) \to (\beta \to \alpha \to \gamma)$（根据 9 和 10 用肯定前件推出）
12. $\vdash \varphi_1 \to ((\alpha \to \beta) \to (\beta \to \gamma) \to (\alpha \to \gamma))$（11 的实例）
13. $\vdash (\alpha \to \beta) \to (\beta \to \gamma) \to (\alpha \to \gamma)$（根据 11 和 12 用肯定前件推出）

□

- 双重否定消去-内定理. $\vdash \neg\neg\alpha \to \alpha$

证明

1. $\vdash (\neg\alpha \to \neg\neg\alpha) \to (\neg\alpha \to \neg\alpha) \to \alpha)$（公理 *A3* 的实例）

2. $\vdash ((\neg\alpha \to \neg\neg\alpha) \to (\neg\alpha \to \neg\alpha \to \alpha)) \to (\neg\neg\alpha \to (\neg\alpha \to \neg\neg\alpha) \to (\neg\alpha \to \alpha) \to \alpha)$（公理 *A1* 的实例）

3. $\vdash \varphi : \neg\neg\alpha \to (\neg\alpha \to \neg\neg\alpha) \to (\neg\alpha \to \alpha) \to \alpha$（根据 *1* 和 *2* 使用肯定前件推出）

4. $\vdash \varphi \to ((\neg\neg\alpha \to \neg\alpha \to \neg\neg\alpha) \to (\neg\neg\alpha \to (\neg\alpha \to \neg\alpha) \to \alpha))$（公理 *A2* 的实例）

5. $\vdash (\neg\neg\alpha \to \neg\alpha \to \neg\neg\alpha) \to (\neg\neg\alpha \to (\neg\alpha \to \neg\alpha) \to \alpha)$（根据 *3* 和 *4* 使用肯定前件推出）

6. $\vdash \neg\neg\alpha \to \neg\alpha \to \neg\neg\alpha$（公理 *A1* 的实例）

7. $\vdash \neg\neg\alpha \to (\neg\alpha \to \neg\alpha) \to \alpha$（根据 *5* 和 *6* 用肯定前件推出）

8. $\vdash (\neg\neg\alpha \to (\neg\alpha \to \neg\alpha) \to \alpha) \to ((\neg\neg\alpha \to \neg\alpha \to \neg\alpha) \to (\neg\neg\alpha \to \alpha))$（公理 *A2* 的实例）

9. $\vdash (\neg\neg\alpha \to \neg\alpha \to \neg\alpha) \to (\neg\neg\alpha \to \alpha)$（根据 *7* 和 *8* 用肯定前件推出）

10. $\vdash \neg\alpha \to \neg\alpha$（自蕴含律的实例）

11. $\vdash (\neg\alpha \to \neg\alpha) \to (\neg\neg\alpha \to \neg\alpha \to \neg\alpha)$（公理 *A1* 的实例）

12. $\vdash \neg\neg\alpha \to \neg\alpha \to \neg\alpha$（根据 *10* 和 *11* 用肯定前件推出）

13. $\vdash \neg\neg\alpha \to \alpha$（根据 *9* 和 *12* 用肯定前件推出）

□

- 双重否定引入-内定理. $\vdash \alpha \to \neg\neg\alpha$.

证明

1. $\vdash \varphi_1 : (\neg\neg\neg\alpha \to \neg\alpha) \to (\neg\neg\neg\alpha \to \alpha) \to \neg\neg\alpha)$（公理 *A3* 的实例）

2. $\vdash \varphi_1 \to (\neg\alpha \to (\neg\neg\neg\alpha \to \neg\alpha) \to (\neg\neg\neg\alpha \to \alpha) \to \neg\neg\alpha)$（公理 *A1* 的实例）

3. $\vdash \varphi_2 : \alpha \to (\neg\neg\neg\alpha \to \neg\alpha) \to (\neg\neg\neg\alpha \to \alpha) \to \neg\neg\alpha$（根据 *1* 和 *2* 使用肯定前件推出）

4. $\vdash \varphi_2 \to ((\alpha \to \neg\neg\neg\alpha \to \neg\alpha) \to (\alpha \to (\neg\neg\neg\alpha \to \alpha) \to \neg\neg\alpha))$（公理 *A2* 的实例）

5. $\vdash (\alpha \to \neg\neg\neg\alpha \to \neg\alpha) \to (\alpha \to (\neg\neg\neg\alpha \to \alpha) \to \neg\neg\alpha)$（根据 *3* 和 *4* 使用肯定前件推出）

6. $\vdash \neg\neg\neg\alpha \to \neg\alpha$（双重否定消去的实例）

7. $\vdash (\neg\neg\alpha \to \neg\alpha) \to \alpha \to (\neg\neg\alpha \to \neg\alpha)$（公理 *A1* 的实例）

8. $\vdash \alpha \to \neg\neg\neg\alpha \to \neg\alpha$（根据 *6* 和 *7* 用肯定前件推出）

9. $\vdash \alpha \to (\neg\neg\neg\alpha \to \alpha) \to \neg\neg\alpha$（根据 *5* 和 *8* 用肯定前件推出）

10. $\vdash \alpha \to ((\neg\neg\neg\alpha \to \alpha) \to \neg\neg\alpha) \to (\alpha \to \neg\neg\neg\alpha \to \alpha) \to (\alpha \to \neg\neg\alpha)$（公理 *A2* 的实例）

11. $\vdash (\alpha \to \neg\neg\neg\alpha \to \alpha) \to (\alpha \to \neg\neg\alpha)$（根据 *9* 和 *10* 用肯定前件推出）

12. $\vdash \alpha \to \neg\neg\neg\alpha \to \alpha$（公理 *A1* 的实例）

13. $\vdash \alpha \to \neg\neg\alpha$（根据 *11* 和 *12* 用肯定前件推出）

□

- 换质换位-内定理. $\vdash (\alpha \to \beta) \to (\neg\beta \to \neg\alpha)$

证明

1. $\vdash \beta \to \neg\neg\beta$

2. $\vdash (\beta \to \neg\neg\beta) \to (\alpha \to \beta) \to (\alpha \to \neg\neg\beta)$

3. $\vdash \varphi_1 : (\alpha \to \beta) \to (\alpha \to \neg\neg\beta)$

4. $\vdash \neg\neg\alpha \to \alpha$

5. $\vdash (\neg\neg\alpha) \to (\alpha \to \neg\neg\beta) \to (\neg\neg\alpha \to \neg\neg\beta)$

6. $\vdash \varphi_2 : (\alpha \to \neg\neg\beta) \to (\neg\neg\alpha \to \neg\neg\beta)$

7. $\vdash \varphi_2 \to \varphi_1 \to ((\alpha \to \beta) \to (\neg\neg\alpha \to \neg\neg\beta))$

8. $\vdash \varphi_1 \to ((\alpha \to \beta) \to (\neg\neg\alpha \to \neg\neg\beta))$

9. $\vdash (\alpha \to \beta) \to (\neg\neg\alpha \to \neg\neg\beta)$

10. ...

□

接下来证明几个元定理

(1) 归纳原理(induction principle). 设 S 是合式公式的集合，$\Gamma \subseteq S, \Lambda \subseteq S$，并且由 S 中的公式通过肯定前件推导出的新公式也在 S 中（即肯定前件在 S 中封闭），则 S 中有 Γ 的全部定理.

(2) 演绎定理(deduction theorem). 若 $\Gamma \cup \{\alpha\} \vdash \beta$，则 $\Gamma \vdash \alpha \to \beta$，反之亦然.

证明 假设 $(\beta_1, \beta_2, \cdots, \beta_n)$ 是 $\Gamma \cup \{\alpha\} \vdash \beta$ 的一条推演序列，并且 $\beta_n = \beta$. 我们可以对 β_i 实行归纳.

归纳第一步，对于 $i = 1$ 的情况，有 $\beta_1 \in \Gamma \cup \{\alpha\} \cup \Lambda$，要么 $\beta_1 = \alpha$. 对于 $\beta_1 \in \Gamma \cup \Lambda$ 的情况，按定义有 $\Gamma \to \beta_1$，注意到 $\beta_1 \to (\alpha \to \beta_1)$ 是公理 *A1* 的实例，使用肯定前件即可得到 $\Gamma \vdash \alpha \to \beta_1$. 对于 $\beta_1 = \alpha$ 的情况，问题就转化为了 $\Gamma \cup \{\alpha\} \vdash \alpha$ 如何推导出 $\Gamma \vdash \alpha \to \alpha$. 别忘了 $\alpha \to \alpha$ 是不依赖 Γ 的内定理.

归纳第二步，假设 $\Gamma \cup \{\alpha\} \vdash \beta_i$ 已经可以导出 $\Gamma \vdash \alpha \to \beta_i$ 了，此时要么 $\beta_{i+1} \in \Gamma \cup \{\alpha\} \cup \Lambda$，要么 β_{i+1} 由 β_j 和 $\beta_l = \beta_j \to \beta_{i+1}$ 使用肯定前件得到（$j, l < i+1$）. 前者就像 $i = 1$ 的情况一样处理. 对于后者，既然我们已经有了 $\Gamma \cup \{\alpha\} \vdash \beta_j$ 和 $\Gamma \cup \{\alpha\} \vdash \beta_l$，又注意到 $(a \to (\beta_j \to \beta_{i+1})) \to ((\alpha \to \beta_j) \to (\alpha \to \beta_{i+1}))$ 是公理 *A2* 的实例，用两次肯定前件就可以得到 $\Gamma \vdash \alpha \to \beta_{i+1}$ 了.

反过来其实就是肯定前件. 已知 $\Gamma \vdash \alpha \to \beta$ 且 $\Gamma \cup \{\alpha\} \vdash \alpha$，根据肯定前件自然有 $\Gamma \cup \{\alpha\} \vdash \beta$. □

(3) 换质换位-元定理(contraposition). 若 $\Gamma \cup \{\alpha\} \vdash \neg\beta$，则 $\Gamma \cup \{\beta\} \vdash \neg\alpha$.

证明 已知 $\Gamma \cup \{\alpha\} \vdash \neg\beta$，根据演绎定理可得 $\Gamma \vdash \alpha \to \neg\beta$，根据换质换位内定理和肯定前件可得 $\Gamma \vdash \beta \to \neg\alpha$，根据演绎定理可得 $\Gamma \cup \{\beta\} \vdash \neg\alpha$. □

(4) 爆炸原理-元定理(ex falso quodlibet). 若 Γ 不一致，则每个公式都是它的内定理，反之亦然.

证明 (⇒) 已知 Γ 不一致，则 $\Gamma \vdash \alpha$ 且 $\Gamma \vdash \neg\alpha$. 任选一个公式 β. 根据公理 *A1*，有 $\Gamma \vdash \neg\alpha \to (\beta \to \neg\alpha)$；根据肯定前件，有 $\Gamma \vdash \beta \to \neg\alpha$；根据换质换位内定理，有 $\Gamma \vdash \alpha \to \neg\beta$；根据肯定前件，有 $\Gamma \vdash \beta$.

(⇐) 显然，如果每个公式都是 Γ 的定理，那么某个公式 α 和其否定 $\neg\alpha$ 也是，所以 Γ 不一致. □

(4) 归谬-元定理(reductio ad absurdum). 若 $\Gamma \cup \{\alpha\}$ 不一致，则 $\Gamma \vdash \neg\alpha$，反之亦然.

证明 (⇒) 若 $\Gamma \cup \{\alpha\}$ 不一致，根据爆炸原理元定理，有 $\Gamma \cup \{\alpha\} \vdash \alpha$；根据演绎定理，有 $\Gamma \vdash \neg\alpha \to \alpha$；根据公理 *A3*，有 $\Gamma \vdash \neg\alpha \to \alpha$.

(⇐) 当 $\Gamma \cup \{\alpha\}$ 不一致时有 $\Gamma \cup \{\alpha\} \vdash \beta$ 且 $\Gamma \cup \{\alpha\} \vdash \neg\beta$，根据演绎定理可得 $\Gamma \vdash \alpha \to \beta$ 且 $\Gamma \vdash \alpha \to \neg\beta$. 容易验证 $\{\alpha \to \beta, \alpha \to \neg\beta\} \vDash \neg\alpha$，所以由规则 T 可得 $\Gamma \vdash \neg\alpha$. □

(5) 冗余假设定理(redundant hypothesis theorem). 若 $\Gamma \cup \{\neg\alpha\} \vdash \alpha$，则 $\Gamma \vdash \alpha$.

证明 已知 $\Gamma \cup \{\neg\alpha\} \vdash \alpha$，根据演绎定理 $\Gamma \vdash (\neg\alpha \to \alpha)$. 又根据公理 *A3*，$\Gamma \vdash (\neg\alpha \to \alpha) \to \alpha$，再用肯定前件就得到了 $\Gamma \vdash \alpha$. □

7.6 语义和语法的统一

上面几节讨论了命题逻辑的语义和语法，但到现在为止还没有说明什么是语义、什么是语法，不过你肯定已经看出来了，讨论语义离不开真假赋值，讨论语法的时候注重的是运用规则进行推演而不管推导出句子的真假. 在语言学中，语法(syntax) 关注的是不同词性的词该怎样按照正确的顺序排列成句子，在形式语言中对应着形式推演；语义(semantics) 则还需要关注每个词的含义以及组成的句子的意义，在形式语言中对应着真值指派. 语法和语义二者有时并不一定兼得，语言学中最著名的例子莫过于 Noam Chomsky 写的

> Colorless green idea sleeps furiously.
> 没有颜色的绿色睡得很激烈。

这句话在语法上是正确的，主谓宾定状的顺序都正确，但这个句子却没什么意义甚至自相矛盾. 到现在为止，我们已经讨论了公式集 Σ 和另一个单独公式 φ 之间可能存在的语义后承和语法后承，语义后承要求无论什么样的真假赋值函数只要满足了 Σ 就要满足 φ，与真假赋值无关；而语法后承却和推理规则有关，不同的推演系统可能会有不同结论. 但是，我们都希望推演系统的结果在意义上是可以信任的，即不能推出假的公式；或更进一步地希望推演系统能推出一切真的句子. 明显后者的要求比前者高得多.

定义 可靠性与完备性

对于某推演系统，其公理集是 Γ，如果 $\Gamma \vdash \varphi$ 就有 $\Gamma \vDash \varphi$，就称该推演系统是可靠的(sound).
反过来，如果 $\Gamma \vDash \varphi$ 就有 $\Gamma \vdash \varphi$，就称该推演系统是完备的(complete).

命题逻辑的希尔伯特推演系统和自然推演系统既是可靠的也是完备的，接下来分别证明之.

定理 命题逻辑的希尔伯特推演系统是可靠的，即 $\Gamma \vdash \alpha$ 则 $\Gamma \vDash \alpha$.

证明 希尔伯特系统的公式 α 无非就三种，Γ 中的公式、Λ 中的公式和由肯定前件推出的内定理.
如果 $\alpha \in \Lambda$，那是最好办的，用真值表的方法很容易验证三条逻辑公理都是重言式.
如果 $\alpha \in \Gamma$，那也好办. 回顾 $\Gamma \vDash \alpha$ 的定义：当真假赋值 v 满足 Γ 中的每一个公式时就满足 α. 既然都满足 Γ 中的每一个公式了，而 $\alpha \in \Gamma$，当然满足 α.
如果 α 是由肯定前件推出的内定理，我们可以使用归纳的方法，假设 $\Gamma \vDash \beta, \Gamma \vDash (\beta \to \alpha)$，若要 $\beta \to \alpha$ 为真，而且 β 为真，那么 α 只能为真，所以 $\Gamma \vDash \alpha$. 归纳完毕，所有的内定理都是 Γ 的语义后承. □

要证明可靠性，我们只需要验证公理是否为重言式，以及满足公理集的真假赋值是否也满足推导法则推出来的新公式. 完备性比可靠性难证明得多，证明思路也非常绕：我们首先需要证明系统的完备性与“一致则可满足”等价，然后把完备性的证明转移到这上面来；接着我们在保持系统一致性的前提下往公理集 Γ 中塞入尽可能多的公式得到 Γ^*，以至于再任意塞入任何一个公式都会使系统不一致；然后我们构造一个具体的真假赋值满足这个 Γ^*，从而说明 Γ 也是可满足的，这样就证明了 Γ“一致则可满足”. 废话不多说，我们开始.

引理 对于具有公理集 Γ 的希尔伯特系统，以下两条命题是等价的.
(1) 若 Γ 一致则可满足.
(2) 系统是完备的.

证明 (1) ⇒ (2) 假设系统满足条件 *(1)*，并且对于某个公式 α，$\Gamma \vDash \alpha$，即满足 Γ 的真假赋值就一定满足 α，一定不满足 $\neg\alpha$，所以 $\Gamma \cup \{\neg\alpha\}$ 就不可满足. 根据条件 *(1)*，$\Gamma \cup \{\neg\alpha\}$ 不一致，所以 $\Gamma \cup \{\neg\alpha\} \vdash \alpha$，根据冗余假设定理就有 $\Gamma \vdash \alpha$.

(2) ⇒ (1) 假设系统满足条件 *(2)* 且一致，反证：现有公式 α，若 Γ 不可满足，那么“某个真假赋值满足 Γ”就是假命题，而“如果某个真假赋值满足 Γ 就满足公式 α”是真命题，即 $\Gamma \vDash \alpha$（这里又是爆炸原理的应用，是哪个层面的爆炸原理呢？值得仔细思考）. 根据条件 *(2)* 就有 $\Gamma \vdash \alpha$. 这里没有限制 α 是什么公式，所以用同样的论证可得 $\Gamma \vdash \neg\alpha$，矛盾，所以系统可满足. □

现在我们成功地转移了焦点，把对完备性的证明转移到了“一致则可满足”上. 以下的林登鲍姆引理是证明的重点，其断言按照某个特定方法构造出的 Γ 的超集 Γ^* 是极大且一致的. 为了更好地刻画 Γ^*，我们多证明几个命题.

引理 林登鲍姆引理 (Lindenbaum's lemma)

设 $\{\alpha_n\}_{n\in\mathbb{N}}$ 枚举 $\mathcal{L}_0$ 的所有公式，按如下方式构造 Γ 的极大一致扩张 Γ^*：

$$\Gamma_0 = \Gamma$$
$$\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{\alpha_n\}, & \Gamma_n \cup \{\alpha_n\}\text{一致} \\ \Gamma_n \cup \{\neg\alpha_n\}, & \Gamma_n \cup \{\alpha_n\}\text{不一致} \end{cases}$$
$$\Gamma^* = \bigcup_{n\in\mathbb{N}} \Gamma_n$$

按照这种方式得到的 Γ^* 有以下几个特点：
(1) Γ^* 是一致的；
(2) Γ^* 是极大的，即对于某个不在 Γ^* 的公式 φ，$\Gamma^* \cup \{\varphi\}$ 不一致；
(3) 对于 $\mathcal{L}_0$ 中的某个公式 φ，要么 $\varphi \in \Gamma$，要么 $(\neg\varphi) \in \Gamma$；
(4) 对于 $\mathcal{L}_0$ 中的某两个公式 φ, ψ，当且仅当 $\varphi \notin \Gamma^*$ 或 $\psi \in \Gamma^*$ 时，$(\varphi \to \psi) \in \Gamma^*$；

(5) Γ^* 对推演封闭，即 Γ 的所有内定理都在 Γ^* 中.

证明 *(1)* 稍微用点归纳就可以证明每个 Γ_n 都是一致的：Γ_0 一致；假设 Γ_n 一致，若 $\Gamma_n \cup \{\alpha_n\}$ 一致，即 Γ_{n+1}，否则 $\Gamma_n \cup \{\neg\alpha_n\}$ 一致，还是能得出 Γ_{n+1} 一致；所以全体 Γ_n 都是一致的. 别忘了 Γ^* 是全体 Γ_n 取并集，只要某公式在 Γ^* 中，那它一定在某个 Γ_n 中. 反设如果 Γ^* 不一致，即 $\Gamma^* \vdash \varphi$ 且 $\Gamma^* \vdash \neg\varphi$，则存在两条推演序列分别推出 φ 和 $\neg\varphi$，所以存在某个 Γ_n 使得两条推演序列中出现的所有公式都在其中，然后这个 Γ_n 就不一致了，矛盾，所以 Γ^* 一致.
(2) 证明其逆否命题. 假设 $\Gamma^* \cup \{\varphi\}$ 一致. 首先 φ 肯定被枚举出来，是某个 α_n，所以 $\Gamma_n \cup \{\varphi\}$ 一致，所以 $\varphi \in \Gamma_{n+1}$，所以 $\varphi \in \Gamma^*$.
(3) $\Gamma \cup \{\varphi\}$ 和 $\Gamma \cup \{\neg\varphi\}$ 至少有一个是一致的，根据极大性必有 $\varphi \in \Gamma^*$ 或 $\neg\varphi \in \Gamma^*$，但是两个不能同时成立否则 Γ^* 就不一致了.
(4) (5) 反证，假设 $\varphi \notin \Gamma^*$ 但 $\Gamma^* \vdash \varphi$. 根据极大性有 $\neg\varphi \in \Gamma^*$，根据推演的定义有 $\Gamma^* \vdash \neg\varphi$，所以 Γ^* 就不一致了，矛盾. □

引理 极大一致的 Γ^* 是可满足的，构造真假赋值：

$$v(A_i) = \begin{cases} T, & A_i \in \Gamma^* \\ F, & A_i \notin \Gamma^* \end{cases}$$

该真假赋值唯一满足 Γ^*，即仅满足所有 Γ^* 内的公式，而其他公式一条也不满足.

证明 我们对合式公式进行归纳，把合式公式可能出现的所有情况全讨论一遍即可. 所有合式公式仅有三种情况：是某个命题符号、形如 $\neg\alpha$、形如 $\alpha \to \beta$.

- 如果公式 α 是命题符号 A_i，看定义立即得到.
- 如果公式 α 形如 $\neg\beta$ 且 β 满足了引理，分两种情况讨论：
  若 $\beta \in \Gamma^*$，则 $v(\beta) = T$，根据林登鲍姆引理有 $\alpha \notin \Gamma^*$，同时有 $v(\alpha) = F$；
  若 $\beta \notin \Gamma^*$，则 $v(\beta) = F$，根据林登鲍姆引理有 $\alpha \notin \Gamma^*$，同时有 $v(\alpha) = F$.
  所以 α 也满足引理.
- 如果公式 α 形如 $\beta \to \gamma$ 且 β 和 γ 都满足引理，要想让 $\alpha \in \Gamma^*$，根据林登鲍姆引理，必须有 $\beta \notin \Gamma^*$ 或 $\gamma \in \Gamma^*$，即 $v(\beta) = F$ 或 $v(\gamma) = T$，这两种情况都能使得 $v(\alpha) = T$. 所以 α 也满足引理.

□

现在万事俱备，只需要小小说明一下即可.

定理 命题逻辑的希尔伯特推演系统是完备的，即 $\Gamma \vDash \alpha$ 则 $\Gamma \vdash \alpha$.

证明 假定 Γ 是一致的，将 Γ 扩张成极大一致的 Γ^*. 因为 Γ^* 是可满足的，而满足 Γ^* 的真假赋值也满足其子集 Γ，所以 Γ 也是可满足的. 这就最终说明了系统是完备的. □

这个长长的证明过程到这里就结束了，不知道你有没有注意到希尔伯特系统的完备性有一个很重要的前提：在林登鲍姆引理中，我们使用一个序列 $\{\alpha_n\}_{n\in\mathbb{N}}$ 枚举了 $\mathcal{L}_0$ 的所有公式，这是因为 $\mathcal{L}_0$ 是一切合法的有限长的公式的集合，是可数无穷的. 剧透一下，一阶逻辑中的语言 $\mathcal{L}_0$ 可没有这个性质，因为一阶逻辑引入了“任意”这个量词，允许我们在其他集合中取其元素组成公式，而这个“其他集合”的势有多大就

不知道了，所以一阶逻辑乃至后面的高阶逻辑语言的推演系统就没有完备性了，这个问题由哥德尔和冯 · 诺伊曼最先发现. 命题逻辑的希尔伯特系统（或者其他具有有限个公理和推理法则的推演系统）具有完备性，意味着如果我们想判别某个公式是不是推演系统内定理，验证它是不是重言式就可以了. 但是一个具有 n 个命题符号的公式，其命题符号的真值有 2^n 种情况，所以如果你想用真值表验证某个命题是不是重言式的话，需要把这 2^n 种情况都过一遍，即需要指数时间 EXPTIME 才能验证. 有没有时间复杂度更低的算法比如多项式时间的算法呢？很可惜，这个问题和计算理论中著名的 “P = NP” 问题等价.

8 一阶逻辑

8.1 一阶逻辑的语言

定义 一阶逻辑的语言

一阶逻辑语言由以下几个部分组成：

- 连接词和括号，作用与命题逻辑相同.
- 变量(variable)，记作 $v_1, v_2, \cdots$.
- 全称量词(universal quantifier)，写作 “$\forall$”，用于修饰变量.
- 谓词符号(predicate symbol)，记作 $P_1, P_2, \cdots$，作用于若干个变量或常量，可以没有.
- 函数符号(function symbol)，记作 $F_1, F_2, \cdots$，作用于若干个变量或常量，可以没有.
- 常量(constant)，记作 $c_1, c_2, \cdots$，可以没有.

一般来说往往还有个等于号(equality symbol)，为了和自然语言中的等于号相区分，以下记作 $\approx$.
常量、谓词符号和函数符号是未规定的部分，统称为图册(signature)，或者非逻辑符号(non-logical symbols)，图册的集合记作 $\mathrm{sig}\mathcal{L}_1 = \{c_1, c_2, \cdots; P_1, P_2, \cdots; F_1, F_2, \cdots\}$.
当然还有括号、逗号、冒号感叹号和存在量词等为了保证无歧义和增强可读性的符号，这些不全是必须的，它们的书写场合接下来会讨论.
规定好以上的部分后，此时该一阶语言可以记作 $\mathcal{L}_1$.

关于以上的定义，我们给出几点说明.

- 作用于 $x_1, x_2, \cdots, x_n$ 这 n 个变量的谓词或函数 P，学术上往往采用波兰写法 $Px_1x_2\cdots x_n$，我们为了好看不妨写作 $P(x_1, x_2, \cdots, x_n)$.
- 等于号其实是作用于两个变量的谓词符号 $\approx(x_1, x_2)$，但是它太常用了，而且往往有固定的解释，所以经常单独拿出来，并且我们把 $\approx(x_1, x_2)$ 写作 $x_1 \approx x_2$.
- 就像命题逻辑一样，我们规定量词的优先级高于一切逻辑连接词，当谓词写成连接词形式时（比如等于号 $\approx$），优先级也高于逻辑连接词，这样可以省掉不必要的括号，当然有时省得太多也不好看，但始终坚持无歧义原则. 有些教材把 $\forall x\alpha$ 写作 $\forall x : \alpha$，可读性当然提高了，但是我们不采用.
- 我们之前常见的存在量词（$\exists$）不在这个列表里，那是因为 “存在” 可以用 “任意” 加上连接词来等价地解释：$\exists x\alpha$ 可以解释为 $\neg(\forall x(\neg\alpha))$，就举个具体的例子帮助理解：

$$\text{桌上的苹果存在有坏的} \Leftrightarrow \text{桌上的苹果不全是好的} \Leftrightarrow \text{“桌上的任意苹果都是好的” 是错的}$$

令 α 解释为"桌上的苹果是坏的"，"桌上的苹果有坏的"即为 $\exists x\alpha$，"'桌上的任意苹果都是好的'是错的"即为 $\neg(\forall x(\neg\alpha))$.

- 有的教材还会引入唯一存在量词（$\exists!$）增强公式的可读性，它当然也可以等价解释：$\exists! x\alpha$ 可以解释为 $\exists x\forall y(\alpha \to x \approx y)$，例如：

  $$\text{桌上的苹果存在只有一个是坏的} \Leftrightarrow \text{如果苹果}x\text{是坏的，那么桌上的任何坏苹果其实都是苹果}x$$

- 剧透：一阶逻辑语言的"阶"表示量词作用的类型的范围，在一阶语言 $\mathcal{L}_1$ 中，量词可以作用于变量，但是不能作用于谓词，谓词只能作用于项；但是在二阶逻辑语言 $\mathcal{L}_2$ 中，量词可以作用于谓词，谓词也能作用于谓词，但是量词不能作用于"作用于谓词的谓词"；在三阶逻辑语言中范围又扩得更宽. 你可能想问为什么不直接学高阶语言 $\mathcal{L}_\infty$，那是出于建立在语言上的推演系统性质的考虑. 我们学过推演系统的四大性质：可靠性、完全性、可判定性和公理的独立性，一阶语言的推演系统将会丢失可判定性，二阶语言的将会丢失完全性. 并且由于历史原因，大多数推演系统其实都是建立在一阶语言上，再加上哥德尔和司寇伦在一阶逻辑上的突出贡献，使得一阶语言在数理逻辑中有不可撼动的地位. 命题逻辑语言连量词都没有，也称为零阶逻辑语言，所以我们之前将所有命题逻辑语言合式公式的集合记作 $\mathcal{L}_0$.

- 变量的集合 $\{v_i\}_{i\in I}$ 是至多可数的,图册是有穷的. 当然也有变量集和图册为无穷的语言,称为<u>无穷语言</u>(infinite language)，在模型论中有独特的用处，但是现在不考虑.

- 谓词符号可以类比朴素集合论中的关系，函数符号可以类比朴素集合论中的函数，因此函数可以理解为一种特殊的谓词，而常量可以看作零元函数，因此图册本质上就是谓词的集合，所以有的教材上也把一阶逻辑语言称为一阶谓词语言，但我们还是坚持把它们都列出来.

- 请注意，$\mathcal{L}_1$ 是一个集合，里面是合式公式，$\mathrm{sig}\mathcal{L}_1 \notin \mathcal{L}_1$. 和命题逻辑的语言做个对比：命题逻辑的语言只有命题符号（对应于一阶逻辑语言的变量）、连接词和括号，公式中能出现什么符号基本都定死了；而一阶逻辑语言的定义只是搭了一个框架，具体有什么谓词、函数和常量可以根据需要确定，这些没有规定死的部分即图册，当图册确定好之后，一阶语言也就完整了.

举几个使用一阶逻辑语言的具体例子：

- 哲学语言：分析哲学这个分支的出现使得一些哲学家喜欢使用数理逻辑的语言重新叙述哲学命题，例如
  - 当今的法国国王是秃子.
    $$\forall x(P_1(x) \to P_2(x))$$
    其中 $P_1(x)$ 用自然语言解释为"x 是当今的法国国王"，$P_2(x)$ 用自然语言解释为"x 是秃子"，当然还可以有不同的解释.
  - 金山不存在.
    $$\neg\exists x(P_1(x) \wedge P_2(x))$$
  - 晨星即暮星.
    $$\forall x(P_1(x) \to \forall y(P_2(x) \to x \approx y))$$
- 纯谓词演算：有多个谓词符号、多个常量，没有等于号、没有函数符号，$\mathrm{sig}\mathcal{L}_1 = \langle c_1, c_2, \cdots; P_1, P_2, \cdots\rangle$.
- 初等数论：有常量 0，有谓词符号"$<$"和等于号，函数符号有一元的后继函数 S 和二元的加号和乘号，$\mathrm{sig}\mathcal{L}_1 = \langle 0; <; S, +, \cdot\rangle$. 给出几个例子.

  - 0 不是任何自然数的后继.
$$\neg\exists x(S(x)\approx 0)$$
  - 当且仅当两个自然数相等时，其后继数相等.
$$\forall x\forall y(x\approx y\leftrightarrow S(x)\approx S(y))$$
  - x 是素数.
$$x>S(0)\wedge\forall y\forall z((y<x\wedge z<x)\to\neg(y\cdot z\approx x)$$

- 一阶公理集合论：仅有“$\in$”一个谓词符号，没有函数符号，没有等于号，没有常量，$\sigma=\langle\in\rangle$，当然有时候把等于号放进来，把空集当常量，$\mathrm{sig}\mathcal{L}_1=\langle\emptyset;\in,\approx\rangle$. 给出几个例子.
  - x 是 y 的子集.
$$\forall z(z\in x\to z\in y)$$
  - x 是空集.
$$\forall y(\neg(y\in x))$$
  - $x=\{y,z\}$.
$$\forall s(s\in x\to(s\approx y\vee s\approx z))$$

  你可能会好奇集合论里为什么没有函数，毕竟不久前才在朴素集合论里学习了函数的概念. 但别忘了那个函数是有序对的集合，是用集合定义出来的，不是附带在语言中的. 也别忘了元语言和对象语言的区别.

接下来我们像命题逻辑一样，递归定义什么是“公式”.

定义　项和公式

设一阶语言 $\mathcal{L}_1$，$\mathrm{sig}\mathcal{L}_1=\{c_1,c_2,\cdots;P_1,P_2,\cdots;F_1,F_2,\cdots\}$，首先定义项(term)
(1) 每个常量和变量都是一个项.
(2) 设 $x_1,x_2,\cdots,x_n$ 是 n 个项，且函数符号 F_i 作用在这些项上，则 $F_i(x_1,x_2,\cdots,x_n)$ 也是一个项.
设 $x_1,x_2,\cdots,x_n$ 是 n 个项，且谓词符号 P_i 作用在这些项上，则 $P_i(x_1,x_2,\cdots,x_n)$ 是一个原子公式(atomic formula)，从原子公式开始我们可以定义合式公式(well-formed formula, WFF).
(1) 原子公式是合式公式.
(2) 若 φ 是合式公式，则 $(\neg\varphi)$ 是合式公式.
(3) 若 φ,ψ 都是是合式公式，则 $(\varphi\to\psi)$ 是合式公式.
(4) 若 φ 是合式公式，则 $\forall x\varphi$ 是合式公式.
一阶逻辑语言全体合式公式的集合记为 $\mathcal{L}_1$.

在以上的例子中，变量出现了两种，就拿 $\forall y(\neg(y\in x))$ 来说，y 被量词约束了，只是一个占位的变量，用 z 换掉 y，即 $\forall z(\neg(z\in x))$ 得到的还是相同的公式；而 x 就不同，如果在上下文中 x 还要用到，那就不能随便替换掉. 我们见过不少类似的情况，比如积分 $\int f(x)\mathrm{d}x$ 和 $\int f(y)\mathrm{d}y$ 是一样的，求和 $\sum_{i=1}^n a_i$ 和 $\sum_{j=1}^n a_j$ 是一样的，当时我们把这种只用来占位的变量称为哑变量，现在我们称它们为受囿的(bounded) 变量，而那些不被量词约束的变量就称为自由的(free) 变量. 现在我们严格递归定义.

定义 自由出现和受囿出现

- 在原子公式中出现的变量是自由出现的.
- 对于公式 $\varphi : \neg\alpha$，如果变量在 α 中自由出现，那么在 φ 也是自由出现的.
- 对于公式 $\varphi : \alpha \to \beta$，如果变量在 α 和 β 中自由出现，那么在 φ 也是自由出现的.

如果变量不是自由出现的，那就是受囿出现的. 如果公式中没有自由出现的变量，就称该公式是一个封闭的(closed) 公式.

习惯上，我们把带有自由变量 x 的公式 φ 写成 $\varphi(x)$，两个变量 x, y 就写成 $\varphi(x, y)$，你一定经常这么写.

细细品味定义中的那句话："如果变量不是自由出现的，那就是受囿出现的"，反过来："如果变量不是受囿出现的，那就是自由出现的" 对吗？也对。看似自由出现和受囿出现是对立的概念，但是请看例子

$$(\forall x(x \approx x)) \wedge (x \approx c)$$

这个 x 既是受囿出现的，因为出现了 "$\forall x$"，但也是自由出现的，因为在量词根本就管不到 "$x \approx c$" . 这个公式是符合句法的合式公式，只不过重复利用了变量 x 而已，相信你在编程中也这么做过. 这个例子是一个极端的情况，告诉我们变量既可以是受囿的也可以是自由的，当然习惯上我们不会这么写，也最好不要这么写，我们约定：接下来书写合式公式的时候，避免一个变量既是受囿又是自由的情况.

既然受囿变量可以换掉而不影响公式表达的意思，那可以随便换吗？一般问出这个问题就表明答案显然是不行. 例如

$$\exists x \neg(x \approx y) \xrightarrow{\text{把}x\text{换成}z} \exists z \neg(z \approx y)$$

这个替换是没问题的，但是

$$\exists x \neg(x \approx y) \xrightarrow{\text{把}x\text{换成}y} \exists y \neg(y \approx y)$$

左边是一个看上去很正确的公式，尤其是把它放在初等数论中解释的时候，有无穷多个自然数，固定一个自然数 y，当然存在与它不等的自然数 x 嘛. 右边的公式就错得离谱了，因为我们期望任何东西都等于它自己. 问题出在哪里呢？第一个替换在替换前后自由出现的 y 还是自由出现的，但第二个替换在替换前 y 是自由出现的，替换后就被量词辖制了，变成了受囿出现的变量，这样的替换肯定是我们不想要的，我们需要规范变量的替换行为.

8.2 一阶逻辑的语法

你一定想问，为什么要先讨论语法而不是语义. 因为我们遵循先易后难的顺序，命题逻辑的语义说白了就是命题符号和公式的真假，围绕着真假赋值讨论；而语法涉及到形式证明，可以导出许多让人眼花缭乱的性质，比语义难得多. 一阶逻辑的语法是命题逻辑逻辑语法的扩展，并且这种扩展是比较保守的，没有加入多少新内容；消除了不确定性后才能讨论意义，命题逻辑的不确定因素是命题符号的真假，所以用真假复制消除其不确定性后就可以了，但一阶逻辑由于有变量和不确定的谓词和函数，所以语义讨论起来要复杂得多，例如公式 $\forall x(\neg P(x))$，想讨论它的真假，那就得弄明白这个谓词 P 在元语言层面上该怎么解释、变量 x 的取值范围有什么规定. 因此需要引入一个外部结构，并把公式解释成外部结构中的命题，而这个 "解释" 究竟是什么，也值得我们细细研究，这就使得一阶逻辑的语义比语法难得多.

好啦，

8.3 一阶逻辑的结构

正如上节开头所说，想讨论一阶逻辑命题的语义，需要引入外部结构，并把公式解释称这个外部结构的命题.

至于为什么要引入外部结构，在给出一阶语言例子的时候给出了几个很好的例子. 仅给出公式 $\forall x(P_1(x) \to P_2(x))$，无法判断它的真假，因为我们只知道 P_1, P_2 是谓词，不清楚谓词 P_1, P_2 作用于 x 会呈现出什么样的真假，没有更多的信息了，换句话说，P_1, P_2 这两个谓词没有意义，这就导致整个公式没有意义，无法判断真假.

我们在命题逻辑中也遇到过类似的情况，公式 $A_1 \to A_2$ 是真是假？答案是不确定，我们只有知道 A_1, A_2 的真假（即这两个命题符号的意义）之后才能判断公式的真假，真假赋值 v 正是帮我们明确了每个命题符号的真假，消除了这种不确定性，我们才能判断公式的真假. 但至于 v 为什么将这个 A_i 赋值成真，将那个 A_i 赋值成假，这种“原子事实”不是数理逻辑关心的事情，我们只关心赋值完成后公式的真假.

回到一阶逻辑，当你把 $P_1(x)$ 用自然语言解释为“x 是当今的法国国王”，$P_2(x)$ 用自然语言解释为“x 是秃子”，P_1, P_2 就有了意义，不确定性就消除了. 整个公式就可以用自然语言解释成“当今的法国国王是秃子”，然后就可以判断真假了. 当然我们不关心为什么谓词要这么解释，我们只关心将不确定的部分（谓词等）解释好之后，整个公式的真假.

无论是命题逻辑还是一阶逻辑，尽管我们可以像搭积木一样摆弄公式（说的就是你，推演系统），但是其语言都有不确定的部分（命题符号、签名），导致整个公式都是没有意义的，我们只有通过“外部力量”（真假赋值、解释函数）消除这种不确定性，才能让公式“有意义”从而讨论真假，尽管我们不关心消除不确定性的过程. 这就是数理逻辑语义研究的基本内容.

所谓“将公式解释成外部结构的命题”，其重点就是把公式中不确定的部分——说的就是签名——解释为外部结构对应的部分，当然外部结构这种“对应的部分”必须要存在才行，接下来我们给出这种解释严格的规定.

定义　一阶结构

一阶逻辑的<u>结构</u>(structure) 由三部分组成：论域 M、签名 σ 和解释函数 I 组成，一并记作 $\mathcal{M} = \langle M, \sigma, I\rangle$，其中

- <u>论域</u>(domain of discourse) 是一个集合，要求是非空的.
- <u>签名</u>(signature) 由常量、关系和函数组成，$\sigma = \langle c_1^M, c_2^M, \cdots; R_1^M, R_2^M, \cdots; F_1^M, F_2^M, \cdots\rangle$，这里加了上标 M 强调这些常量都是 M 的元素、关系和函数都建立在集合 M 上，从而区分一阶语言的常量、谓词和函数.
- <u>解释函数</u>(interpretation function)，将语言的签名解释成结构中的签名，$I : \mathrm{sig}\mathcal{L}_1 \to \sigma$，分以下三种情况：
  - 对于语言中的某个常量 c_i，结构中存在某个常量 c_j^M，使得 $I(c_i) = c_j^M$.
  - 对于语言中的某个谓词符号 P_i，结构中存在某个关系 R_j^M，使得 $I(P_i) = R_j^M$.
  - 对于语言中的某个函数符号 F_i，结构中存在某个函数 F_j^M，使得 $I(F_i) = F_j^M$.

给出一个具体的例子，自然数的结构 $\mathcal{N} = \langle \mathbb{N}, \{0; <, =; S, +, \cdot\}, I\rangle$，建立在集合论的语言 $\mathcal{L}_1 = \langle \emptyset; \in, \approx\rangle$ 上，并且规定解释函数做出以下解释：

$$\begin{array}{ccc} \mathcal{L}_1 \text{ 的签名} & & \mathcal{N} \text{ 的签名} \\ \emptyset & \longmapsto & 0 \\ \in & \longmapsto & < \\ \approx & \longmapsto & = \end{array}$$

那么集合论中的公式（这其实是后面 ZFC 的空集公理）

$$\exists x \forall y \neg y \in x$$

就应该解释成

$$\exists x \forall y \neg y < x$$

"存在自然数 x，所有自然数 y 都不比它小."

这当然是对的（现在还没讨论到满足赋值，尚不严谨），取 $x=0$，那么就不存在任何小于 x 的 y 了. 集合论中的公式（其实这是后面 ZFC 的无序对公理）

$$\forall x \forall y \exists z \forall t (t \in z \to (t \approx x \lor t \approx y))$$

就应该解释成

$$\forall x \forall y \exists z \forall t (t < z \to (t = x \lor t = y))$$

"对于自然数 x, y，都存在自然数 z，只要某个自然数小于 z，那么它不是 x 就是 y."

这个公式是对的吗？当然错误嘛，取 $x=2, y=3$，这个 z 显然不存在，分几种情况讨论很快就证明出来了. 公理居然是错的，换个严谨的说法，自然数的结构 $\mathcal{N}$ 不满足 ZFC 公理集.

那么，不确定性就此消失了吗？我们改写一下空集公理，把 x 改成自由变量，得到 $\varphi(x) : \forall y \neg y \in x$，翻译之后得到 $\varphi^{\mathcal{N}} : \forall y \neg y < x$，这个公式是正确的吗？取 $x=0$ 时为真，但 x 取其它数的时候就是假的了，也就是说，我们还需要进一步处理自由变量，将自由变量赋值为论域中的某个元素. 剧透：在这个例子中，当自由变量 x 赋值为 0 即 $s(x)=0$ 时，自然数的结构 $\mathcal{N}$ 就满足 $\varphi^{\mathcal{N}}$ 了，记作 $(\mathcal{N}, s) \vDash \varphi$. 实际上，我们就像真假赋值那样，把每个变量都赋值为论域中的元素，并将其扩张以处理所有项.

定义　满足赋值及其扩张

设语言 $\mathcal{L}_1$ 中的全体变量的集合记为 $\mathcal{L}_1^0$，全体项的集合记为 T，结构 $\mathcal{M} = \langle M, \sigma, I \rangle$，则赋值函数形如 $s : \mathcal{L}_1^0 \to M$，将所有变量赋值为论域 M 中的元素，即对于变量 v_i，有 $s(v_i) \in M$.
赋值函数 s 按照如下法则扩张成 $\bar{s} : T \to M$

- 对于变量 v_i，$\bar{s}(v_i) = s(v_i)$.
- 对于常量 c_i，$\bar{s}(c_i) = I(c_i)$.
- 对于函数 $F_i(x_1, x_2, \cdots, x_n)$，$\bar{s}(F_i(x_1, x_2, \cdots, x_n)) = I(F_i)(\bar{s}(x_1), \bar{s}(x_2), \cdots, \bar{s}(x_n))$.

当确定了满足赋值以后，我们就可以正式定义一阶逻辑下的"满足"了.

定义　真值模式 (T-scheme)，塔斯基，1933

给定结构 $\mathcal{M} = \langle M, \sigma, I \rangle$ 和扩张的满足赋值函数 $\bar{s}$，递归定义 $(\mathcal{M}, s) \vDash \varphi$.

- 如果 $\varphi : P(x_1, x_2, \cdots, x_n)$，当且仅当 $(\bar{s}(x_1), \bar{s}(x_2), \cdots, \bar{s}(x_n)) \in I(P)$ 时，有 $(\mathcal{M}, s) \vDash \varphi$.
- 如果 φ 形如 $\neg\alpha$，当且仅当 $(\mathcal{M}, s) \nvDash \varphi$ 时，有 $(\mathcal{M}, s) \vDash \varphi$.
- 如果 φ 形如 $\alpha \to \beta$，当且仅当 $(\mathcal{M}, s) \nvDash \alpha$ 或 $(\mathcal{M}, s) \vDash \beta$ 时，有 $(\mathcal{M}, s) \vDash \varphi$.
- 如果 φ 形如 $\forall x \alpha$，当且仅当对于时

定义 语义后承

设 Γ 为一个公式集，φ 是一个公式，如果对于所有的结构 $\mathcal{M}$ 和满足赋值 s，只要 $(\mathcal{M}, s)$ 满足 Γ 就一定满足 φ，就称 φ 是 Γ 的语义后承，记作 $\Gamma \vDash \varphi$.

定义 模型

给定结构 $\mathcal{M}$，对于公式 φ，如果所有满足赋值均有 $(\mathcal{M}, s) \vDash \varphi$，就称 $\mathcal{M}$ 是 φ 的模型(model)，记作 $\mathcal{M} \vDash \varphi$.

9 公理化的集合论

9.1 ZFC 集合论

策梅洛-弗兰克尔集合论 (Zermelo-Frankel set theory, ZFC) 是一个关于集合的理论，它建立在一阶逻辑之上，使用希尔伯特风格的推演系统（也可以是自然推演系统，总之是形式主义的），谓词只有 $\in$，读作“属于”，没有常量和函数，也没有等于号（有时为了方便可以有），$\mathcal{L}_1 = \langle \in \rangle$，或 $\mathcal{L}_1 = \langle \in, = \rangle$. 每个变量在直觉上具有“集合”的类型. 它的公理集有 7 条公理，以下逐条叙述，正文给出最原始的形式，具体的意思之后阐述. 在书写上，暂时使用大写字母表示集合，小写字母表示元素，尽管实际上这两者其实是无法区分的，并且为了可读性，我们使用中括号承担与小括号相同的职能.

(1) 外延公理 (axiom of extensionality)：

$$\forall X \forall Y [\forall x (x \in X \wedge x \in Y) \rightarrow X = Y]$$

两个集合是否相等取决于元素是否相同，你有我也有，我有你也有，你我就相等了. 这条命题与其说是公理，不如说是补充了“等于”这个谓词的定义. 其逆命题为 $\forall X \forall Y [X = Y \rightarrow \forall x (x \in X \wedge x \in Y)]$，是一阶逻辑希尔伯特系统的内定理.

如果一阶语言没有等于号，那这个公理就是等于号的定义.

(2) 并集公理 (axiom of union)：

$$\forall X \exists Y \forall u [u \in Y \leftrightarrow \exists v (u \in v \wedge v \in X)]$$

并集公理可以改写为 $\forall X \exists Y (Y = \bigcup X)$，即任意集合的并集都存在.

(3) 幂集公理 (axiom of power set)：

$$\forall X \exists Y \forall x \forall z [(x \in z \rightarrow x \in X) \leftrightarrow z \in Y]$$

幂集公理可改写为 $\forall X \exists Y [Y = \mathcal{P}(X)]$，即任意集合的幂集都存在. 将来我们可以看到，幂集公理形式上虽然很简单，但正是它赋予了 ZFC 极大的证明论强度.

(4) 正规公理 (axiom of regularity)：

$$\forall X \exists x [x \in X \rightarrow \neg \exists u (u \in x \wedge u \in X)]$$

这个公理说的是“任意集合 X 中都存在这样一个元素 x，使得 $x \cap X = \emptyset$”，你一定会觉得这个公理很莫名其妙，但它实际上非常巧妙地断言了“$\in$”的反自反性、反对称性和良基性，从而规避了罗素悖论. 如果我们允许 $x \in x$，那么

$\{x\}$ 这个集合就不满足正规公理了，因为它唯一的元素 $x \cap \{x\} \neq \emptyset$；同理，如果我们允许 $x \in y \in x$，那么 $\{x, y\}$ 这个集合就不满足正规公理了；同理，如果我们允许 $x_0 \in x_1 \in \cdots \in x_\alpha \in x_0$，那么 $\{x_0, x_1, \cdots, x_\alpha\}$ 这个集合就不满足正规公理了.

正规公理还告诉我们，$\mathbb{ON}$ 不是一个集合，因为如果它是一个集合，那么 $\langle \mathbb{ON}, \in \rangle$ 将是一个良序集，而且任意序数 $\alpha = \{\beta | \beta < \alpha\} \subset \mathbb{ON}$，意味着 $\mathbb{ON}$ 也是传递集，所以 $\mathbb{ON}$ 是一个序数，所以有 $\mathbb{ON} \in \mathbb{ON}$，违背了正规公理.

同理，包含所有集合的集合也是不存在的，即 $\mathbb{V}$ 不是一个集合，因为如果是的话，就有 $\mathbb{V} \in \mathbb{V}$，违背正规公理.

(5) 替换公理模式 (axiom scheme of replacement)：

$$\forall X[\forall x(x \in X \land \exists y \forall z[\phi \leftrightarrow z = y]) \rightarrow \exists Y \forall x(x \in X \rightarrow \exists y[y \in Y \land \phi])]$$

其中 ϕ 是一阶合式公式.

之所以称为“公理模式”是因为其中的合式公式 ϕ 可以任选，但在一阶逻辑中全称量词不能约束公式，所以不能将其写入公理中. 将该合式公式写成 $\phi(x, y)$，式中的 $\exists y \forall z[\phi \leftrightarrow z = y]$ 说的是“存在 y 使 $\phi(x, y)$ 成立，如果另外有 z 也能使得 $\phi(x, z)$ 成立，那么 y 和 z 是同一个对象”，意味着使得 ϕ 成立的对象是唯一的，但由于 ϕ 成立与否由 x, y 共同确定，所以这个唯一的 y 由 x 确定.

你可能突然发现这个 ϕ 好像确定了一个函数 $y = f(x)$，将 $\forall x(x \in X \land \exists y \forall z[\phi \leftrightarrow z = y])$ 改写为 $\forall x(x \in X \land \exists! y[y = f(x)])$，其中 $\exists!$ 表示“唯一存在”的意思，你就会意识到这个 $X \subseteq \mathrm{dom}(f)$.

往后看，$\exists Y \forall x(x \in X \rightarrow \exists y[y \in Y \land \phi])$ 改写为 $\exists Y \forall x(x \in X \rightarrow \exists y[y \in Y \land y = f(x)])$，你又发现这个 $Y \subseteq \mathrm{ran}(f)$. 所以这个公理说的是：“如果 f 是一个函数关系，X 是一个集合，那么 $f[X]$ 也是一个集合”.

当然有的资料上将替换公理模式开头的“$\forall X$”写成“$\forall X \forall \vec{w}$”，其实就相当于用了一次一阶逻辑的概括公理. 两者区别不大.

(6) 无穷公理 (axiom of infinity)：

$$\exists X[\exists x(\neg \exists y[y \in x] \land x \in X) \land \forall x[x \in X \rightarrow \exists y(y \in X \land \forall u[u \in y \leftrightarrow (u = x \lor u \in x)])]$$

这条公理又臭又长，是因为把它严格按照一阶逻辑语言写出来了而已. $\exists x(\neg \exists y[y \in x] \land x \in X)$ 中的 x 不含元素，这说的是 $\emptyset \in X$，$\forall u[u \in y \leftrightarrow (u = x \lor u \in x)]$ 表示 y 的元素只能是 x 或 x 的元素，即 $y = x \cup \{x\}$，如果把 x 当作序数看的话那么 $y = x^+$.

所以这条公理可以改写为 $\exists X[0 \in X \land \forall x(x \in X \rightarrow x^+ \in X)]$，属于 X 的元素，要么是 0（即空集），要么是某个已属于 X 的序数的后继数——这说的不就是自然数嘛！所以无穷公理实际上断言了 $\mathbb{N}$ 的存在性.

如果联合正规公理一起看，那么无穷公理还可以写作 $\mathbb{N} \neq \mathbb{ON}$，有的集合论书上就是这么表述无穷公理的.

以上便是 ZF 集合论的公理，一般我们会加上选择公理，形成 ZFC 集合论.

(7) 选择公理 (axiom of choice)：

$$\forall X[\forall x \forall y([x \in X \land y \in X] \rightarrow [x \neq y \leftrightarrow \neg \exists z(z \in x \land z \in y)]) \rightarrow \exists Y \forall x(x \in X \rightarrow \exists! z[z \in x \land z \in Y])]$$

这条公理更长了，我们一点点来拆解它. $[x \in X \land y \in X] \rightarrow [x \neq y \leftrightarrow \neg \exists z(z \in x \land z \in y)]$ 说的是 X 中的两个不同的元素没有共同的元素，即两两不交. $\forall x(x \in X \rightarrow \exists! z[z \in x \land z \in Y])$ 意思是 X 中的每一个元素与 Y 只有一个共同元素. 如果将集合 X 写作 $\{x_i\}_{i \in I}$，那么选择公理可以翻译成一句话：“给定一系列两两不相交的集合 $\{x_i\}_{i \in I}$，存在一个集合 Y，每个 x_i 与 Y 恰好共享一个元素”.

不说完全理解，这句话至少让我们看懂了选择公理，但还是没有体现出“选择”的意思. 回忆你去商店买衣服的过程，你要从众多的衣服中挑出一件称心的，而挑出的衣服是商店里有的（废话）. 如果记商店里的衣服集合为 X，用函数 f 表

示你挑衣服的过程，你挑选出的衣服为 $f(X)$，自然有 $f(X) \in X$. 对于 $S = \{x_i\}$ 中的每个 x_i，满足 $f(x_i) \in x_i$ 的函数称为 S 的选择函数(choice function)，选择公理断言，每个非空集合都有其选择函数. 用买衣服的例子来理解，S 是一家大商场，x_i 是其中的每个商店，$f(x_i)$ 是你在商店 x_i 中挑出的那件衣服，选择公理断言你总能在每一家店里都挑出一件衣服，最后出商场时买下的所有衣服能构成一个集合. 这样解释起来就自然多了.

ZFC 原本还有以下几条公理，不过它们是多余的，能被上面那几条公理推出来.

- 空集公理 (axiom of empty set)：

$$\exists X[\neg x(x \in X)]$$

这条公理断言空集的存在性，可以被证明：无穷公理断言了 $\mathbb{N}$ 的存在性，根据正规公理和下面的概括公理模式，有

$$\emptyset = \{x \in \mathbb{N} | x \in x\}$$

- 概括公理模式/子集公理模式/分离公理模式/分类公理模式 (axiom of comprehension / subset / separation / specification)

$$\forall X \exists Y \forall x \forall \vec{w}[(x \in X \wedge \phi) \leftrightarrow x \in Y]$$

其中 ϕ 是合式公式.

这条公理是朴素集合论中概括原则的削弱版本，概括原则断言形如 $\{x|\phi\}$ 的是集合，概括公理模式则需要先准备好一个集合 X，然后才断言 $Y = \{x \in X | \phi\}$ 是一个集合. 它可以被证明：在替换公理模式中，令合式公式

$$\phi(x, y, w) : x = y \wedge x \in w \wedge \psi$$

立即知道这个公式对参数 y 有外延性. 替换公理模式断言 $\{y | \phi(x, y, X), x \in X\}$ 是一个集合，稍微改写一下，这个集合就是 $\{x \in X | \psi\}$.

- 无序对公理 (axiom of pairing)：

$$\forall x \forall y \exists z[\forall w(w \in z \leftrightarrow [w = x \vee w = y])]$$

这条公理说的是："对于任意集合 x, y，$\{x, y\}$ 是一个集合"，看起来有点无厘头，也可以被证明：首先空集是一个集合，根据幂集公理，$\mathcal{P}(\mathcal{P}(\emptyset)) = \{\emptyset, \{\emptyset\}\}$ 也是一个集合. 对于任意集合 u, v，在替换公理模式中，令

$$\phi(x, y, u, v) := (x = \emptyset \wedge y = u) \vee (\neg(x = \emptyset) \wedge y = v)$$

可以注意到这个公式对 y 有外延性. 替换公理模式断言 $\{y | \phi(x, y, u, v), x \in \{\emptyset, \{\emptyset\}\}\}$ 是一个集合，稍微整理一下，这个集合就是 $\{u, v\}$.

我们之所以选择公理集合论，是因为朴素集合论出现了瑕疵，于是使用公理化的方法扬弃朴素集合论的内容. 但不少教材并没有提到公理化的集合论，一个重要的原因是 ZFC 抛弃掉的朴素的那一部分，在大多数数学对象的构造中并不会触碰到. 相反，ZFC 下集合的行为仍然是相当自由的，以下证明一个命题来刻画这种自由.

定理 对于两个集合 A, B，$A \cup B, A \cap B, A - B, B - A, A \triangle B, (A, B), A \times B$ 都是存在的.

证明 $A\cup B=\bigcup\{A,B\}$，由并集公理和无序对公理保证其存在性.
$A\cap B=\{x\in A|x\in B\}$，由分离公理模式保证其存在性.
$A-B=\{x\in A|x\notin B\}$，由分离公理模式保证其存在性，$B-A$ 类似，而 $A\triangle B=(A-B)\cup(B-A)$，额外需要并集公理.
$(A,B)=\{\{A\},\{A,B\}\}\subset\mathcal{P}(\mathcal{P}(A\cup B))$，由并集公理和无序对公理保证其存在性.
对于 $A\times B$，任选 $a\in A,b\in B$，由于 $\{a\}\subseteq A$，所以 $\{a\}\in\mathcal{P}(A)$. 同理，由于 $\{a,b\}\subseteq A\cup B$，所以 $\{a,b\}\in\mathcal{P}(A\cup B)$. 因此 $\{\{a\},\{a,b\}\}\subset\mathcal{P}(A\cup B)$，得到 $(a,b)=\{\{a\},\{a,b\}\}\in\mathcal{P}(\mathcal{P}(A\cup B))$. 最终我们有 $A\times B=\{p\in\mathcal{P}(\mathcal{P}(A\cup B))|\exists a\exists b(a\in A\wedge b\in B\wedge p=(a,b))\}$，这由并集公理、幂集公理、分离公理模式保证其存在性. □

10 再次讨论正规公理

10.1 冯 · 诺伊曼层垒谱系

一个集合是怎么得到的？历史上最初引入集合论，就是为了将某些具有相同性质的对象集中讨论，这些对象不一定是一个集合. 我们把所有不是集合但可以属于某个集合的东西称为原子(atmo)，以表示这些对象不宜使用集合论的语言再次分割，比如样本空间中的样本点、向量空间中的向量等.

把所有原子的集合记作 V_0，把目光看向 $V_1=\mathcal{P}(V_0)$ 这个集合上，它以 V_0 的子集作为元素，而 V_0 的子集要么是空集，要么以赤裸裸的原子作为元素，所以 V_1 的元素可以理解为“给若干原子穿上一层集合的衣服”的所有可能情况. 同理，$V_2=\mathcal{P}(\mathcal{P}(V_0))$ 的元素为“给若干原子穿上两层集合的衣服”的所有可能情况，由此归纳下去，V_n 即“给若干原子穿上 n 层集合的衣服”的所有可能情况. 举个例子，假如一共有 2 个原子，记作 $1,2$，那么

$$V_0=\{1,2\}$$

是赤裸裸的原子.

$$V_1=\{\emptyset,\{1\},\{2\},\{1,2\}\}$$

是“给若干原子穿上一层集合的衣服”的所有可能情况.

$$\begin{aligned}V_2=\{&\\&\emptyset,\\&\{\emptyset\},\{\{1\}\},\{\{2\}\},\{\{1,2\}\},\\&\{\emptyset,\{1\}\},\{\emptyset,\{2\}\},\{\emptyset,\{1,2\}\},\{\{1\},\{2\}\},\{\{1\},\{1,2\}\},\{\{2\},\{1,2\}\},\\&\{\emptyset,\{1\},\{2\}\},\{\emptyset,\{1\},\{1,2\}\},\{\emptyset,\{2\},\{1,2\}\},\{\{1\},\{2\},\{1,2\}\},\\&\{\emptyset,\{1\},\{2\},\{1,2\}\}\\&\}\end{aligned}$$

是“给若干原子穿上两层集合的衣服”的所有可能情况（不要在意空集）.

经过以上恐怖的列举，你也许会有一种模模糊糊的想法：既然正规公理断言了“属于”关系的反自反性、反对称性和良基性，那么任意一个集合，无论它的结构有多复杂，把它的衣服层层脱去（可以用并集实现这个过程），最终都会得到原子. 这个想法很好，在理论上也能得到肯定，集合论是研究无穷的数学分支，之前讨论过的超穷递归也说明了无穷多次操作的可行性，所以我们允许“脱去”的次数为任意序数多次. 基于这个模糊的想法，我们定义层垒谱系(cumulative hierarchy).

定义　层垒谱系

层垒谱系由超穷递归的形式定义.

$$V_0 = \text{所有原子的集合}$$
$$V_{\alpha^+} = \mathcal{P}(V_\alpha)$$
$$V_\alpha = \bigcup_{\beta<\alpha} V_\beta$$

定义良基宇宙(well-founded universe) 或冯 · 诺伊曼宇宙(von Neumann universe)

$$\mathbb{WF} = \bigcup_{\alpha\in\mathbb{ON}} V_\alpha$$

你一定有一种强烈的直觉：既然任意集合中元素的“集合衣服的层数”至多超穷多，而这正是 $\mathbb{WF}$ 所允许的，那么所有的集合都应该属于 $\mathbb{WF}$，即 $\mathbb{V} = \mathbb{WF}$ 才对. 接下来我们就会发现这个结论是正确的，并证明一个更加优美的定理.

首先定义集合的秩(rank).

定义　集合的秩

对于任意 $x \in \mathbb{WF}$，如果 $x \subseteq V_\alpha$，而且对于任意 $\beta < \alpha$，都有 $x \not\subseteq V_\beta$，那么序数 α 称为 x 的秩(rank)，记作 $\operatorname{rank}(x) = \alpha$.

注意在 $\mathbb{WF}$ 中，$\operatorname{rank}(x)$ 是一定存在的. 这样定义集合的秩有一个优点：对于任意序数 α，都有 $\operatorname{rank}(\alpha) = \alpha$，这一点可以用超穷归纳法来证明，只要注意到 α^+ 和 $\mathcal{P}(\alpha)$ 的秩都恰好比 α 的秩多 1 即可.

这一小节的标题是“再次讨论正规公理”，这暗示了 $\mathbb{V} = \mathbb{WF}$ 能由正规公理导出，现在我们来证明这个命题.

证明　在承认正规公理的前提下，假设 $\mathbb{V} \neq \mathbb{WF}$，即某些集合没有秩，选取一个没有秩的集合，记作 c，接下来把目光聚焦在 $A = \{x \in \mathrm{TC}(\{c\}) | x\text{没有秩}\}$ 这个集合上. 根据正规公理，存在 $a \in A$ 使得 $a \cap A = \emptyset$，这时对于 a 的任一元素 x，立即有 $x \notin A$，然而由于 $\mathrm{TC}(\{c\})$ 是传递集，根据传递集的性质 $x \in \mathrm{TC}(\{c\})$，这说明 x 有秩. a 中的每个元素都有秩，根据秩的性质，得到 a 有秩，所以 $a \notin A$，和题设矛盾了. 因此这个没有秩的集合 c 是不存在的，每个集合都有秩，即 $\mathbb{V} = \mathbb{WF}$. □

然而当我们将充分必要条件反过来，可以发现也成立.

证明　如果 $\mathbb{V} = \mathbb{WF}$，即每个集合都有秩，那么对于任意集合 A，我们可以选出秩最小的元素 a，这时候对于任意集合 x，如果 $x \in a$，根据秩的性质，有 $\operatorname{rank}(x) < \operatorname{rank}(a)$，但由于 a 已经是秩最小的元素了，所以 x 不可能也是是 A 的元素，所以 $a \cap A = \emptyset$，这就导出了正规公理. □

经过以上推导，我们得到了这样一个结论：

定理　　正规公理与 $\mathbb{V} = \mathbb{WF}$ 等价，每个集合都有秩.

回忆传递集，传递集 A 是这样一种集合，它的结构足够特殊，以至于 $x \in A$ 和 $x \subset A$ 等价. 序数就是典型的传递集. 但对于任意集合，我们是否能构造出关于它的传递集呢？

定义 传递闭包

递归定义一个序列 $\{\mathrm{TC}_n(x)\}_{n\in\mathbb{N}}$：

$$\mathrm{TC}_0(x) = x$$
$$\mathrm{TC}_{n+1}(x) = \bigcup \mathrm{TC}_n(x)$$

而 x 的传递闭包(transitive closure) 就是该序列取并集：

$$\mathrm{TC}(x) = \bigcup_{n\in\mathbb{N}} \mathrm{TC}_n(x)$$

$\mathrm{TC}(A)$ 一定是包含 A 的最小传递集. 这句话包含了四个命题：对于任意集合 A，$\mathrm{TC}(A)$ 一定存在；$A \subseteq \mathrm{TC}(A)$；$\mathrm{TC}(A)$ 是传递集；只要 B 是传递集且 $A \subseteq B$，则 $\mathrm{TC}(A) \subseteq B$.

证明 首先证明存在性. A 是一个集合，那么 $\mathrm{TC}_0(A)$ 也是集合，根据并集公理，$\mathrm{TC}_{n+1}(A)$ 也是集合，再根据数学归纳定理，所有的 $\mathrm{TC}_n(A)$ 都是集合，再根据并集公理即可证明 $\mathrm{TC}(A)$ 是集合.

第二个命题只要注意到 $A = \mathrm{TC}_0(A) \subset \mathrm{TC}(A)$ 即可.

至于传递性，首先任选 $x \in \mathrm{TC}(A)$，根据并集的定义，必然存在 n 使得 $x \in \mathrm{TC}_n(x)$，即 $x \subseteq \mathrm{TC}_{n+1}(A)$. 此时对于任意 $y \in x$，必然有 $y \in \mathrm{TC}_{n+1}(A)$，根据并集定义，$y \in \mathrm{TC}(A)$，所以 $\mathrm{TC}(A)$ 是传递集.

第四个命题比较有意思，接下来用数学归纳法证明. 首先 $\mathrm{TC}_0(A) = A \subseteq B$；只要 $\mathrm{TC}_n(A) \subseteq B$，就有 $\mathrm{TC}_{n+1}(A) = \bigcup \mathrm{TC}_n(A) \subseteq \bigcup B$，由于 B 是传递集，有 $\bigcup B \subseteq B$，所以 $\mathrm{TC}_{n+1}(A) \subseteq B$；因此所有的 $\mathrm{TC}_n(A) \subseteq B$，所以 $\mathrm{TC}(A) \subseteq B$. □

10.2 推广的归纳和递归

11 选择公理的另一面

回顾选择公理，它断言每个非空集族都有选择函数. 具体说来，非空集族指的是集合 $X = \{x_i\}_{i\in I}$，其中每个 x_i 都是非空的. 而它的选择函数指的是函数 $f : X \to \bigcup X$，使得 $f(x_i) \in x_i$. 在上文已经使用逛商场买衣服的例子通俗介绍了选择公理. 就像很晦涩的正规公理都能推出 $\mathbb{V} = \mathbb{WF}$ 这个优美的等价命题一样，我们也希望选择公理能推出更优美的等价命题.

数学家们经常把选择公理单独拿出来，不让它和 ZF 的其他公理放一起讨论，是因为选择公理似乎并没有其他公理那么“自然”. 选择公理断言了某个集合（选择函数）存在，却没有告诉我们该集合（选择函数）如何构造，这是与 ZFC 其他公理相异的一点. 这种“管杀不管埋”的特点让很多数学家不喜欢它. 不过就像数学分析中的中值定理一样，“断言某事物存在但不明确构造”的命题也是很有研究价值的.

11.1 良序定理

选择公理是策梅洛首先提出来的，不过当时提出的是选择公理的等价命题：“任意集合都可以被良序化”. 不少集合存在“自然的”良序，例如自然数的大小关系就是 $\mathbb{N}$ 上的良序. 对于 $\mathbb{Z}$，我们可以规定任意两个整数 a, b，若 $|a| < |b|$ 则 $a \prec b$，若 $|a| = |b|$ 且 $a < 0$ 且 $b > 0$ 则 $a \prec b$，此时“$\prec$”就是 $\mathbb{Z}$ 上的良序. 而对于 $\mathbb{R}$ 来说就没那么简单了，不但你想不出怎样构造良序，数学家们也想不出.

但是有了选择公理，情况就不一样了. 联想选择函数的定义，对于任意一个集合 A，如果存在某个函数使得 $f(A) \in A$，我们可以记 $a_0 = f(A)$，然后将此元素取出，剩下的集合就是是 $A - \{a_0\}$，如果这时候这个选择函数还能发挥作用，即 $f(A - \{a_1\}) \in A - \{a_0\}$，那么我们又可以记 $a_1 = f(A - \{a_1\})$，将此元素拿走，剩下的集合就是 $A - \{a_0, a_1\}$. 如此循环往复下去直至超穷次，所有元素都会被取光，然后取出来

的元素就组成了一条序列 $\{a_\xi\}_{\xi<\theta}$，借助序数的大小比较法则我们可以规定 $a_\xi \prec a_\eta \leftrightarrow \xi < \eta$，此时“$\prec$”就是 A 上的良序.

为了避免“取出超穷多次”这种定义模糊的语句，接下来给出完整证明过程.

证明 令 S 为 A 的非空子集组成的集合（这个 S 存在性是毫无疑问的，因为 $S \subset \mathcal{P}(A)$，由幂集公理和子集公理模式断言存在），根据选择公理，S 上存在选择函数 f 使得对于任意 $X \in S$ 有 $f(X) \in X$. 接下来令

$$a_\xi = f(A - \{a_\eta | \eta < \xi\}) \qquad (A - \{a_\eta | \eta < \xi\} \neq \emptyset)$$

由此得到一个序列 $\{a_\xi\}_{\xi<\theta}$ 枚举了 A 的所有元素，此时序列的先后顺序即为 A 的良序. □

刚才我们为了把 A 排好序，构造出了一个集合 S 包含 A 的所有非空子集，然后选择公理断言了 S 有选择函数. 现在反过来，如果 A 的内部已经被良序关系“$\prec$”排好了，那么对于 A 的每个非空子集 X，其内部也被“$\prec$”排好了，并且容易证明这个“$\prec$”也是 X 上的良序——也就是说有最小元 $\min(X)$. 此时如果定义一个函数 f，满足 $f(X) = \min(X)$，咦，这不就是 S 上的选择函数嘛，这不就推出了选择公理嘛. 现在我们把这个过程严格地表述一遍.

证明 设 S 为任意非空集族，根据并集公理，$\bigcup S$ 存在. 根据良序定理，存在序关系“$\prec$”使得 $\langle \bigcup S; \prec \rangle$ 为良序集. 对于 $X \in S$，有 $X \subseteq \bigcup S$，此时“$\prec\restriction X$”是 X 上的良序关系，定义函数 $f : S \to \bigcup S$ 满足 $f(X) = \min(X)$，那么 f 即为 S 的选择函数. □

> **定理 良序定理 (well-ordering theorem)**
>
> 对于任意集合 A，均存在序关系“$\prec$”使得 $\langle A; \prec \rangle$ 为良序集.
> 良序定理是选择公理的等价命题.

11.2 佐恩引理

直接上命题.

(1) <u>选择公理</u>(axiom of choice, AC)：非空集组均有选择函数.

(2) 两个非空集合的直积非空.

证明 说得更明白一些，若 $\{x_i\}_{i\in I}$ 是非空集族，则 $\prod_{i\in I} x_i \neq \emptyset$. 证明思路很简单，考察 $\prod_{i\in I} x_i$ 的元素，都是有序组 $(a_i)_{i\in I}$，而且 $a_i \in x_i$. 这个有序组恰好枚举了某个选择函数的值域！由此确定的选择函数就是 $f = \{(i, a_i) | i \in I\}$. 有序组和选择函数一一对应，因此只要直积非空，选择函数就存在，反之亦然. □

(3) <u>良序定理</u>(well-ordering theorem, by Zermelo)：任意集合均可建立良序.

证明 要想建立 A 上的良序，只要构造一个序列 $\{x_\xi\}_{\xi<\alpha}$ 枚举 A，然后定义二元关系 $\leq$: $x_\alpha \leq x_\beta \Leftrightarrow \alpha \leq \beta$，此即为 A 上的良序关系，因此重点是映射 $f : \xi \mapsto x_\xi$ 的构造.

令 $P = \mathcal{P}(A) - \{\emptyset\}$，即 A 的非空子集的集合. 该映射就是 $x_\xi = f(A - \{a_\eta\}_{\eta<\xi})$ □

(4) <u>佐恩引理</u>(Zorn's Lemma)：若偏序集 $\langle A; \leq \rangle$ 所有连通的子集都有上界，那么 A 有最大元.

(5) 对于任意两个集合，存在一个映射，以一方为定义域，并射满另一方.

11.3 基数运算的遗留问题

12 自然数、整数和有理数的构造

12.1 自然数的性质

历史上刻画自然数大多从皮亚诺公理开始，但如之前所述，自然数就是有限的序数，无穷公理断言了它们的存在性，序数运算法则把加法、乘法和乘方定义好了，集合论从更一般的角度刻画了自然数，因此我们不必再引入皮亚诺公理，而可以直接讨论自然数及其运算的性质.

(1) 加法交换律：$a+b=b+a$.

(2) 加法结合律：$(a+b)+c=a+(b+c)$.

(3) 加法消去律：若 $a+c=b+c$，则 $a=b$，反之成立.

(4) 加法保序性：若 $a+c>b+c$，则 $a>b$，反之成立.

(5) 乘法交换律：$a\cdot b=b\cdot a$.

(6) 乘法结合律：$(a\cdot b)\cdot c=a\cdot(b\cdot c)$.

(7) 乘法对加法的分配律：$(a+b)\cdot c=a\cdot c+b\cdot c$.

(8) 乘法消去律：若 $a\cdot c=b\cdot c$ 且 $c\neq 0$，则 $a=b$，反之成立.

(9) 乘法保序性：若 $a\cdot c>b\cdot c$ 且 $c\neq 0$，则 $a>b$，反之成立.

(10) 欧几里得定理(Euclidean theorem)：设 $n\geq 0$ 且 $q>0$，则存在 $m\geq 0$ 和 $0\leq r<q$ 使得 $n=mq+r$.

(11) 0 是零元：若 $a\cdot b=0$，则 $a=0$ 或 $b=0$，反之成立.

12.2 整数和有理数的构造

这一节我们要构造整数集 $\mathbb{Z}$ 和有理数集 $\mathbb{Q}$. 这里有一个非常重要的问题，在初等数学中 1 既是自然数，又是整数，也是实数和复数，无论是什么数，$1=1$ 都是成立的. 然而在集合论中，作为自然数的 1 和作为整数的 1 是两个不同的集合，后者甚至是一个无穷集合[4]，为了区分这几者，作为自然数的 1 写作 $1_{\mathbb{N}}$，整数则为 $1_{\mathbb{Z}}$，其它数字类似，这种表示方法将在定义完成之后抛弃.

之所以会出现整数，是因为减法运算把自然数捅出了漏洞，为了使减法封闭，我们引入了负数，并把负数和自然数统称为整数. 这话说得轻巧，想要严格地定义构造出整数集，工作量比想象中的要大一些（尽管这些工作的细节大多都是“为满足我们心目中的整数所具有的性质不得不做的矫揉造作的体力活”）. 在初等数学的语境中，随便拿出一个整数，例如 -1，有 $-1=0-1=1-2=2-3=\cdots$，可以说这些式子涉及的可以都是自然数，但它刻画了 -1 这个负数. 回到集合论语境，将这些减数和被减数依次放入有序对中，得到 $(0_{\mathbb{N}},1_{\mathbb{N}}),(1_{\mathbb{N}},2_{\mathbb{N}}),(2_{\mathbb{N}},3_{\mathbb{N}}),\cdots$，我们可以将这些有序对的集合作为整数的集合论构造.

对于同一个整数，任意选两个有序对 $(m_{\mathbb{N}},n_{\mathbb{N}}),(p_{\mathbb{N}},q_{\mathbb{N}})$，根据定义有 $m_{\mathbb{N}}-n_{\mathbb{N}}=p_{\mathbb{N}}-q_{\mathbb{N}}$，即 $m_{\mathbb{N}}+q_{\mathbb{N}}=n_{\mathbb{N}}+p_{\mathbb{N}}$. 我们可以由此定义出一个等价关系.

---

[4]这一点使得初等数学的底层细节多了很多不必要的细节，提高了研究难度，所以某些数学家对集合论构造数系嗤之以鼻，更倾向于使用公理规定它们的性质.

定义 整数

定义 $\mathbb{N}^2$ 上的等价关系 "$\equiv$":

$$(m_{\mathbb{N}}, n_{\mathbb{N}}) \equiv (p_{\mathbb{N}}, q_{\mathbb{N}}) \Leftrightarrow m_{\mathbb{N}} + q_{\mathbb{N}} = n_{\mathbb{N}} + p_{\mathbb{N}}$$

整数(integer) 是 "$\equiv$" 的一个等价类，整数集是 $\mathbb{N}^2$ 中所有的等价类的集合，即

$$\mathbb{Z} := \mathbb{N}^2 / \equiv$$

- 对于整数 $A = [(m_{\mathbb{N}}, n_{\mathbb{N}})]_{\equiv}, B = [(p_{\mathbb{N}}, q_{\mathbb{N}})]_{\equiv}$，定义其加法、乘法、相反数(negation) 和减法(subtraction) 为

$$\begin{aligned} A + B &:= [(m_{\mathbb{N}} + p_{\mathbb{N}}, n_{\mathbb{N}} + q_{\mathbb{N}})]_{\equiv} \\ A \cdot B &:= [(m_{\mathbb{N}} \cdot p_{\mathbb{N}} + n_{\mathbb{N}} \cdot q_{\mathbb{N}}, m_{\mathbb{N}} \cdot q_{\mathbb{N}} + n_{\mathbb{N}} \cdot p_{\mathbb{N}})]_{\equiv} \\ -A &:= [(n_{\mathbb{N}}, m_{\mathbb{N}})]_{\equiv} \\ A - B &:= A + (-B) \end{aligned}$$

- $\mathbb{N}$ 可作为 $\mathbb{Z}$ 的子结构，其同态映射 μ 为

$$\mu(n) := [(n_{\mathbb{N}}, 0_{\mathbb{N}})]_{\equiv}$$

  $\mu(n)$ 姑且称为 "在整数意义下的自然数 n"，今后不再区分 n 和 $\mu(n)$，除非讨论其集合论意义的差别.

- 对于整数 p, q，若存在自然数 n，使得 $p + n = q$，则记 $p \leq q$，若 $n \neq 0$，则记 $p < q$.

整数的加法同样是交换、结合、消去、保序的，乘法也是交换、结合、消去、分配的，但相对于自然数，整数有了以下这些性质:

- 整数的乘法不保序，进一步地，若 $c < 0$ 且 $a > b$，则 $ac < bc$.
- 减法在整数集中是封闭的，即若 $a, b \in \mathbb{Z}$，则 $a - b \in \mathbb{Z}$.
- 自然数和自然数的相反数恰好构成了整数，即对于一个整数 p，一定存在一个自然数 n，$p = n$、$p = -n$ 两者至少成立其一；进一步地，当两者都成立时，$p = 0$.
- 序关系 "$<$" 在 $\mathbb{Z}$ 中是全序但不是良序，即对于任意整数 n，都存在另外的整数 p, q，使得 $p < n < q$.

有理数集的构造类似，我们之所以对整数不够满意，是因为除法可以把整数集捅出漏洞. 在初等数学中，能够写成两个整数相除形式的都是有理数，反之成立. 然而对于同一个有理数，存在无穷多对整数相除得到，例如 $1/2 = 2/4 = 4/8 = \cdots$. 因此我们可以像构造整数那样构造有理数，即定义一个等价类，当 $m/n = p/q$ 即 $mq = pn$ 时属于同一类.

定义 有理数

定义 $\mathbb{Z}^2$ 上的等价关系 “$\equiv$”:

$$(m_\mathbb{Z}, n_\mathbb{Z}) \equiv (p_\mathbb{Z}, q_\mathbb{Z}) \Leftrightarrow m_\mathbb{Z} \cdot q_\mathbb{Z} = n_\mathbb{Z} \cdot p_\mathbb{Z}$$

有理数(rational number) 是 “$\equiv$” 的一个等价类，有理数集是 $\mathbb{Z}^2$ 中所有等价类的集合，即

$$\mathbb{Q} := \mathbb{Z}^2/\equiv$$

- 对于有理数 $A = [(m_\mathbb{Z}, n_\mathbb{Z})]_\equiv, B = [(p_\mathbb{Z}, q_\mathbb{Z})]_\equiv$，定义有理数加法、乘法、相反数、减法、倒数(reciprocal) 和除法(division) 为

$$\begin{aligned}
A + B &:= [(m_\mathbb{Z} \cdot q_\mathbb{Z} + p_\mathbb{Z} \cdot n_\mathbb{Z}, n_\mathbb{Z} \cdot q_\mathbb{Z})]_\equiv \\
A \cdot B &:= [(m_\mathbb{Z} \cdot p_\mathbb{Z}, n_\mathbb{N} \cdot q_\mathbb{N}]_\equiv \\
-A &:= [(-m_\mathbb{Z}, n_\mathbb{Z})]_\equiv \\
A - B &:= A + (-B) \\
A^{-1} &:= [(n_\mathbb{Z}, m_\mathbb{Z})]_\equiv \\
A/B &:= A \cdot B^{-1}
\end{aligned}$$

- $\mathbb{Z}$ 可作为 $\mathbb{Q}$ 的子结构，其同态映射 μ 为

$$\mu(n) := [(n_\mathbb{Z}, 1_\mathbb{Z})]_\equiv$$

  $\mu(n)$ 姑且称为 “在有理数意义下的整数 n”，今后不再区分 $n_\mathbb{Z}$ 和 $\mu(n)_\mathbb{Q}$，除非讨论其集合论意义的差别.

- 若存在 $m_\mathbb{Z}, n_\mathbb{Z} > 0$ 使得 $p_\mathbb{Q} = m_\mathbb{Z}/n_\mathbb{Z}$，则称 p 是正有理数，其相反数是负有理数. 对于两个有理数 p, q，若存在正有理数 n 使得 $p + n = q$，则记 $p < q$.

相比于整数，有理数系多了一些性质

(1) $<$ 是 $\mathbb{Q}$ 上的稠密序关系，即任意两个不同的有理数之间都存在其他的有理数.

证明 设 p, q 是有理数且 $p \neq q, q \neq 0$. 考察表达式 $\frac{p+q}{2}$. 首先这是有理数. 不妨设 $p < q$，那么 $p = \frac{p+p}{2} < \frac{p+q}{2} < \frac{q+q}{2} = q$，由此体现 “$<$” 是稠密的. □

(2) $x \cdot x^{-1} = 1$，1 是有理数中的乘法逆元.

13 标准实数集的结构

这里强调接下来构造的实数集是 “标准” 实数集，即构造的实数集遵从牛顿/莱布尼兹-柯西-魏尔斯特拉斯的路线，其拓扑结构采用标准拓扑，有别于鲁滨逊非标准分析的构造，此时**假设有理数集已经构造好了**，而每个有理数都形如 p/q（p, q 为整数且 $q \neq 0$）. 从有理数构造实数有几种方法，例如柯西数列、戴德金分割和公理化定义，接下来使用柯西数列构造，因为完整地定义实数的十进制表示法需要用到柯西数列.

13.1 从有理数构造实数

有理数集 $\mathbb{Q}$ 上的大小关系（序关系）是稠密的，这是自然数集 $\mathbb{N}$ 和整数集 $\mathbb{Z}$ 都不具备的性质.

定义 稠密性和完备性

设 $\langle A;<\rangle$ 为全序集，
若对于任意两个元素 a,b，只要 $a<b$，就存在 $c\in A$，使得 $a<c<b$，则称该序关系是稠密的(dense)；
若 $D\subseteq A$，对于任意 $a,b\in A$ 且 $a<b$，都存在 $d\in D$ 使得 $a<d<b$，则称 D 是 A 的稠密的子集.
若 A 的有界子集都有确界，则称 $\langle A;<\rangle$ 是序完备的(order-complete).

有理数集上的序关系就是稠密的，也就是说任取两个有理数，无论它们相差多小，只要不相等，都可以在它们之间找到有理数，它们的算术平均就是一个例子. 接下来任取两个不等的有理数 a,b，然后取它们之间的有理数 b_1，再取 a 和 b_1 之间的有理数 b_2，再取 a 和 b_2 之间的有理数 $b_3\cdots\cdots$如此操作下去，有理数列 $\{b_n\}$ 必然会收敛于某一点（柯西审敛原理），而这一点不一定是有理数，比如构造两个有理数列 $\{a_n\}_{n=0}^{\infty}$ 和 $\{b_n\}_{n=0}^{\infty}$：

$$a_n=\frac{k}{10^n},\quad \left(\frac{k}{10^n}\right)^2<2<\left(\frac{k+1}{10^n}\right)^2,\quad k\in\mathbb{N}$$
$$b_n=\frac{k+1}{10^n},\quad \left(\frac{k}{10^n}\right)^2<2<\left(\frac{k+1}{10^n}\right)^2,\quad k\in\mathbb{N}$$

两个数列一个有界严格递增，一个有界严格递减，可以列出前几项：

$$\begin{array}{llll}
a_0=1 & b_1=2 & a_1^2=1 & b_1^2=4;\\
a_1=1.4 & b_2=1.5 & a_2^2=1.96 & b_2^2=2.25;\\
a_2=1.41 & b_3=1.42 & a_3^2=1.9881 & b_3^2=2.0164;\\
a_3=1.414 & b_4=1.415 & a_4^2=1.99396 & b_4^2=2.002225;\\
a_4=1.4142 & b_5=1.4143 & a_5^2=1.99996164 & b_5^2=2.00024449;\\
\cdots & \cdots & \cdots & \cdots
\end{array}$$

随着 n 的增加，a_n^2 和 b_n^2 都越来越接近 2，而且精确程度可以任意高，但 a_n^2 总是不足的，b_n^2 总是过剩的，因此数列 $\{b_n\}$ 中的每一项都可以作为数列 $\{a_n\}$ 的上界. 然而：

- 不存在 $x\in\mathbb{Q}$，使得 $x^2=2$.

证明 *反证法，若存在 $x\in\mathbb{Q}$ 使得 $x^2=2$，那么可以找到两个互质的整数 p,q 使得 $x=p/q$，此时有 $2=p^2/q^2$，即 $p^2=2q^2$. 因为 q^2 是整数，所以 p^2 为偶数，数论进一步告诉我们，p 也为偶数，因此可以设 $p=2k$，其中 k 也是整数. 根据互质的要求，q 不能为偶数. 然而 $q^2=p^2/2=(2k)^2/2=2k^2$，这意味着 q^2 也是偶数，所以 q 也是偶数，矛盾了. 因此这样的 x 是不存在的.* □

我们考察一下 $\sup\{a_n\}$，肯定不是任意一个 b_n，因为 $\{b_n\}$ 递减但没有最小值，如果哪个 b_n 敢说自己是 $\sup\{a_n\}$，b_{n+1} 肯定不服，所以只能是那个神秘的 $\sqrt{2}$ 了，这意味着 $\sup\{a_n\}\notin\mathbb{Q}$，有理数集被 $\sup,\inf$ 运算捅出了漏洞.

以上构造的两个数列有非常明显的特征，它们不一定需要后一项与前一项之差越来越小，但必须控制在一个范围内，更近一步说，任意两项之差都必须控制在一个范围内，这一类数列就是柯西数列.

定义 柯西数列

在度量空间 (S,d) 中定义序列 $\{a_n\}_{n\in\mathbb{N}}$, 若对于任意的 $\varepsilon\in\mathbb{Q}^+$, 都存在正整数 N, 使得任意 $m,n>N$, 都有 $d(a_m,a_n)<\varepsilon$, 则称该序列为柯西序列(Cauchy sequence).
用有理数域上的语言来叙述：对于有理数列 $\{a_n\}_{n=0}^{\infty}$, 若对于任意的 $\varepsilon\in\mathbb{Q}^+$, 都存在正整数 N, 使得任意 $m,n>N$, 都有 $|a_m-a_n|<\varepsilon$, 则称数列 $\{a_n\}$ 为有理数域上的柯西数列.

需要强调，以上构造的两个数列定义在有理数的范围内，现在也可以说，有理数集被柯西数列捅出了漏洞. 构造实数的关键一步在于把这些漏洞补上，使得无论怎么构造柯西数列，都不会收敛到集合外面，即所谓柯西完备性(Cauchy-complete). 在将要构造的实数域内，序完备和柯西完备是等价的.

尽管 $\{a_n\}$ 和 $\{b_n\}$ 不收敛于有理数，但随着 n 的增大，它们的差距越来越小，用有理数的 $\varepsilon-N$ 语言来说：$\forall\varepsilon>0$, $\exists N\in\mathbb{N}$, 当 $n>N$ 时，有 $|a_n-b_n|<\varepsilon$. 直观理解，尽管它们不收敛到有理数，但似乎收敛于同一个目标，这个“目标”可能是有理数，也可能不是，但至少可以知道，这些“目标”组成的集合可以使得有理数的大小关系扩充为一个完备的序关系，这样的“目标”，在直觉上可以理解为实数，但严谨起见，我们不能把一个模糊而抽象的“目标”当做被定义的对象，所以干脆用等价类本身算了，因此其大小关系和相等关系也需要重新定义.

定义 实数

定义 $\mathbb{Q}$ 上柯西数列 $\{a_n\}_{n=0}^{\infty}$ 和 $\{a_n\}_{n=0}^{\infty}$ 之间的等价关系：

$$\{a_n\}_{n\in\mathbb{N}}\equiv\{b_n\}_{n\in\mathbb{N}}\Leftrightarrow(\forall\varepsilon\in\mathbb{Q}^+)(\exists N\in\mathbb{N})(\forall n\in\mathbb{N}^+)(n>N\to|a_n-b_n|<\varepsilon)$$

实数(real number) 是“$\equiv$”的等价类，实数集是 $\mathbb{Q}^\omega$ 中所有等价类的集合，即

$$\mathbb{R}:=\mathbb{Q}^\omega/\equiv$$

对于两个实数 a,b, 从其中分别任意抽出一个柯西数列 $\{a_n\}$ 和 $\{b_n\}$, 若存在 N, 使得对于任意 $n>N$, 都有 $a_n\geq b_n$, 则记 $a\geq b$.
对于两个实数 a,b, 从其中分别任意抽出一个柯西数列 $\{a_n\}$ 和 $\{b_n\}$, 若 $\{a_n\}\equiv\{b_n\}$, 则记 $a=b$.

13.2 实数的十进制表示

接下来要解决实数的十进制表示问题. 用十进制表示实数是一个很显然的过程，但完整地解释起来并不是那么简单，尤其是出现省略号的时候. 为什么 $12.333\cdots$ 是实数的合法表达，而 $\cdots 333.21$ 却不是？为什么 00123 和 123 表示的是同一个实数，而 123 和 12300 却不是？我们是否能跨越无穷多位小数，更改“最后”几位数，捏造出 $12.333\cdots 334$ 这样的数出来，并认为它有别于 $12.333\cdots$? 以及贴吧的经验密码，为什么 $0.999\cdots$ 和 1 是同一个实数？接下来依次定义整数的十进制表示、有限小数的十进制表示和无限小数的十进制表示来解决这个问题.

定义 位

实数十进制表示的位(digit) 是一个自然数，而且只能取 $0,1,2,3,4,5,6,7,8,9$ 之一，再定义自然数的符号 拾 表示 9 的后继数.

这里使用符号“拾”而不是 10 是因为现在尚未定义出十进制表示法，使用 10 会导致逻辑循环. 现在使用一个有限长的由位组成的数列来表示整数部分.

定义　整数的十进制表示

对于一个由位组成的数列 $\{a_i\}_{i=0}^{n}$，定义字符串

$$\pm a_n a_{n-1} a_{n-2} \cdots a_1 a_0 := \pm \sum_{i=0}^{n} a_i \times 拾^i$$

称为整数的十进制表示；

若 $a_n \neq 0$，则该字符串又称为整数的约化十进制表示.

整数存在唯一的约化十进制表示，这点可用带余除法证明；而十进制表示可以有任意多个，因为如果 $a_n a_{n-1} \cdots a_1 a_0$ 是整数的十进制表示，那么 $0 a_n a_{n-1} \cdots a_1 a_0$ 也是. 因此现在可以说 00123 和 123 表示同一个整数，而 123 和 12300 表示两个不同整数. 此外 $10 = 1 \times 拾^1 + 0 = 拾$，因此现在可以用 10 来代替字符“拾”了.

定义　有限小数的十进制表示

对于一个由位组成的数列 $\{a_i\}_{i=-m}^{n}$，定义字符串

$$\pm a_n a_{n-1} a_{n-2} \cdots a_1 a_0 . a_{-1} a_{-2} \cdots a_{-m} := \pm \sum_{i=-m}^{n} a_i \times 10^i$$

称为有限小数的十进制表示；

若 $a_n \neq 0$ 且 $a_{-m} \neq 0$，则该字符串又称为有限小数的约化十进制表示.

同样地，有限小数也存在唯一的约化十进制表示，但有任意多个十进制表示.

从有限到无限的拓展必须十分小心，无穷多个数的和必须要另外定义[5].

定义　无限小数的十进制表示

对于一个由位组成的数列 $\{a_i\}_{i=-\infty}^{n}$，定义字符串

$$\pm a_n a_{n-1} a_{n-2} \cdots a_1 a_0 . a_{-1} a_{-2} \cdots := \pm \sum_{i=-m}^{n} a_i \times 10^i$$

称为无限小数的十进制表示；

若 $a_n \neq 0$，则该字符串又称为无限小数的约化十进制表示.

无限小数的约化十进制表示并不是唯一的，但每一个实数都至少有一个约化十进制表示.

命题　百度贴吧经验密码

在标准实数系下，

$$0.999 \cdots = 1$$

[5]对于数列 $\{a_n\}_{n=0}^{\infty}$，若对于任意 $\varepsilon > 0$，都存在正整数 N，使得当 $n > N$ 时，有 $\left|\sum_{i=0}^{n} a_i - L\right| < \varepsilon$，则称该级数是收敛的，且收敛于 L，写作

$$\sum_{n=0}^{\infty} a_n = L$$

此为级数的柯西和，当无穷出现在求和号下方时定义类似.

证明：这里需要说明，$0.999\cdots$ 中省略号省略了无穷多位 9. 因此根据定义，

$$0.999\cdots := \sum_{n=-\infty}^{-1} 9\times 10^{n} = \sum_{n=1}^{\infty} 9\times 10^{-n}$$

任意给定 $\varepsilon>0$，取 $N=-\lceil \lg\varepsilon\rceil$，当 $n>N$ 时，根据等比数列求和公式

$$\left|1-\left(\sum_{i=1}^{n} 9\times 10^{-i}\right)\right| = \left|1-\frac{9\times 10^{-1}(1-10^{-i})}{1-10^{-1}}\right| = 10^{-n} < 10^{-N} < \varepsilon$$

即该级数的柯西和收敛于 1，因此 $0.999\cdots=1$.

13.3 可数与不可数

(4) $\mathrm{card}(\mathbb{R})=\mathrm{card}(\mathcal{P}(\mathbb{N}))$

我们不直接证明原命题，而证 $\mathrm{card}(\mathbb{R})=\mathrm{card}\left(\{0,1\}^{\mathbb{N}}\right)$，原命题即可根据上个性质得证.

根据双射 $f(x)=\tan\left(\frac{\pi}{2}x-\frac{1}{2}\right)$，易知 $\mathrm{card}(0,1)=\mathrm{card}(\mathbb{R})$，而 $(0,1)$ 中的每个实数都可以用二进制小数 $0.a_1a_2a_3\cdots\quad(a_i=0,1)$ 来表示，令 $S=\{i|a_i=1\}$，则 $\chi_S(n)$ 便是一个根据该小数构造出来的示性函数. 和上一个性质证明同样的思路，易得 $\mathrm{card}(\mathbb{R})=\mathrm{card}\left(\{0,1\}^{\mathbb{N}}\right)$. 根据这个结果，$n$ 维空间中的点、连续几何曲线上的点、无理数集、有理柯西数列的集合都不可数，并与 $\mathbb{R}$ 等势.

实数不可数最初由康托尔 (Georg Cantor) 通过对角线证法得出，这个证明发表的那一天标志着集合论的诞生，尽管以今天的视角来看，当时的证明有些瑕疵（无限小数 p 进制表示的不唯一性），但这些瑕疵都很好解决，比如约定禁止 9 循环出现，或者用连分数展开的线性表示式来代替无限小数等.

13.4 实数完备性定理

初学数学分析时，书上都会提到所谓“关于实数完备性的六大基本定理”，当然不一定是“六大”. 不少教材只是讨论了这几大定理如何相互推导，由此刻画了实数的完备性，却没有说清楚“实数”为何物. 现在说清楚其定义了，接下来证明就显得较为“自然”. 可以用以下几个命题描述实数完备性：

- 柯西完备性(Cauchy completeness)
  $\mathbb{R}$ 上的数列 $\{a_n\}_{n=0}^{\infty}$ 收敛的充要条件是对于任意实数 $\varepsilon>0$，存在 N 使得 $m,n>N$ 时，$|a_n-a_m|<\varepsilon$.
  这点是显然的，因为每一个实数都是有理数的柯西数列.
- 康托尔交集定理(Cantor intersection theorem)
  对于一系列闭区间 $\{[a_n,b_n]\}_{n=0}^{\infty}$，若 $[a_n,b_n]\supseteq[a_{n+1},b_{n+1}]$，而且 $\lim\limits_{n\to\infty} a_n-b_n=0$，则存在唯一实数 L，使得 $\lim\limits_{n\to\infty} a_n=\lim\limits_{n\to\infty} b_n=L$.
- 戴德金完备性(Dedekind completeness)
  把 $\mathbb{R}$ 的元素分到两个集合 A,B 中，并使 A 中的任意数小于 B 中的任意数，那么一定存在实数 c，对于 A 中的元素 a 和 b 中的元素 b，都满足 $a\le c\le b$.
- 海涅-波莱尔定理(Heine-Borel theorem)
  如果有一系列开区间 $S=\{(a_\xi,b_\xi)\}_{\xi<\alpha}$ 覆盖了一个闭区间 $[a,b]$，即 $[a,b]\subseteq\bigcup S$，那么可以从这个系列中选出有限个开区间覆盖这个闭区间，即存在 S 的有限子集 S' 使得 $[a,b]\subseteq\bigcup S'$
- 波尔查诺-魏尔斯特拉斯定理(Bolzano-Weierstrass theorem)
  $\mathbb{R}$ 有界数列必有收敛子列. 即对于任意数列 $\{a_n\}_{n=1}^{\infty}$，如果存在某个常数 M 使数列每一项 $|a_n|<M$，

- 介值定理(intermediate value theorem)
  $\mathbb{R}$ 上定义的连续函数，若能取到两个不同的值 a,b，则可以取到闭区间 $[a,b]$ 内的所有实数.

- 苏斯林性质(Suslin property)
  若有一系列开区间两两不相交，则这些开区间至多是可数的.

13.5 关于实数认识的局限性

实数集是不可数无穷的，它是比自然数的可数无穷更高阶的无穷，人们对实数的认识目前还有很多局限性，关于实数性质仍有很多奇妙的地方.

之前的讨论确认了 $\operatorname{card}(\mathbb{N}) < \operatorname{card}(\mathbb{R})$，于是自然有这样的问题：是否存在一个集合 S，使得 $\operatorname{card}(\mathbb{N}) < \operatorname{card}(S) < \operatorname{card}(\mathbb{R})$ 呢？连续统假设给出了否定回答.

命题　连续统假设 (continuum hypothesis)

不存在集合 S，使得 $\operatorname{card}(\mathbb{N}) < \operatorname{card}(S) < \operatorname{card}(\mathbb{R})$.

连续统假设尚未被证明或证伪，而且研究还发现，这个问题实在太过于特殊，现有的公理体系以及将来进行的一切构造性拓展都不足以证明或证伪它，除非承认一些纲领性的命题；甚至在某些哲学流派看来，连续统假设甚至都不是一个定义良好的问题. 目前学界倾向于否定它.

实数集以及定义在其上的四则运算和大小关系构成一个完备的有序域，其重要特征是 (1) 无最大或最小的元素；(2) 稠密性；(3) 完备性；(4) 苏斯林性质. 符合这四个条件的完备有序域不止有 $\mathbb{R}$，还可以构造出很多. 问题来了，这四条性质是否足够刻画实数？对于另一个符合这四个特征的完备有序域 S，是否一定存在一个双射 $f: S \to \mathbb{R}$，使得若 $a \le b$，则 $f(a) \le f(b)$ 呢？苏斯林假设给出了肯定回答.

命题　苏斯林假设 (Suslin hypothesis)

满足这四个特征的完备有序域一定存在到实数集上的保序映射.

苏斯林假设也尚未被证明或证伪，目前的研究结果是，如果把连续统假设当成公理加入现有的公理体系，再放入一些构造性的命题，就可以导出苏斯林假设. 如果苏斯林假设成立，就说明不必借助代数结构刻画实数.

无论是算法还是普通的描述性语句，它们都是有限长的符号序列，而前者是后者的一种特殊情况. 所有算法组成的集合和所有精确描述一个实数的语句组成的集合都是可数无穷多个的，面对不可数无穷多个的实数，必然有一些是无法计算和无法描述的，它们是比一般的无理数更加“无理”的数. 从有理数扩张到实数可以依次经过以下这几个具有里程碑意义的集合：

(1) 代数数 (algebric number). 代数数是整系数高次多项式方程的实根，有限次使用四则运算和乘方开方得到的数都是代数数，比如常见的 $\sqrt{2}$、黄金分割比 $\dfrac{\sqrt{5}-1}{2}$ 等都是代数数. 全体代数数集合的势为 $\operatorname{card}(\mathbb{N})$，代数数在实数内的补集称为超越数 (transcendental number). 我们接触到的具体的数大多都是代数数，不过超越数也很容易借助 e 得到：

定理　林德曼-魏尔斯特拉斯定理 (Lindermann-Weierstrass theorem)

对于非零代数数 $a_1, a_2, \cdots, a_n$ 和代数数 $b_1, b_2, \cdots, b_n$，均有

$$a_1 e^{b_1} + a_2 e^{b_2} + \cdots + a_n e^{b_n} \neq 0$$

以下定理则更进一步，说明超越数可以用两个代数数生成.

定理 格尔丰德-施耐德定理 (Gelfond–Schneider theorem)

若 a 是不为 $0, 1$ 的代数数，b 是无理数，则 a^b 是超越数.

这两个定理直接说明了对于 0 以外的代数数 x，e^x 均为超越数，它的反函数 $\ln(x)$ 也只经过 $(1, 0)$ 这个唯一的横纵坐标都是代数数的点，套入欧拉公式，可以得出除了几个特殊的点以外，三角函数和双曲函数也不经过横纵坐标都是代数数的点，比如 $\sin(1)$ 就是一个超越数.

(2) 可计算数 (computable number). 尽管超越数不能通过解整系数的高次多项式方程得到，比如 π, e 等，但是它们也许存在解析形式的表达式，或者有一个算法可以把它们计算到任意精确的程度，通俗的说就是“知道它怎么算”. 全体可计算数的势为 $\mathrm{card}(\mathbb{N})$. 不可计算意味着无从得知某些量精确的值，但是可以证明这些量是一个定值，比如蔡廷常数 (Chaitin's constant)Ω，不严谨地定义为随机一段可以运行的程序最终能停下来的概率，根据概率的性质可以知道 $0 < \Omega < 1$，但没有更准确的范围了.

(3) 可定义数 (definable number). 对于这类数，只能通过非构造式的方法意识到它们的存在，但它们中的任意一个元素都无法被定义出来，因为一旦做到了，就不是不可定义数了. 证明的思路很简单[6]，任意一个“定义”都是一串长度有限的字符串，而字符串中的每个不可分割的字符可能的选取都是有限的（我们一般把包含所有不可分割的字符的集合称为字母表或者字典），把字典中的字符编号，并用编号代替字符本身，那么每一个字符串就可以写成一个长长的自然数，这样就建立了所有“定义”的集合到自然数集的单射，实数集远远要大于自然数集，因此有些实数是不可被定义的.

假设有一个集合 A，对于一个实数，如果我们能证明它是无理数，就把它加入集合 A，由于“证明”的过程也是一个长度有限的字符序列，所以可以得到 $A \cup \mathbb{Q} \neq \mathbb{R}$，并且 $\mathrm{card}(A \cup \mathbb{Q}) < \mathrm{card}(\mathbb{R})$. 若往现有的公理体系中加入更多更强的公理，就可以扩充 A，但并不会改变它的势，我们完全可以找到一个实数，在现有公理体系的某个模型下证明它是无理数，在另一个模型下证明它是无理数. 说到底，有理和无理的性质已经触及到了数学大厦的总体面貌，非常依赖于讨论时定下的哲学纲领，关于它的研究目前是一个较为热门的方向.

13.6 复数域的构造

回顾抽象代数中关于“域”的定义：

集合 F 连同定义在其上的两种运算 $(+, \cdot)$，如果满足以下这 5 条规则，则称 $\langle F; +, \cdot \rangle$ 为一个域：

(1) 交换律：$a + b = b + a; \quad a \cdot b = b \cdot a.$

(2) 结合律：$a + (b + c) = (a + b) + c; \quad a \cdot (b \cdot c) = (a \cdot b) \cdot c.$

(3) 存在唯一且不同的单位元：$\exists! \theta : a + \theta = a, \quad \exists! \tau : a \cdot \tau = a, \quad \theta \neq \tau$

(4) 存在逆元：$\forall a \exists b : a + b = \theta, \quad \forall a \neq \theta \exists b : a \cdot b = \tau$

[6] 这种把字符串重新编码为一个自然数的证明思路称为哥德尔编码 (Godel number) 方法

> (5) 分配律：$(a+b)\cdot c=(a\cdot c)+(b\cdot c)$.
>
> 在两个代数结构建立映射 $h:\langle\mathbb{F};\vec{\mu}_{\mathbb{F}}\rangle\mapsto\langle\mathbb{G};\vec{\mu}_{\mathbb{G}}\rangle$，如果 h 能保持每个运算 μ^k，即 $h(\mu^k_{\mathbb{F}}(\vec{a}))=\mu^k_{\mathbb{G}}(h(\vec{a}))$，则该映射 h 称为这两个代数结构的同态映射，如果 h 为双射，则称为同构映射.

实数集连同其上定义的加法和乘法 $\langle\mathbb{R};+,\cdot\rangle$ 显然是一个域，交换律、结合律和分配律的证明略去，任意实数 a 的加法逆元是 $-a$，乘法逆元是 $1/a$，加法单位元是 0，乘法单位元是 1.

现在推广一下. 在 $\mathbb{R}^2$ 平面上定义坐标之间的两种运算：

$$(a,b)+(c,d):=(a+b,c+d)$$
$$(a,b)\cdot(c,d):=(ac-bd,bc+ad)$$

容易验证 $\langle\mathbb{R}^2;+,\cdot\rangle$ 也是一个域，交换律、结合律、分配律证明略，加法的单位元是 $(0,0)$，乘法的单位元是 $(1,0)$，任意坐标 (a,b) 的加法逆元是 $(-a,-b)$，乘法逆元是 $\left(\dfrac{a}{a^2+b^2},\dfrac{-b}{a^2+b^2}\right)$.

更进一步地，映射 $\lambda a.(a,0):\langle\mathbb{R};+,\cdot\rangle\mapsto\langle\mathbb{R}^2;+,\cdot\rangle$ 是一个同构映射. 考虑到 $(a,b)=a\cdot(1,0)+b\cdot(0,1)$，而 $h(1)=(1,0)$，是乘法的单位元，所以我们可以记 $i=(0,1)$，这样 (a,b) 就可以写作 $a+b\cdot i$，再简写为 $a+bi$，现在我们已经把所有的符号都约定好了，接下来给出正式定义.

定义　复数域

复数集定义为 $\mathbb{C}=\{(a,b)|a,b\in\mathbb{R}\}$，
复数的加法定义为 $(a,b)+(c,d):=(a+b,c+d)$，乘法定义为 $(a,b)\cdot(c,d):=(ac-bd,bc+ad)$.
(a,b) 简写为 $a+bi$，$i=(0,1)$ 称为<u>虚数单位</u>(imaginery unit).

回到最初的起点，我们在 $\mathbb{R}^2$ 平面上定义了加法和乘法，然后才导出了复数的概念. 众所周知，想在 $\mathbb{R}^2$ 平面上确定一个点，除了使用直角坐标以外，还可以使用极坐标，直角坐标 (x,y) 和极坐标 (ρ,θ) 的转换关系如下：

$$\begin{cases}x=\rho\cos\theta\\ y=\rho\sin\theta\end{cases}\Longleftrightarrow\begin{cases}\rho=\sqrt{x^2+y^2}\\ \theta=\operatorname{atan2}(x,y)\end{cases}$$

其中 ρ 称为极径，$\theta=\operatorname{atan2}(x,y)$ 称为极角，定义为

$$\operatorname{atan2}(x,y)=\begin{cases}\arctan\dfrac{y}{x}, & x>0\\ \arctan\dfrac{y}{x}+\pi, & x<0,y\geq 0\\ \arctan\dfrac{y}{x}-\pi, & x<0,y<0\\ \dfrac{\pi}{2}, & x=0,y>0\\ -\dfrac{\pi}{2}, & x=0,y<0\\ \text{无定义}, & x=0,y=0\end{cases}$$

atan2 函数尽管是分段定义的，但它却是连续函数. 根据该定义，可见该极角的范围是 $(-\pi,\pi]$. 对于复数，我们也有类似定义.

定义　复数的极坐标形式

令 $z=a+bi$，则 $z=\rho\angle\mathrm{Arg}(z)$，其中 $\rho=\sqrt{a^2+b^2}$ 是复数的模(modulo)；$\mathrm{Arg}(z)=\mathrm{atan2}(a,b)$ 称为复数的辐角(argument).

这只是暂时的一种记法，接下来介绍完指数函数以后，将使用更好的指数形式来代替极坐标形式.

我们还注意到，复数和 $\mathbb{R}^2$ 上向量的加法规则是相同的，如果把"数量"仅限于实数，那么数量乘法的规则也是相同的. 因此复数也有所谓"三角形法则"和"平行四边形法则". 还可以进一步注意到，如果把 $i=i\cdot 1$ 中等号左边的 i 和右边的 1 写成列向量的形式，就是

$$\begin{bmatrix}0\\1\end{bmatrix}=i\cdot\begin{bmatrix}1\\0\end{bmatrix}$$

同时对比向量的线性变换式：

$$\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}$$

可以发现虚数单位 i 和顺时针旋转 90 度的矩阵作用是相似的，接下来我们将会发现更多这种相似.

第二部分　线性代数（remake，预计至少 45 页）

线性代数研究的对象是向量和线性变换. 这句话非常笼统，但这也意味着线性代数的覆盖范围其实非常广泛. 如果从历史的角度来看，线性代数是向量运算的形而上学，向量在线性代数中有至关重要的地位.

什么是向量？在初等数学中，向量是一个有限长的数字列表，例如 $(2,3,1)$，和坐标的写法相同. 按照形式主义的观点来看，"写起来完全相同"意味着"是同一个东西"，而真正不同的地方在于向量能进行各种运算（不过接下来会经常混淆两者）. 当我们秉持着数形结合的思想引入坐标系后，坐标往往表示坐标系中的一个点，箭头表示坐标系中的有向线段，也可以描述坐标系中某种有限长直线运动的过程.

经过了初等数学的洗礼之后，我们可以很自然地理解"坐标系""坐标系中的点""有向线段"等概念表达的意思，这当然是好事，但这同时也意味着我们容易对这些概念产生固有印象，认为某个名词必须要这么理解. 这一章将逐步破除这些桎梏，但请注意，将自己从思想牢笼中解放出来不代表完全抛弃牢笼中的思想，而是在其之上推广和进步，利用其来帮助理解推广后的概念.

不过我们首先还是要全面回顾一下"牢笼中的思想"：向量代数.

14　回顾：最基础的向量

如前言所述，狭义的向量(vector) 是一个有限长的数字列表，向量的元数就是列表的项数，在集合论的层面上可以诠释为有序组. 很多教科书里都把一般的 n 元的向量写成 $(a_1,a_2,\cdots,a_n)$，我们可以从集合论里借鉴一些表达得更明确的写法，比如 $(a_i)_{i=1}^n$，如果不确定或不强调有 n 元，那么可以简写成 $(a_i)_i$，使用加粗的希腊字母 $\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\cdots$ 来表示向量.

特别地，如果 $i=1$，即只有一元的向量 (a_i)，可以认为它就是一个数 a_i，不必再当作向量看待（你会发现此时向量的运算法则和数量的运算法则是兼容的）.

向量可以进行加法运算，定义为 $(a_i)_i+(b_i)_i:=(a_i+b_i)_i$，在几何上表现为：如果将 $(a_i)_i$ 和 $(b_i)_i$ 首尾相接，那么加法得到的向量的起点是 $(a_i)_i$ 的起点、终点是 $(b_i)_i$ 的终点，这称为向量相加的三角形法则. 你可以进一步发现，$(a_i)_i$ 和 $(b_i)_i$ 谁的首接谁的尾并不影响结果，对于多个向量相加的时候也是如此，此即为向量相加的多边形法则，这说明向量的加法是良定义的.

就像实数可以取相反数一样，向量 $\boldsymbol{\alpha}$ 也可以取“相反向量” $-\boldsymbol{\alpha}$，相反向量在几何上表现为头尾调换的原向量. 有了相反向量的概念，我们可以定义向量的减法 $\boldsymbol{\alpha}-\boldsymbol{\beta}:=\boldsymbol{\alpha}+(-\boldsymbol{\beta})$. 在几何上，把两个向量的尾部靠在一起，从减向量的终点指向被减向量的终点即为结果.

向量还可以进行数乘运算，用一个数量 c 乘以向量 $\boldsymbol{\alpha}$ 的结果 $c\cdot\boldsymbol{\alpha}$ 是一个向量，在几何上相当于把向量的长度改为原来的 c 倍，仅做了伸缩而没有旋转. 如果将 $\boldsymbol{\alpha}$ 和 $c\boldsymbol{\alpha}$ 的起点放在一起，那么你可以画一条直线将这两个向量完全覆盖，对于这种情况，我们说这两个向量是<u>共线的</u>(colinear) 或<u>平行的</u>(parallel).

我们还可以计算向量的<u>内积</u>(inner product)：

$$\langle(a_i)_{i=1}^n,(b_i)_{i=1}^n\rangle:=\sum_{i=1}^n a_ib_i$$

即两个向量对应的元相乘后求和，结果是一个数量，根据乘法的交换律和分配律，内积也满足交换律和分配律. 内积在几何上表现为两个向量（有向线段）的长度和它们之间的夹角余弦值三者相乘. 由于非零向量和自己的夹角是 0，所以自己和自己做内积的结果是自身长度的平方，由此正式定义出向量的<u>模</u>(modulo)：$\|\boldsymbol{\alpha}\|:=\sqrt{\langle\boldsymbol{\alpha},\boldsymbol{\alpha}\rangle}$，这时向量<u>夹角</u>(angle) 的余弦值就可以定义为

$$\cos\angle(\boldsymbol{\alpha},\boldsymbol{\beta}):=\frac{\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle}{\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|}$$

有两个（类）值得注意的向量，零向量 $\mathbf{0}$ 是一个特殊的向量，它的模是 0，它与其他向量的夹角不可计算，所以我们说零向量没有方向，或者任意一个方向都是它的方向. <u>单位向量</u>(unit vector) 则是模长为 1 的向量，往往起着“指向”的作用. 向量除以自身的模长之后得到的就是与同方向的单位向量，记作 $\hat{\boldsymbol{\alpha}}:=\dfrac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|}$，这个过程称为<u>归一化</u>(normalization). 此时原本的向量就可以写作 $\|\boldsymbol{\alpha}\|\hat{\boldsymbol{\alpha}}$，体现向量具有方向和大小这两个属性，这是一个非常重要的技巧，在接下来很多地方都会用到.

当两向量夹角为 $\pi/2$ 时，余弦值为 0，因此内积也为 0，此时称这两个向量是<u>正交的</u>(orthogonal)，在几何上画出来确实如此. 同理，当夹角为 0 时，$\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle=\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|$，即共线时两个向量的内积即为模长相乘，夹角为 π 则变成相反数.

如果用夹角来表示内积，则有 $\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle=\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|\cos\angle(\boldsymbol{\alpha},\boldsymbol{\beta})$. 注意红色部分，在几何上它表示向量 $\boldsymbol{\beta}$ 投影在向量 $\boldsymbol{\alpha}$ 上的长度，投影后的向量与 $\boldsymbol{\alpha}$ 同向，所以应该乘以 $\hat{\boldsymbol{\alpha}}$，据此我们可以定义<u>投影</u>(projection)：

$$\begin{aligned}\operatorname{proj}_{\boldsymbol{\alpha}}(\boldsymbol{\beta})&:=\|\boldsymbol{\beta}\|\cos\angle(\boldsymbol{\alpha},\boldsymbol{\beta})\hat{\boldsymbol{\alpha}}\\&=\|\boldsymbol{\beta}\|\frac{\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle}{\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|}\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|}\\&=\frac{\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle}{\langle\boldsymbol{\alpha},\boldsymbol{\alpha}\rangle}\boldsymbol{\alpha}\end{aligned}$$

投影也是一个很重要的概念，接下来讨论到基底的时候会经常用到.

基底是一个神奇的东西. 以空间直角坐标系为例，某个向量 $\boldsymbol{\alpha}$ 的表达式 $(2,4,3)$ 究竟代表着什么？如果我们规定三个向量 $\hat{\boldsymbol{\imath}}=(1,0,0),\hat{\boldsymbol{\jmath}}=(0,1,0),\hat{\boldsymbol{k}}=(0,0,1)$，那么显然有

$$\boldsymbol{\alpha}=(2,4,3)\iff\boldsymbol{\alpha}=2\boldsymbol{i}+4\boldsymbol{j}+3\boldsymbol{k}$$

两个式子究竟哪个更本质？左边的式子将向量看作一个坐标，右边的式子将则先给出基底，然后让向量分解为基底向量的加权和. 如果认为左边的式子更本质，那么可以进入离散数学的世界，如果认为右边的式子更本质，那么走进的才是线性代数的世界.

15 向量空间

15.1 公理化的向量

公理集合论没有解释“集合是什么”，而直接把集合当作不可被更基本的概念定义的原始观念. 同理，在公理化的系统下“向量”这个概念也被当作原始观念. 更进一步地，通过对集合论公理的演绎以及对集合运算的诠释，我们在印象中大概会将“集合”的概念同现实中的“袋子”“盒子”等容器联系起来，即认为“集合”就是一个从各种容器中抽象出来的概念，“元素”则是“容器”中装的物品；麻烦的是向量公理比集合论更不直观：在上一节中，我们将“向量”同几何中“有向线段”的概念联系起来，而向量公理则抛弃了这种联系，凡是能进行某种运算的对象都能称为向量. 尽管公理层面的“向量”比“集合”更难理解，但托集合论的福，向量公理叙述起来更为方便.

定义 线性空间

设 $\mathbb{F}$ 是一个数域，S 是一个集合；$\mathbb{F}$ 中的元素使用正常字体表示，S 中的元素使用加粗字体表示；$\mathbb{F}$ 中的单位元记作 1，S 中存在常量 $\mathbf{0}$；S 上装备了两种运算“$+ : S \times S \to S$”和“$\cdot : \mathbb{F} \times S \to S$”，这两种运算满足以下 8 条向量公理

$$
\begin{array}{lll}
\forall \boldsymbol{\alpha}, \boldsymbol{\beta}: & \boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\beta}+\boldsymbol{\alpha} & \text{“}+\text{”的交换律} \\
\forall \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}: & (\boldsymbol{\alpha}+\boldsymbol{\beta})+\boldsymbol{\gamma}=\boldsymbol{\alpha}+(\boldsymbol{\beta}+\boldsymbol{\gamma}) & \text{“}+\text{”的结合律} \\
\forall \boldsymbol{\alpha}: & \boldsymbol{\alpha}+\mathbf{0}=\boldsymbol{\alpha} & \text{“}+\text{”存在零元} \\
\forall \boldsymbol{\alpha} \exists \boldsymbol{\beta}: & \boldsymbol{\alpha}+\boldsymbol{\beta}=\mathbf{0} & \text{“}+\text{”存在逆元} \\
\forall a, b, \boldsymbol{\alpha}: & (ab) \cdot \boldsymbol{\alpha}=a \cdot(b \cdot \boldsymbol{\alpha}) & \text{“}\cdot\text{”的结合律} \\
\forall \boldsymbol{\alpha}: & 1 \cdot \boldsymbol{\alpha}=\boldsymbol{\alpha} & \text{“}\cdot\text{”存在单位元} \\
\forall a, b, \boldsymbol{\alpha}: & (a+b) \cdot \boldsymbol{\alpha}=a \cdot \boldsymbol{\alpha}+b \cdot \boldsymbol{\alpha} & \text{数量分配律} \\
\forall a, \boldsymbol{\alpha}, \boldsymbol{\beta}: & a \cdot(\boldsymbol{\alpha}+\boldsymbol{\beta})=a \cdot \boldsymbol{\alpha}+a \cdot \boldsymbol{\beta} & \text{向量分配律}
\end{array}
$$

那么 $\langle\mathbb{F}, S ;+, \cdot ; 1, \mathbf{0}\rangle$ 统称为线性空间(linear space)，其中的元素称为向量(vector).

这里要注意：

(1) 线性空间并没有规定数量 0 和向量 $\boldsymbol{\alpha}$ 相乘后一定是某个向量 $\mathbf{0}$，但在大多数情况下我们都会这么规定，即

> $\mathbb{F}$ 中的零元记作 0，对于任意 $\boldsymbol{\alpha}$，有 $0 \cdot \alpha=\mathbf{0}$；对于任意 c，有 $c \cdot \mathbf{0}=\mathbf{0}$.

(2) 书写时常常省略“·”符号；

(3) 与其说 $\langle\mathbb{F}, S ;+, \cdot ; 1, \mathbf{0}\rangle$ 是一个向量空间，一般更倾向于直接简称 S 是线性空间.

(4) 线性空间的“线性”究竟是什么意思，将在接下来讨论线性变换时详细说明. 而“空间”一词则没有什么可说的，定义在某个集合上的运算，只要运算的结果跳不出这个集合，那么这个集合都可以称为“空间”. 线性空间中，向量经过加法和数乘运算的结果仍在定义域中，所以称为“空间”.

就像集合有子集，线性空间也有线性子空间. 对象如其名，看到“子”这个字，你应当立即意识到子空间是原空间的子集. 但除了“是原空间的子集”以外，它还是“空间”，应当满足以上这 8 条公理.

定义 线性子空间

设 $\langle \mathbb{F}, S; +, \cdot; 1, \mathbf{0}\rangle$ 是一个线性空间，而 $T \subseteq S$ 且 $\mathbf{0} \in T$，此外 “$+$” 和 “$\cdot$” 在 T 中封闭，则 T 是 S 的线性子空间(linear subspace).

15.2 线性相关性

有了公理化的向量，我们来回顾本章开始时定义的狭义的向量，它是一个数字列表，在空间直角坐标系中，$\boldsymbol{\alpha} = (2, 4, 3)$ 意味着 $\boldsymbol{\alpha} = 2\hat{\boldsymbol{\imath}} + 4\hat{\boldsymbol{\jmath}} + 3\hat{\boldsymbol{k}}$，在这里，向量 $\boldsymbol{\alpha}$ 被向量 $\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}$ 使用加法数量乘法表示出来了. 问题来了，一定需要 $\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}\}$ 这三个向量吗？就不能去掉一个吗？

我们可以这么理解，$\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}\}$ 之中的每一个向量都引入了独特的方向. $\hat{\boldsymbol{\imath}}$ 向量数乘以后得到的向量只能分布在一条直线上，$\hat{\boldsymbol{\jmath}}$ 也是如此，但 $\hat{\boldsymbol{\imath}}$ 和 $\hat{\boldsymbol{\jmath}}$ 得到的直线并不重叠或平行；$\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}\}$ 数乘后求和得到的向量可以分布在一个平面上，$\{\hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}\}$ 也如此，但这两组向量得到的平面并不平行或重叠；$\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}\}$ 数乘求和后的向量可以遍布整个立体空间. 每个向量都是独特的，不同的向量组合也都是不同的，并且随着向量的增加，数乘求和得到的向量丰富性大大增加.

给个反面教材，$\boldsymbol{a} = (0, 1, 1), \boldsymbol{b} = (1, 0, 1), \boldsymbol{c} = (3, 2, 5)$ 这三个向量，我们很容易发现 $\boldsymbol{c} = 2\boldsymbol{a} + 3\boldsymbol{b}$，那这就意味着 $\{\boldsymbol{a}, \boldsymbol{b}\}$ 和 $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ 数乘后相加所得到的向量是一样的，$\boldsymbol{c}$ 的出现并没有引入新的方向，没有提升向量空间的丰富度.

对于以上两个例子，我们定义几个概念.

定义 线性相关

对于一组向量 $\{\boldsymbol{\alpha}_i\}_{i=1}^n$：

- 表达式 $\displaystyle\sum_{i=1}^n c_i \boldsymbol{\alpha}_i = c_1\boldsymbol{\alpha}_1 + c_2\boldsymbol{\alpha}_2 + \cdots + c_n\boldsymbol{\alpha}_n$ 称为这组向量的线性组合(linear combination)，其中 c_i 是任取的数量，**可以全为零**.
- 由该组向量的线性组合的来的所有向量的集合，称为该组向量张成的空间，记作 $\operatorname{span}\{\boldsymbol{\alpha}_i\}_{i=1}^n$.
- 如果某个向量 $\boldsymbol{\beta}$ 等于该组向量的某个线性组合（即属于该张成的空间），那么就说 $\boldsymbol{\beta}$ 能被该组向量线性表出.
- 如果零向量被该组向量线性表出时系数 c_i 全为零，那么就说该组向量线性无关(linearly independent)，否则为线性相关(linearly dependent).

现在我们可以说，向量 $\hat{\boldsymbol{\imath}}$ 张成的空间是一条直线，$\operatorname{span}\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}\}$ 是一个平面，$\operatorname{span}\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}\}$ 则是一个立体的空间. 若想得到立体空间中的 $\mathbf{0}$ 向量，唯一的组合是 $0\hat{\boldsymbol{\imath}} + 0\hat{\boldsymbol{\jmath}} + 0\hat{\boldsymbol{k}}$，系数全为 0，所以 $\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}\}$ 线性无关.

接下来通过几个命题来刻画以上这些概念

(1) $\{\boldsymbol{\alpha}_i\}_{i=1}^n$ 线性无关的充要条件是任一 $\boldsymbol{\alpha}_i$ 都不能被剩余 $n-1$ 个向量线性表出.

证明 反证法.（右推左）反设 $\{\boldsymbol{\alpha}_i\}_{i=1}^n$ 线性相关，那么就有不全为零的数 $\{c_i\}_{i=1}^n$，使得 $\mathbf{0} = \displaystyle\sum_{i=1}^n c_i\boldsymbol{\alpha}_i$，把 $\boldsymbol{\alpha}_1$ 移到左边然后除以 $-c_1$，得到 $\boldsymbol{\alpha}_1 = -\displaystyle\sum_{i=2}^n \frac{c_i}{c_1}\alpha_i$，这就是 $\boldsymbol{\alpha}$ 的线性表出，矛盾了.

（左推右）反设 □

(2) 若某组向量 $\mathcal{A}$ 线性无关，则其任意非空子集都线性无关；若 $\mathcal{A}$ 线性相关，则其任意超集都线性相关.

(3) 若 $\mathbf{0}\in\mathcal{A}$，则 $\mathcal{A}$ 一定线性相关.

(4) 若 $\mathcal{A}$ 线性无关，但 $\mathcal{A}\cup\{\boldsymbol{\beta}\}$ 线性相关，那么 $\boldsymbol{\beta}$ 可用 $\mathcal{A}$ 中的向量线性表出，且表出的方式是唯一的.

根据以上命题 (2)，我们定义极大无关组

定义　极大线性无关组

对于向量组 $\mathcal{A}$, 若 $\mathcal{B}\subseteq\mathcal{A}$ 线性无关, 而且任意 $\mathcal{C}\supseteq\mathcal{B}$ 都线性相关, 那么 $\mathcal{B}$ 称为 $\mathcal{A}$ 的<u>极大线性无关组</u>, $\operatorname{card}(B)$ 称为 $\mathcal{A}$ 的<u>秩</u>(rank)，记作 $r(\mathcal{A})$.

(5) 若 S 是线性空间，向量 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\in S$，则 $\operatorname{span}\{\boldsymbol{\alpha}_i\}_1^n$ 是 S 的线性子空间.

(6) 若 $\mathcal{B}$ 是 $\mathcal{A}$ 的极大线性无关组，那么 $\operatorname{span}\mathcal{A}=\operatorname{span}\mathcal{B}$，且 $r(\mathcal{A})=r(\mathcal{B})$.

对于狭义的向量，又可以补充以下两个命题.

(7) n 元狭义向量的线性无关组中至多有 n 个向量.

回到本节开头的例子，空间直角坐标系中只需要 $\{\hat{\boldsymbol{\imath}},\hat{\boldsymbol{\jmath}},\hat{\boldsymbol{k}}\}$ 就可以表出所有空间向量，而平面直角坐标系只需要 $\{\hat{\boldsymbol{\imath}},\hat{\boldsymbol{\jmath}}\}$ 两个向量. 你也许也发现了对于狭义的向量，线性无关组似乎有上界：

(8) 设 $(a_i)_{i=1}^n$ 为一个狭义向量，而另一个狭义向量 $(b_i)_{i=0}^m$ 满足 $m>n$，且对于 $1\le i\le n$ 均有 $a_i=b_i$，那么 $(b_i)_{i=1}^m$ 称为 $(a_i)_{i=1}^n$ 的延伸向量. 对于向量组 $\mathcal{A}=\{\boldsymbol{\alpha}_i\}_{i=1}^n$ 和 $\mathcal{B}=\{\boldsymbol{\beta}_i\}_{i=1}^n$，若 $\boldsymbol{\beta}_i$ 是 $\boldsymbol{\alpha}_i$ 的延伸向量，那么称 $\mathcal{B}$ 是 $\mathcal{B}$ 的延伸组，$\mathcal{A}$ 是 $\mathcal{B}$ 的缩短组. 如果某一组向量线性无关，那么其延伸组线性无关；如果某一向量组线性相关，那么其缩短组也线性相关.

15.3　空间的基与坐标

上一节详细叙述了线性相关性，却一直没有捅破窗户纸：我们只将 $\{\hat{\boldsymbol{\imath}},\hat{\boldsymbol{\jmath}},\hat{\boldsymbol{k}}\}$ 看作是一个线性无关的向量组，也讲明了 $\operatorname{span}\{\hat{\boldsymbol{\imath}},\hat{\boldsymbol{\jmath}},\hat{\boldsymbol{k}}\}$ 为整个立体空间，但一直没有出现“基底”一词，似乎刻意避开了这个概念——其实是为了留到现在.

从直觉上来说，有了基底，空间中的向量就可以表示为基底向量的线性组合，就像 $\boldsymbol{\alpha}=2\hat{\boldsymbol{\imath}}+4\hat{\boldsymbol{\jmath}}+3\hat{\boldsymbol{k}}$ 一样. 除此之外，我们还希望基底向量是线性无关的，即每个向量都能引出独特的方向.

定义　空间的基与坐标

若 $\mathcal{A}=\{\boldsymbol{\alpha}_i\}_{i=1}^n$ 是一个线性无关的向量组，而且线性空间 S 中的每一个向量都可以用 $\mathcal{A}$ 中向量的线性组合来表示（允许系数全为 0，那么称 $\mathcal{A}$ 是 S 的一个基底.

若某个向量 $\boldsymbol{\beta}\in S$ 为 $\boldsymbol{\alpha}=\displaystyle\sum_{i=1}^n c_i\boldsymbol{\alpha}_i$，那么狭义向量 $(c_i)_{i=1}^n$ 称为向量 $\boldsymbol{\beta}$ 在基底 $\mathcal{A}$ 下的<u>坐标</u>(coordinate)，记作 $[\boldsymbol{\beta}]_{\mathcal{A}}$.

坐标一词终于出现了，我们发现坐标就是狭义向量，而系数的任意性又保证了每个狭义向量都是一个坐标，所以现在我们规定：**形如 $(a_i)_i$ 这样的有序组，今后不再称为狭义向量，而称为坐标**. 值得注意的是，我们常常将坐标竖着写，例如 $\mathcal{A}=\{\hat{\boldsymbol{\imath}},\hat{\boldsymbol{\jmath}},\hat{\boldsymbol{k}}\}$ 是空间直角坐标系中的基，那么

$$\boldsymbol{\beta}=2\hat{\boldsymbol{\imath}}+4\hat{\boldsymbol{\jmath}}+3\hat{\boldsymbol{k}}\Longleftrightarrow[\boldsymbol{\beta}]_{\mathcal{A}}=\begin{bmatrix}2\\4\\3\end{bmatrix}$$

这种写法已经有了矩阵的影子，至于为什么这样写，那是历史遗留问题了⋯⋯但我们将会发现该写法在方程组的简化表示上起到非常大的作用.

每个线性空间是否都存在基底？我们希望它是存在的. 激动人心的一点是，证明每个线性空间都有基底依赖于选择公理——如此底层的公理居然在这里派上了用场.

命题 每个线性空间都有基底.

15.4 线性代数的核心：线性变换

* 这是非常重要的内容，除了线性变换以外，矩阵也将首次出现.

15.4.1 定义和性质

什么是“线性”? 在初等数学中，如果某个因变量正比于自变量，或者某事物的发展与另一事物总是相差某个倍数，亦或者某个量在等长的两段时间间隔内的变化是相等的，那么我们就会为这种情况打上“线性”的标签. 不难发现，这三个例子中，“线性”的背后总暗含着关系：因变量与自变量的关系、两个事物发展情况的关系、某个量与时间的关系，因此“线性”一词实际上是对于联系的描述. 联系在数学中体现为函数，我们可以为函数定下线性的定义，既严格又符合我们的直觉.

定义 线性映射

设 S,T 为线性空间，如果一个映射 $f:S\to T$ 满足以下两点

(1) $f(\boldsymbol{\alpha}+\boldsymbol{\beta})=f(\boldsymbol{\alpha})+f(\boldsymbol{\beta})$

(2) $f(c\cdot\boldsymbol{\alpha})=c\cdot f(\boldsymbol{\alpha})$

则称 f 为线性映射(linear mapping)；如果 $S=T$，那么一般又称为线性变换(linear transformation).

15.4.2 矩阵：”Hello World!”

既然向量空间中都存在基底，那么每一个向量都可被基底表现出来. 设某个向量空间中的基底是 $\mathcal{A}=\{\boldsymbol{\alpha}_i\}_{i=1}^n$，某个向量 $\boldsymbol{\beta}=\sum\limits_{i=1}^n c_i\boldsymbol{\alpha}_i$，现在把 $\boldsymbol{\beta}$ 放入一个线性变换 f 中：

$$f(\boldsymbol{\beta})=f\left(\sum_{i=1}^n c_i\boldsymbol{\alpha}_i\right)=\sum_{i=1}^n c_i f(\boldsymbol{\alpha}_i)$$

可见，我们只要知道每个基底向量经过线性变换后的结果，就可以知道任意一个向量的结果. 更进一步地，既然基底向量数量有限，其经过线性变换后也是有限的，元数也有限，那么我们可以将基底向量线性变换后的基底记录下来形成一个列表，其他向量即可根据该列表记录下来.

继续这个例子，记 $f(\boldsymbol{\alpha}_i)$ 的坐标为 $(a_{j,i})_{j=1}^n$，不要问我为什么 j 要放前面，这是历史使然. 继续上式的推导：

$$f(\boldsymbol{\beta})=\sum_{i=1}^n c_i f(\boldsymbol{\alpha}_i)=\sum_{i=1}^n c_i\sum_{j=1}^n a_{ji}\boldsymbol{\alpha}_j=\sum_{j=1}^n\left(\sum_{i=1}^n c_i a_{ji}\right)\boldsymbol{\alpha}_j \tag{*}$$

接下来我们把每个 $f(\boldsymbol{\alpha}_i)$ 竖着写的坐标拼起来，得到一张矩形表格：

$$\left[\begin{array}{c|c|c|c} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{21} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{array}\right]_{n\times n}$$

在给定空间的基底 $\mathcal{A}$ 的情况下，这张矩形表可以确定线性变换 f 本身，所以称为线性变换的矩阵表示(transformation matrix)，也称作基变换矩阵(change of basis)．我们可以从集合论中借鉴更先进的表达式来表示矩阵，如上面这个矩阵可以表示为 $[a_{ij}]_{n\times n}$，坐标可以看成 $n\times 1$ 的矩阵，所以可以写为 $[a_i]_{n\times 1}$．接下来使用粗体来表示矩阵.

就像向量可以进行加法和乘法一样，我们也要给矩阵制定加法和乘法的规则．这里只介绍规则背后的思想，更深入的内容将在后面讨论．在定下具体的表达式之前，先来说一说我们希望它应有的样子．既然矩阵的功能是确定某个线性变换，那么我们可以规定矩阵相加相当于线性变换相加；在某些文献上，一个函数 f 作用于某个变量 x 有时会直接写作 fx——是相乘的形式．由此我们可以**将矩阵与单个坐标的乘积定义为矩阵对应的线性变换作用于该向量所得到的坐标**，即 $f(\boldsymbol{\alpha}) := \boldsymbol{A}\boldsymbol{\alpha}$，其中 $\boldsymbol{A}$ 是变换矩阵.

具体是多少呢？上面其实推导了出来了，新坐标的第 j 元就是 $\sum\limits_{i=1}^{n} c_i a_{ji}$．所以我们可以规定

$$[a_{ij}]_{n\times n}\times[c_i]_{n\times 1} := \left[\sum_{i=1}^{n} c_i a_{ji}\right]_{n\times 1}$$

贴心地展开写出来：

$$\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{21} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}_{n\times n} \times \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1n}c_n \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2n}c_n \\ \vdots \\ a_{n1}c_1 + a_{n2}c_2 + \cdots + a_{nn}c_n \end{bmatrix}_{n\times 1}$$

问题又来了，我们定义了矩阵乘以坐标的结果，那么矩阵乘以矩阵呢？回顾集合论中复合函数的写法：$f(g(x)) = (f\circ g)(x)$，右式居然写成了两个函数的某种运算的形式．在左式中，我们要先算 $g(\cdot)$，再算 $f(\cdot)$，而右式让我们先算 $f\circ g$，再算这个整体作用于 x 的结果．由于 $g(\boldsymbol{\alpha})$ 可以写成一个矩阵乘以一个向量的形式，$(f\circ g)(x)$ 也可以写成一个矩阵乘以一个向量的形式，所以 $f\circ g$ 应当可以写成一个矩阵.

至于该矩阵的每个元是多少，接下来推导一下．设基底是 $\mathcal{A} = \{\boldsymbol{\alpha}_i\}_{i=1}^{n}$，$\boldsymbol{\beta} = \sum\limits_{i=1}^{n} c_i\boldsymbol{\alpha}_i$，两个线性变换 f, g 的矩阵表示是 $\boldsymbol{A} = [a_{ij}]_{n\times n}$ 和 $\boldsymbol{B} = [b_{ij}]_{n\times n}$，由上文已知，放入线性变换 $g(\cdot)$ 后

$$g(\boldsymbol{\beta}) = \sum_{j=1}^{n}\left(\sum_{i=1}^{n} c_i a_{ji}\right)\boldsymbol{\alpha}_j$$

令 $d_j = \sum\limits_{i=1}^{n} c_i a_{ji}$，接下来再计算 $f(g(\boldsymbol{\beta}))$：

$$f(g(\boldsymbol{\beta})) = f(\sum_{j=1}^{n} d_j\boldsymbol{\alpha}_j) = \sum_{j=1}^{n} d_j f(\boldsymbol{\alpha}_j) = \sum_{j=1}^{n} d_j \sum_{k=1}^{n} b_{kj}\boldsymbol{\alpha}_k$$

再把 d_j 写出来：

$$f(g(\boldsymbol{\beta})) = \sum_{j=1}^{n}\sum_{i=1}^{n} c_i a_{ji} \sum_{k=1}^{n} b_{kj}\boldsymbol{\alpha}_j = \sum_{k=1}^{n}\left[\sum_{i=1}^{n} c_i\left(\sum_{j=1}^{n} a_{ji}b_{kj}\right)\right]\boldsymbol{\alpha}_k$$

对照该式与上文的 * 式，可以发现该式的红色部分对应 * 式中的 a_{ji}，所以我们可以规定矩阵乘法为

$$[a_{ij}]_{n\times n}\times[b_{ij}]_{n\times n}:=\left[\sum_{j=1}^{n}a_{ij}b_{jk}\right]_{n\times n}$$

而加减法就简单了，由于加减号可以从线性变换中拿出来，所以可以简单地定义为

$$[a_{ij}]_{n\times n}\pm[b_{ij}]_{n\times n}:=[a_{ij}\pm b_{ij}]_{n\times n}$$

本节仅仅叙述了矩阵的加法和乘法的由来，并没有对其性质深入讨论，这些工作将在后面的章节完成.

16 矩阵运算

17 行列式

18 特征值理论

19 二次型理论

第三部分 级数和初等函数

1 基本初等函数

函数的初等与否是人为划分的概念，最初由刘维尔引入. 基本初等函数(basic elementary function) 包括常数函数、幂函数、指数函数、三角函数、双曲函数以及它们的反函数，而基本初等函数经过有限次四则运算和复合得到的函数称为初等函数(elementary function). 分出初等函数是为了固定这些函数的记号以便于交流，同时初等函数也是生活中最常见的函数.

1.1 指数函数

在复数域中，最重要最基础的函数是指数函数，特指以下这个函数

定义 指数函数

$$\exp(z):=\sum_{n=0}^{\infty}\frac{z^n}{n!}$$

你一定知道这个函数在实数域中就是 $f(x)=e^x$，但是不要心急，在复数域中姑且写成 exp 的形式，以强调它有一些性质有别于实数域的情况，接下来看看这个函数是怎样在复数域中保持大部分实数幂运算的良好性质的：

(1) 该函数在整个复平面上都有定义.
证明：利用系数模比值法，由于 $\lim\limits_{n\to\infty}\dfrac{n!}{(n+1)!}=\lim\limits_{n\to\infty}\dfrac{1}{n+1}=0$，所以收敛半径为无穷大，即该函数项幂级数在整个复平面上一致收敛.

(2) $\exp(z)$ 满足“外相乘内相加”的特性，即 $\exp(a)\exp(b)=\exp(a+b)$.
证明：利用柯西乘积

$$\begin{aligned}\exp(a)\exp(b) &= \left(\sum_{n=0}^{\infty}\frac{a^n}{n!}\right)\left(\sum_{n=0}^{\infty}\frac{b^n}{n!}\right) = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{a^k}{k!}\frac{b^{n-k}}{(n-k)!}\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{k=0}^{n}C_n^k a^k b^{n-k} = \sum_{n=0}^{\infty}\frac{(a+b)^n}{n!}\\ &= \exp(a+b)\end{aligned}$$

(3) 不存在 z，使得 $\exp(z)=0$.
证明：对于任意 z，均有 $\exp(z)\exp(-z)=\exp(0)=1$，若 $\exp(z)$ 能等于 0，则两者乘积为 0，矛盾.

(4) $\exp(z)$ 的导数是自己，即 $\exp'(z)=\exp(z)$.
证明：直接套定义，并且利用外乘内加的性质

$$\exp'(z) := \lim_{h\to 0}\frac{\exp(z+h)-\exp(z)}{h} = \exp(z)\lim_{h\to 0}\frac{\exp(h)-1}{h} = \exp(z)$$

然而 $\exp(z)$ 也具有 e^x 不具有的新奇的性质

(1)

定理　欧拉公式 (Euler's formula)

对于任意实数 x，均有 $\exp(ix)=\sin(x)+i\cos(x)$

证明：直接带入定义就好了

$$\begin{aligned}\exp(ix) = \sum_{n=0}^{\infty}\frac{(ix)^n}{n!} &= 1+\frac{ix}{1!}-\frac{x^2}{2!}-\frac{ix^3}{3!}+\frac{x^4}{4!}+\frac{ix^5}{5!}-\frac{x^6}{6!}-\frac{ix^7}{7!}+\cdots\\ &= \left(1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots\right)+i\left(\frac{x}{1!}-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots\right)\\ &= \cos(x)+i\sin(x)\end{aligned}$$

(2) $\exp(z)$ 具有周期 $2\pi i$，即 $\exp(z+2\pi i)=\exp(z)$.
证明：$\exp(z+2\pi i)=\exp(z)\exp(2\pi i)=\exp(z)(\cos(2\pi)+i\sin(2\pi))=\exp(z)$

(3) 映射 $f(x)=\exp(ix)$ 将实轴上的点映射到复平面上的单位圆内，并且复平面单位圆上的每一点都会被映射到.
这是一条非常重要的性质，它把复数与复平面上的旋转向量结合到了一起.
证明：首先来证明 $\exp(ix)$ 的值必然在复平面的单位圆上. 证明模长为 1 即可.

$$|\exp(ix)| = \exp(ix)\overline{\exp(ix)} = [\cos(x)+i\sin(x)][\cos(x)-i\sin(x)] = \left[\sin^2(x)+\cos^2(x)\right] = 1$$

接下来证明复平面单位圆上的每一点都会被映射到. 换种说法，随意选定任意一个模长为 1 的复数 z，都存在 x 使得 $\exp(ix)=z$. 这好办，可设 $z=a+ib$，并首先讨论 $a\geq 0, b\geq 0$ 的情况. 由于 $a\leq 1$，所以存在 $0\leq x\leq \pi/2$ 使得 $\cos(x)=a$，同时根据模的计算有 $b=\sqrt{1-a^2}=\sqrt{1-\cos^2(x)}=\sin(x)$，这种情况是成立的. 如果 $a<0, b\geq 0$ 呢？那么 $-iz=-a+ib$ 就回到了前一种情况，并且易得此时 $\exp(i(x+\pi/2))=z$. 如果 $b<0$ 呢？那么 $-z=a-ib$ 就回到了前一种情况，并且此时 $\exp(i(x+\pi))=z$.

(4) (复数的指数形式) 对于任意实数 a,b，其中 $a\neq 0$，都有 $a+ib=\sqrt{a^2+b^2}\exp\left(i\arctan\frac{b}{a}\right)$

证明：代入欧拉公式即可，需要用到如下两个三角恒等式：

$$\sin(\arctan(x))=\frac{x}{\sqrt{1+x^2}},\qquad \cos(\arctan(x))=\frac{1}{\sqrt{1+x^2}}$$

$$\begin{aligned}\sqrt{a^2+b^2}\exp\left(i\arctan\frac{b}{a}\right)&=\sqrt{a^2+b^2}\left[\cos\left(\arctan\frac{b}{a}\right)+i\sin\left(\arctan\frac{b}{a}\right)\right]\\&=\sqrt{a^2+b^2}\left[\frac{a}{\sqrt{a^2+b^2}}+\frac{ib}{\sqrt{a^2+b^2}}\right]=a+ib\end{aligned}$$

由于 $|\exp(ix)|\equiv 1$，所以只要限定 a^2+b^2 为常数，那么复数 $a+ib$ 取值的集合就是一个圆心在原点、半径为 $\rho=|a+ib|$ 的圆，此时可以将复平面转化为极坐标系，复数（相当于向量）就可以由极径和极角来确定，因此可以做出以下定义。

(5)

定理　棣莫弗公式 (de Moivre's formula)

设

$$z_1=\rho_1\exp(i\mathrm{Arg}(z_1)),\qquad z_2=\rho_2\exp(i\mathrm{Arg}(z_2))$$

则

$$z_1z_2=\rho_1\rho_2\exp[i(\mathrm{Arg}(z_1)+\mathrm{Arg}(z_2))]$$

此即为：复数相乘模相乘角相加

以上给出了 $\exp(z)$ 的各种性质，并发现它是实数域上 $f(x)=e^x$ 的良好拓展，所以再也没有什么理由可以阻止我们把 $\exp(z)$ 写成 e^z 了.

在某些限定的条件下给出某样东西的定义之后，数学家们就是喜欢打破这些条件，同时让定义强行成立看看会发生什么. 定义出了复数的指数函数 $\exp(z)$ 之后，数学家们便尝试把任意的**算子** (operator) 塞入这个定义中：

定义　算子的指数函数

$$\exp(\cdot)=\sum_{n=0}^{\infty}\frac{\cdot^n}{n!}$$

1.2　对数函数

对数函数是指数函数的反函数，考虑到 $\exp(z)$ 是个周期函数，而不是单射，所以它的反函数必然是个多值函数. 设 $\exp(z+2k\pi i)=w$，则有 $\exp^{-1}(w)=z+2k\pi i$. 但是我们可以发现，辐角函数 $\mathrm{Arg}(z)$ 恰好具有 $\mathrm{Arg}(z+2k\pi i)=\mathrm{Arg}(z)$ 的性质，设 $y=|y|e^{i\mathrm{Arg}(y)}$，则 $\mathrm{Ln}(y)=\ln|y|+i\mathrm{Arg}(y)$，这就是对数函数.

定义　定义（复变对数函数）

$$\mathrm{Ln}(z) := \ln|z| + i\mathrm{Arg}(z)$$

$$\ln(z) := \ln|z| + i\mathrm{arg(z)}$$

其中等式右端的 $ln(x)$ 定义在实数域上.

考虑到指数函数的周期性，复数域上的 Ln 是多值函数，即 $\mathrm{Ln}(z+2k\pi i) = \mathrm{Ln}(z)$，而 ln 是单值函数，取前者的主值.

定义好了指数和对数函数，复数的幂也可以定义了.

定义　定义（复数的幂）

对于任意 $z, w \in \mathbb{C}$，

$$z^w := e^{w\mathrm{Ln}(z)}$$

复数的幂也保留了实数幂的绝大多数性质，但考虑到它是多值函数，所以会出现一些在实数域不可能出现的现象. 比如，在实数域中，1 的任意次方都是 1，但是在复数域中，$1^z = e^{z\mathrm{Ln}(1)} = e^{i2k\pi z}, k \in \mathbb{Z}$，只要 $k \neq 0$，即可得到非 1 的数，再调整 z 的值，可以得到 0 以外的任意复数.

当把幂结构从实数推广到复数之后，“根号”的定义需要特别留意，毕竟在实数域内，根号也不等价于 1/2 次方，而是一个特别定义的符号. 所以现在就在给复数域的根号下定义，并使它兼容实数域内根号的定义. 为避免不必要的麻烦，这里仅为平方根的根号下定义，任意次方根的根号暂不推广至复数域.

定义　复根号

暂定符号“$\sqrt[\mathbb{R}]{\ }$”表示实数域内的根号（已有良好定义），“$\sqrt[\mathbb{C}]{\ }$”表示复数域内的根号，对于任意复数 z，定义

$$\sqrt[\mathbb{C}]{z} := \sqrt[\mathbb{R}]{|z|}\exp\left(\frac{i}{2}\arg(z)\right)$$

可见复根号是单值函数，当不引起混淆时，使用“$\sqrt{\ }$”来代表“$\sqrt[\mathbb{C}]{\ }$”.

1.3 三角函数和双曲函数

同指数函数，复数域内的正弦函数和余弦函数也利用无穷级数来定义.

定义　正弦函数和余弦函数

$$\sin(z) := \sum_{n=0}^{\infty}\frac{z^{2n+1}}{(2n+1)!} \qquad \cos(z) := \sum_{n=0}^{\infty}\frac{z^{2n}}{(2n)!}$$

之所以让 $\sin(z), \cos(z)$ 如此定义，是为了使得欧拉公式在复数域内依然成立.

定理　欧拉公式 (Euler’s formula)

对于任意复数 z，都有

$$\exp(iz) = \cos(z) + i\sin(z)$$

将其显化，得到

$$\sin(z) = \frac{e^{iz} - e^{-iz}}{2i}, \qquad \cos(z) = \frac{e^{iz} + e^{-iz}}{2}$$

证明同实数域. 实际上，也可以通过复数域的欧拉公式隐式地定义正弦函数和余弦函数，换句话说，欧拉公式和以上的级数表示是等价的. 但之所以用级数来定义，是为了强调它的解析性（能展成幂级数即为解析）. 也许你会觉得欧拉公式比级数更加“自然”，那接下来就用这种“自然”的方法来定义双曲函数.

定义　双曲正弦函数和双曲余弦函数

$$\sinh(z) := \frac{e^{z} - e^{-z}}{2}, \qquad \cosh(z) := \frac{e^{z} + e^{-z}}{2}$$

当定义好了正弦和余弦函数之后，其他的三角函数和双曲函数也可以将定义域延伸至复数域了，而且它们在极点之外处处解析：

定义　其他三角函数和双曲函数

$$\begin{aligned}
\tan(z) &:= \frac{\sin(z)}{\cos(z)} & \tanh(z) &:= \frac{\sinh(z)}{\cosh(z)} \\
\cot(z) &:= \frac{\cos(z)}{\sin(z)} & \coth(z) &:= \frac{\cosh(z)}{\sinh(z)} \\
\sec(z) &:= \frac{1}{\cos(z)} & \operatorname{sech}(z) &:= \frac{1}{\cosh(z)} \\
\csc(z) &:= \frac{1}{\sin(z)} & \operatorname{csch}(z) &:= \frac{1}{\sinh(z)}
\end{aligned}$$

虽然一下子定义了 12 个函数，但它们都是只是指数函数 $\exp(z)$ 的某种简单的四则运算结果而已，因此它们和 $\exp(z)$ 也许会共享某些性质，最重要的一点是它们都保持了实数域内的很多良好性质，以下列举一二.

(1) (两角平方关系) 对于任意 z，均有

$$\begin{aligned}
\sin^2(z) + \cos^2(z) &= 1 \\
\sinh^2(z) - \cosh^2(z) &= 1
\end{aligned}$$

证明：利用欧拉公式，再用平方差公式即可

$$\begin{aligned}
\sin^2(z) + \cos^2(z) &= \left(\frac{e^{iz} - e^{-iz}}{2i}\right)^2 + \left(\frac{e^{iz} + e^{-iz}}{2}\right)^2 = \frac{1}{4}\left[(e^{iz} + e^{-iz})^2 - (e^{iz} - e^{-iz})^2\right] \\
&= \frac{1}{4}(e^{iz} + e^{-iz} + e^{iz} - e^{-iz})(e^{iz} + e^{-iz} - e^{iz} + e^{-iz}) \\
&= e^{iz - iz} = 1
\end{aligned}$$

(2) (余弦两角和公式) 对于任意 z，均有

$$\begin{aligned}
\cos(z_1 + z_2) &= \cos(z_1)\cos(z_2) - \sin(z_1)\sin(z_2) \\
\cosh(z_1 + z_2) &= \cosh(z_1)\cosh(z_2) + \sinh(z_1)\sinh(z_2)
\end{aligned}$$

证明：还是套用欧拉公式，但从右边出发推出左边

$$\begin{aligned}
\text{RHS} &= \frac{e^{iz_1}+e^{-iz_1}}{2}\frac{e^{iz_2}+e^{-iz_2}}{2} - \frac{e^{iz_1}-e^{-iz_1}}{2i}\frac{e^{iz_2}-e^{-iz_2}}{2i} \\
&= \frac{1}{4}\left[e^{i(z_1+z_2)}+e^{i(z_1-z_2)}+e^{i(-z_1+z_2)}+e^{i(-z_1-z_2)}+e^{i(z_1+z_2)}-e^{i(z_1-z_2)}-e^{i(-z_1+z_2)}+e^{i(-z_1-z_2)}\right] \\
&= \frac{e^{i(z_1+z_2)}+e^{-i(z_1+z_2)}}{2} = \cos(z_1+z_2) = \text{LHS}
\end{aligned}$$

众所周知，我们常接触的三角恒等变换公式，如诱导公式、倍角公式、升降幂公式、和差化积以及积化和差等，都是由 $\sin^2(x)+\cos^2(x)=1$ 和余弦函数的两角和差公式推出来的，因此绝大多数三角恒等变换公式在复数域内也成立．同理，双曲恒等变换公式在复数域内仍然适用．

但它们也失去了另一些良好的性质：

(1) $\sin(z),\cos(z)$ 是无界的，它们的值域为整个复数域.

证明：以 sin 为例，任取复数 w，假设存在 z 使得 $\sin(z)=w$，现在来解出这个 z．代入欧拉公式

$$w=\sin(z)=\frac{e^{iz}-e^{-iz}}{2i}$$

令 $t=e^{iz}$，则

$$2iw=t-\frac{1}{t}$$

由于 $t\neq 0$，因此可随意乘除．整理一下即得到

$$t^2-2iwt+1=0$$

这是一个一元二次方程，不必再考虑根号带来的无意义问题．解开即得到

$$t=iw\pm\sqrt{1-w^2},\qquad z=-i\mathrm{Ln}\left(iw\pm\sqrt{1-w^2}\right)$$

以上的例子告诉我们，只要在给定 w 的情况下令 $w=\sin(z)$，就可以显式地解出 z 关于 w 的表达式.尽管 z 有无数种取值，但这足以定义正弦函数的反函数了，即令

$$\mathrm{Arcsin}(z):=-i\mathrm{Ln}\left(iw\pm\sqrt{1-w^2}\right)$$

注意这是首字母大写的反正弦函数，就像首字母大写的对数函数一样是个多值函数，我们还得给它找个主值,以确定 $\arcsin(z)$ 的表达式. 首先 Ln 需要改写成 ln 是肯定的了,但正负号应如何选取呢?我们希望复数域中的 arcsin 函数能兼容实数域中的性质,导数就是一个切入点. 在实数域中,有 $\dfrac{\mathrm{d}}{\mathrm{d}x}\arcsin(x)=\dfrac{1}{\sqrt{1-x^2}}$,我们也希望在复数域也是如此.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}z}\mathrm{Arcsin}(z) &= \frac{1\pm iz/\sqrt{1-z^2}}{iz\pm\sqrt{1-z^2}} = \frac{\left(1\pm iz/\sqrt{1-z^2}\right)\left(\sqrt{1-z^2}\mp iz\right)}{\left(\sqrt{1-z^2}\pm iz\right)\left(\sqrt{1-z^2}\mp iz\right)} \\
&= \frac{1-z^2\mp z^2}{\sqrt{1-z^2}}
\end{aligned}$$

答案很明显，应该取加号，这样就得到了复数域中 $\arcsin(z)=-i\ln\left(iz+\sqrt{1-z^2}\right)$. 还没完，还可以进一步化简，代入对数函数 $\ln(z)$ 的表达式中，得到

$$\begin{aligned}
\arcsin(z) &= -i\ln\left(iz+\sqrt{1-z^2}\right) = -i\left(\ln\sqrt{z^2+1-z^2}+i\arg\left(iz+\sqrt{1-z^2}\right)\right) \\
&= \arg\left(iz+\sqrt{1-z^2}\right)
\end{aligned}$$

由于 arg 函数已经是关于复数的基本函数了，不必再解构成更基本的实反正切函数，所以复反正弦函数定义的探索到此为止，以下给出 6 个反三角函数在复数域内的定义.

定义 反三角函数

$$\arcsin(z) := -i\ln\left(\sqrt{1-z^2}+iz\right) \quad \arcsin(z) := -i\ln\left(i\sqrt{1-z^2}+z\right)$$
$$\arctan(z) := \frac{1}{2i}\ln\left(\frac{i-z}{i+z}\right) \quad \operatorname{arccot}(z) := \frac{1}{2i}\ln\left(\frac{z+i}{z-i}\right)$$
$$\operatorname{arcsec}(z) := \arccos\left(\frac{1}{z}\right) \quad \operatorname{arccsc}(z) := \arcsin\left(\frac{1}{z}\right)$$

与反三角函数相似，6 个反双曲函数[7]的定义也可用相同方法求出，根据定义可以看出，它们的定义域比仅在实数域上讨论时要广阔得多.

定义 反双曲函数

$$\operatorname{arsinh}(z) := \ln\left(\sqrt{z^2+1}+z\right) \quad \operatorname{arcosh}(z) := \ln\left(\sqrt{z^2-1}+z\right)$$
$$\operatorname{artanh}(z) := \frac{1}{2}\ln\left(\frac{1+z}{1-z}\right) \quad \operatorname{arcoth}(z) := \frac{1}{2}\ln\left(\frac{z+1}{z-1}\right)$$
$$\operatorname{arsech}(z) := \operatorname{arcosh}\left(\frac{1}{z}\right) \quad \operatorname{arcsch}(z) := \operatorname{arsinh}\left(\frac{1}{z}\right)$$

2 幂级数专题

2.1 形式幂级数

为什么是“形式”幂级数？幂级数的重点在于级数，而级数是要求收敛的. 形式幂级数具有幂级数的结构，但是不要求其收敛，例如

$$f(x) = 1 - 3x + 5x^2 - 7x^4 + 9x^6 - 11x^8 + \cdots$$

从幂级数的角度来说，该级数的收敛半径是 1，而且收敛到某个超几何函数上. 但形式幂级数可不管这么多，它就是抽象出来的、可以进行某些运算的结构，可以把上式写作 $[1,-3,5,-7,\cdots]$，以强调只关注形式幂级数的系数. 又例如 $[1,1,2,6,24,\cdots,n!,\cdots]$，如果从幂级数的角度来关注它，就会发现它在 $x=0$ 意外的地方都不收敛，但是这没关系，它照样可以进行某些运算. 把幂级数放开到“形式”幂级数，可为接下来讨论拉格朗日反演定理带来便利.

严格来说，形式幂级数只定义了 3 种运算：取系数、加法和乘法. 但是我们可以推导出它的减法、除法和复合.

[7]双曲函数及其反函数在不同的地方记法可能不同，例如有的地方把双曲正弦函数记作 $\operatorname{sh}(z)$，反双曲正弦函数记作 $\operatorname{arcsinh}(z)$ 或 $\operatorname{arsh}(z)$，这里按照国际标准 ISO 80000-2 中的记号，上文反三角函数的前缀”arc” 表示“弧”，反双曲函数的前缀”ar” 表示“面积”（area）.

定义 形式幂级数的运算公理

给定两个形式幂级数：$A(x)=\sum_{n=0}^{\infty}a_nx^n$，$B(x)=\sum_{n=0}^{\infty}b_nx^n$，定义：

加法：$A(x)+B(x):=\sum_{n=0}^{\infty}(a_n+b_n)x^n$

乘法：$A(x)\cdot B(x):=\sum_{n=0}^{\infty}\left(\sum_{m=0}^{n}a_mb_{n-m}\right)x^n$

取系数：$[x^n]A(x):=a_n$

接下来推导形式幂级数的倒数. 令 $A(x)=\sum_{n=0}^{\infty}a_nx^n$，假设 $A(x)$ 的倒数也能展开成形式幂级数，即 $A(x)^{-1}=\sum_{n=0}^{\infty}b_nx^n$，自然有 $A(x)A(x)^{-1}=1$，然而根据形式幂级数的乘法：

$$A(x)A(x)^{-1}=\left(\sum_{n=0}^{\infty}a_nx^n\right)\left(\sum_{n=0}^{\infty}b_nx^n\right)=\sum_{n=0}^{\infty}\left(\sum_{m=0}^{n}a_mb_{n-m}\right)x^n$$

对比右侧 x^n 的系数，得到 $[x^0]\left[A(x)A(x)^{-1}\right]=1,[x^n]\left[A(x)A(x)^{-1}\right]=0$，因此

$$\begin{aligned}a_0b_0=1&\Longrightarrow b_0=\frac{1}{a_0}\\ a_0b_1+a_1b_0=0&\Longrightarrow b_1=-\frac{a_1b_0}{a_0}\\ a_0b_2+a_1b_1+a_2b_0=0&\Longrightarrow b_2=-\frac{a_1b_1+a_2b_0}{a_0}\end{aligned}$$

最后得到

$$b_n=-\frac{1}{a_0}\sum_{i+j=n}a_ib_j$$

如何刻画一个函数？解析式直接表达当然是其中一种方法，但未必是唯一的一种. 例如 n 次多项式函数在复数域有 n 个零点，这 n 个零点就唯一地刻画了一个高次多项式函数；若一个函数在某个区间内有任意高阶导数，则区间内内任意一点的高阶导数值就唯一地刻画了一个这样的函数. 函数展开的目的就在于把满足某些条件的函数（无论它有多复杂）表示为有限或无穷多个简单部分的基本运算，展开的思路有很多种，其思想大多都是找到一个数列 $\{a_n\}_{n=0}^{\infty}$，使得该数列能承载被展开的函数的全部信息.

2.2 渐进展开

同正交展开，渐进展开(asymptotic expansion) 是一种将函数展开的思想. 傅里叶级数是典型的正交展开，其特征是部分和在整个定义域上“贴近原函数的程度差不多”；而泰勒级数和洛朗级数就是典型的渐进展开，其特征是选取一个极限点 x_0，x 越接近 x_0，部分和就越贴近原函数.

如果在极限点 L 处（L 可能是无穷远点），存在一系列函数 $\{\varphi_n(x)\}_{n=0}^{\infty}$，使得

$$\lim_{x\to L}\frac{\varphi_{n+1}(x)}{\varphi_n(x)}=0 \quad 或写作 \quad \varphi_{n+1}(x)=o(\varphi_n(x))\ (x\to L)$$

则这一系列的函数可以作为渐进程度的标尺(asymptotic scale)，n 越大，在 L 的某个邻域内 $\varphi_n(x)$ 就变化得越慢.

定义 渐进展开

设被展开的函数为 $f(x)$，极限点为 L. 若一系列函数 $\{\varphi_n(x)\}_{n=0}^{\infty}$ 可以作为渐进尺度的标尺，且存在一系列常数 $\{a_n\}_{n=0}^{+\infty}$，使得

$$\lim_{x\to L}\frac{f(x)-\sum\limits_{n=0}^{N}a_n\varphi_n(x)}{\varphi_N(x)}=0 \quad \text{或写作} \quad f(x)-\sum_{n=0}^{N}a_n\varphi_n(x)=o(\varphi_N(x))\ (x\to L)$$

则称幂级数（不一定收敛）$\sum\limits_{n=0}^{\infty}a_n\varphi_n(x)$ 为 $f(x)$ 的渐近级数(asymptotic series)，记作

$$f(x)\sim\sum_{n=0}^{\infty}a_n\varphi_n(x) \quad (x\to L)$$

一般来说，当 $L=0$ 时，常选 $\varphi_n(x)=x^n$，此时展开式即为麦克劳林级数；L 为任意复数 x_0 时，常选 $\varphi_n(x)=(x-x_0)^n$，此时展开式为一般的泰勒级数或洛朗级数；当 $L=\infty$ 时，常选 $\varphi_n(x)=x^{-n}$，此时展开式即为无穷远处的洛朗级数.

尤其需要注意的是 $L=\infty$ 的情况，因为这时级数可能不收敛. 如果 $f(z)$ 在某个圆域之外 $|z|>R$ 之外解析，则可以展开成洛朗级数.

2.3 泰勒展开

2.3.1 泰勒定理

一次函数没有极值，二次函数有一个极值，三次函数至多有两个极值，随着次数的增加，极值和拐点的个数可以任意多甚至是可数无穷多个，函数图像可以任意弯曲，由此可以以任意高的精确度逼近任何一个连续函数. 在连续函数上取一个点作为起点，用高次多项式函数逼近连续函数，就是泰勒公式的思想，有一个定理保证了可行性.

定理 魏尔斯特拉斯逼近定理 (Stone-Weierstrass theorem)

若 $f(z)$ 是在区间 $[a,b]$ 上连续的复函数，则存在一个多项式序列 $\{P_n\}_{n=1}^{\infty}$，使得 $\lim\limits_{n\to\infty}P_n(z)$ 一致收敛于 $f(z)$.

假设可能实现这样的逼近，那么对于任意一个连续函数 $f(x)$，取它定义域内的一点 x_0，设高次多项式 $p_n(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$，按照渐进展开的思想，当 $x\to x_0$ 时，我们希望 $f(x)-p_n(x)=o(p_n(x))$，如果能达成 $f(x)-p_n(x)=a_{n+1}x^{n+1}$ 就更好了，这样此时式子就可以改写为

$$\lim_{x\to x_0}\frac{f(x)-p_n(x)}{a_{n+1}x^{n+1}}=1$$

不必讨论 $p_n(x)$ 的唯一性，毕竟我们的目标只是找到这样一个实用的高次多项式。

- 当 $n=0$ 时，$p_0(x)=a_0$，为了使 $\lim\limits_{x\to x_0}\dfrac{f(x)}{a_0}=1$，自然有 $a_0=f(x_0)$.
- 当 $n=1$ 时，为了使 $\lim\limits_{x\to x_0}\dfrac{f(x)-a_0}{a_1x}=1$，运用洛必达法则，得到 $\lim\limits_{x\to x_0}\dfrac{f'(x)}{a_1}=1$，所以 $a_1=f'(x_0)$.
- 当 $n=2$ 时，为了使 $\lim\limits_{x\to x_0}\dfrac{f(x)-a_0-a_1x}{a_2x^2}=1$，洛！得到 $\lim\limits_{x\to x_0}\dfrac{f''(x)}{2a_2}=1$，所以 $a_2=\dfrac{f''(x_0)}{2}$.

- 当 $n=3$ 时，为了使 $\lim\limits_{x\to x_0}\dfrac{f(x)-a_0-a_1x-a_2x^2}{a_3x^3}=1$，洛! 得到 $\lim\limits_{x\to x_0}\dfrac{f'''(x)}{6a_2}=1$，所以 $a_3=\dfrac{f'''(x_0)}{6}$.
- 由此推广到更大的 n 值，为了使 $\lim\limits_{x\to x_0}\dfrac{f(x)-p_n(x)}{a_{n+1}x^{n+1}}=1$，运用 $n+1$ 次洛必达法则，洛到函数祖坟上，得到 $\lim\limits_{x\to x_0}\dfrac{f^{(n)}(x)}{n!a_n}=1$，即 $a_n=\dfrac{f^{(n)}(x_0)}{n!}$，这样就得到了

定理　一元函数的泰勒定理 (Taylor's theorem)

若函数在 x_0 处解析，且存在 n 阶导数，则

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n(x)$$

其中 $R_n(x)$ 是泰勒公式的 n 阶余项，它有几种形式:

(1) 皮亚诺余项 (Peano remainder)

$$R_n(x)=o(x^n)$$

(2) 拉格朗日余项 (Lagrange remainder)

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1},\quad \xi\text{介于}x_0\text{和}x\text{之间}$$

(3) 柯西余项 (Cauchy remainder)

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n(x-x_0),\quad \xi\text{介于}x_0\text{和}x\text{之间}$$

(4) 施莱米尔奇余项 (Schlömilch remainder)

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{n!p}(x-x_0)^{n+1-p}(x-x_0)^p,\quad \xi\text{介于}x_0\text{和}x\text{之间}$$

(5) 积分余项 (integral remainder)

$$R_n(x)=\int_{x_0}^{x}f^{(n+1)}(\xi)\frac{(x-\xi)^n}{n!}\mathrm{d}\xi$$

如果 $x_0=0$，则泰勒公式又称为麦克劳林公式.

2.3.2 积分余项和列表积分公式

泰勒定理的积分余项可以明确地给定余项表达式，而且它的推导也非常简洁. 在接下来的推导中，还可以顺便推出泰勒定理. 首先改写 $f(x)$.

$$f(x)=f(x_0)+f(x)-f(x_0)=f(x_0)+\int_{x_0}^{x}f'(t)\mathrm{d}t$$

由于 x 是一个与 t 无关的量，所以 $\mathrm{d}t=\mathrm{d}(t-x)$，然后使用分部积分，让 $f'(t)$ 为被积函数，把其他部分放到微分号后面.

$$\int_{x_0}^{x}f'(t)\mathrm{d}(t-x)=f'(t)(t-x)\Big|_{t=x_0}^{t=x}-\int_{x_0}^{x}f''(t)(t-x)\mathrm{d}t$$

这时就有

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\int_{x_0}^{x}f''(t)(x-t)\mathrm{d}t$$

对其中的积分项再使用分部积分，让 $f''(t)$ 为被积函数，其他部分放到微分号后面.

$$\int_{x_0}^{x} f''(t)(x-t)\mathrm{d}t = \int_{x_0}^{x} f''(t)\mathrm{d}\frac{(t-x)^2}{2} = f''(t)\frac{(t-b)^2}{2}\Big|_{t=x_0}^{t=x} + \int_{x_0}^{x} f'''(t)\mathrm{d}\frac{(t-b)^2}{2}\mathrm{d}t$$

这时就有

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2}(x-x_0)^2 + \int_{x_0}^{x} f'''(t)\frac{(x-t)^2}{2}\mathrm{d}t$$

重复上述过程，用分部积分法对操作下列积分，让 $f^{(n)}(t)$ 为被积函数，把其他部分放到微分号后面.

$$\int_{x_0}^{x} f^{(n)}(t)\frac{(x-t)^{n-1}}{(n-1)!}\mathrm{d}t = \int_{x_0}^{x} f^{(n)}(t)\mathrm{d}\left(-\frac{(x-t)^n}{n!}\right) = f^{(n)}(t)\frac{(x-t)^n}{n!}\Big|_{t=x}^{t=x_0} + \int_{x_0}^{x} f^{(n+1)}(t)\frac{(x-t)^n}{n!}\mathrm{d}t$$

代入原式，就得到了

定理　带积分余项的泰勒定理

若函数 $f(x)$ 在 $x=x_0$ 的某个邻域 U 内有 $n+1$ 阶连续导数，则对于任意 $x\in U$，有

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \int_{x_0}^{x} f^{(n+1)}(t)\frac{(x-t)^n}{n!}\mathrm{d}t$$

以上的推导过程非常具有启发性，原来分部积分法还可以这样用！回想分部积分法最初是用来解决两函数相乘的积分问题，它由乘积的导数推出：

$$(uv)' = u'v + uv' \longrightarrow uv = \int u\mathrm{d}v + \int v\mathrm{d}u \longrightarrow \int f(x)g(x)\mathrm{d}x = F(x)g(x) - \int F'(x)g(x)\mathrm{d}x$$

如果再使用一次分部积分呢？为方便起见，把 $f(x)$ 和 $g(x)$ 简记为 f 和 g，$f(x)$ 求 1 次积分得到的函数记为 $f^{(-1)}$（不需要补上任意常数 C），求 2 次积分得到的函数记为 $f^{(-2)}$，以此类推，那么分部积分法就可以写成

$$\int fg\mathrm{d}x = f^{(-1)}g - \int f^{(-1)}g^{(1)}\mathrm{d}x$$

再用一次分部积分，得到

$$\int fg\mathrm{d}x = f^{(-1)}g - f^{(-2)}g^{(1)} + \int f^{(-1)}g^{(1)}\mathrm{d}x$$

重复上述过程，就可以得到

定理　列表积分公式

设函数 $f(x), g(x)$ 均有 n 阶连续导数和反导数，则

$$\int fg\mathrm{d}x = f^{(-1)}g - f^{(-2)}g^{(1)} + f^{(-3)}g^{(2)} + \cdots + (-1)^{n-1}f^{(-n)}g^{(n-1)} + (-1)^n\int f^{(-n)}g^{(n)}\mathrm{d}x$$

列表积分公式是分部积分的扩展，适用于如下两种情况

(1) g 经过有限次求导后变为 0，例如多项式；

(2) fg 经过有限次操作后能变为原来的式子（的倍数），即积分和求导得到的函数形式存在循环，例如 $e^{\alpha x}\cos(\beta x)$.

举两个例子，用列表积分法可以口算出答案，首先是第一种情况

$$\int e^{2x}(3x^2+2x)\mathrm{d}x = \frac{1}{2}e^{2x}(3x^2+2x) - \frac{1}{4}e^{2x}(6x+2) + \frac{3}{4}e^{2x} + C = \frac{1}{4}e^{2x}(6x^2-2x+1) + C$$

其次是第二种情况

$$\int e^{2x}\sin(3x)\mathrm{d}x=\frac{1}{2}e^{2x}\sin(3x)-\frac{3}{4}e^{2x}\cos(3x)-\frac{9}{4}\int e^{2x}\sin(3x)\mathrm{d}x$$

得到

$$\int e^{2x}\sin(3x)\mathrm{d}x=\frac{1}{13}\left(2e^{2x}\sin(3x)-3e^{2x}\cos(3x)\right)+C$$

2.3.3 泰勒级数

泰勒定理是用有限高次的多项式来逼近连续函数，泰勒级数则是让连续函数严格地等于某个幂级数，即把余项 R_n 改成省略号:

$$\begin{aligned}f(x)&\sim e^{(x-x_0)\frac{\partial}{\partial x}}f(x_0)=\sum_{n=0}^{\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n\\&=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+\cdots\end{aligned}$$

从有限扩展到无限的过程可能会丧失一些好的性质，泰勒定理的适用条件是函数在某邻域内有若干阶导数，但是想把函数展成泰勒级数，不但要保证有任意高阶导数，还要保证 $\lim_{x\to x_0}R_n(x)\to 0$ 才行，这就需要考虑收敛域. 当 $x_0=0$ 时，泰勒级数又称为麦克劳林级数(Maclaurin series)，此时 $f(x)\sim e^{x\frac{\partial}{\partial x}}f(0)$.

2.3.4 快查表：常用的麦克劳林级数

首先是指数函数和对数函数的. 如 2.1 节所述，适用于整个复数域的指数函数的麦克劳林展开式实际上是它的定义.

$$
\begin{aligned}
e^z &= \sum_{n=0}^{\infty}\frac{1}{n!}z^n &&= 1+z+\frac{1}{2!}z^2+\frac{1}{3!}z^3+\cdots && z\in\mathbb{C}\\
\ln(1+z) &= \sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}z^{n+1} &&= z-\frac{1}{2}z^2+\frac{1}{3}z^3-\frac{1}{4}z^4+\cdots && |z|<1\text{或}z=1
\end{aligned}
$$

$(1+x)^\alpha$ 的展开也很有来头，将通式以及常用的几个 α 列举如下，为了查阅方便，一并给出洛朗级数. 其中牛顿二项式系数 C_α^m 是组合数 C_n^m 的推广，它允许 n 不是整数. $(1+x)^\alpha$ 展开式的特点就是它的麦克劳林级数和洛朗级数的适用范围恰好覆盖整个定义域（除了 $x=1$ 处）

$$
\begin{aligned}
(1+z)^\alpha &= \sum_{n=0}^{\infty}C_\alpha^n z^n &&= 1+\alpha z+C_\alpha^2z^2+C_\alpha^3z^3+\cdots && |z|<1\\
(1+z)^\alpha &= \sum_{n=0}^{\infty}C_\alpha^n z^{n+\alpha} &&= z^\alpha+\alpha z^{1+\alpha}+C_\alpha^2z^{2+\alpha}+C_\alpha^3z^{3+\alpha}+\cdots && |z|>1\\
\frac{1}{1+z} &= \sum_{n=0}^{\infty}(-1)^n z^n &&= 1-z+z^2-z^3+\cdots && |z|<1\\
\frac{1}{1+z} &= \sum_{n=0}^{\infty}(-1)^n z^{-n-1} &&= z^{-1}+z^{-2}-z^{-3}+z^{-4}-\cdots && |z|>1\\
\frac{1}{1-z} &= \sum_{n=0}^{\infty}z^n &&= 1+z+z^2+z^3+\cdots && |z|<1\\
\frac{1}{1-z} &= \sum_{n=0}^{\infty}-z^{-n-1} &&= -z^{-1}-z^{-2}-z^{-3}-z^{-4}-\cdots && |z|>1
\end{aligned}
$$

接下来是 6 个三角函数的麦克劳林级数或洛朗级数，其中 B_n 表示第 n 个伯努利数，E_n 表示第 n 个欧拉数，后面将会介绍. 不同于 $(1+x)^\alpha$ 的展开式，由于 $\tan,\cot,\sec,\csc$ 具有周期性，而且有无穷多个极点，所以麦克劳林级数和洛朗级数都只能在一个连续区间内生效，后面将要介绍的极点展开就很好地解决了这个问题.

$$
\begin{aligned}
\sin(z) &= \sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)!}z^{2n+1} &&= z-\frac{1}{3!}z^3+\frac{1}{5!}z^5-\frac{1}{7!}z^7+\cdots && z\in\mathbb{C}\\
\cos(z) &= \sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}z^{2n} &&= 1-\frac{1}{2!}z^2+\frac{1}{4!}z^4-\frac{1}{6!}z^6+\cdots && z\in\mathbb{C}\\
\tan(z) &= \sum_{n=1}^{\infty}\frac{(4^n-1)B_{2n}}{4^{-n}(2n)!}z^{2n-1} &&= z+\frac{1}{3}z^3+\frac{2}{15}z^5+\frac{17}{315}z^7+\cdots && |z|<\frac{\pi}{2}\\
\cot(z) &= \sum_{n=0}^{\infty}\frac{B_{2n}2^{2n}}{(-1)^n(2n)!}z^{2n-1} &&= \frac{1}{z}-\frac{1}{3}z-\frac{1}{45}z^3-\frac{2}{945}z^5-\cdots && |z|<\pi\\
\sec(z) &= \sum_{n=0}^{\infty}\frac{(-1)^nE_{2n}}{(2n)!}z^{2n} &&= 1+\frac{1}{2}z^2+\frac{5}{24}z^4+\frac{61}{720}z^6+\cdots && |z|<\frac{\pi}{2}\\
\csc(z) &= \sum_{n=0}^{\infty}\frac{(2-2^{2n})B_{2n}}{(-1)^n(2n)!}z^{2n-1} &&= \frac{1}{z}+\frac{1}{6}z+\frac{7}{360}z^3+\frac{31}{15120}z^5+\cdots && |z|<\pi
\end{aligned}
$$

接下来是双曲函数，它们的展开式系数和对应的三角函数正好差了因式 $(-1)^n$. 注意欧拉数的生成函数就是 $\operatorname{sech}(z)$，所以这个展开式只是把欧拉数的定义抄了一遍.

$$\begin{array}{rclcll}
\sinh(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{1}{(2n+1)!}z^{2n+1} & = & z+\frac{1}{3!}z^3+\frac{1}{5!}z^5+\frac{1}{7!}z^7+\cdots & z\in\mathbb{C} \\
\cosh(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{1}{(2n)!}z^{2n} & = & 1+\frac{1}{2!}z^2+\frac{1}{4!}z^4+\frac{1}{6!}z^6+\cdots & z\in\mathbb{C} \\
\tanh(z) & = & \displaystyle\sum_{n=1}^{\infty}\frac{(4^n-1)B_{2n}}{(-4)^n(2n)!}z^{2n-1} & = & z-\frac{1}{3}z^3+\frac{2}{15}z^5-\frac{17}{315}z^7+\cdots & |z|<\frac{\pi}{2} \\
\coth(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{B_{2n}2^{2n}}{(2n)!}z^{2n-1} & = & \frac{1}{z}-\frac{1}{3}z+\frac{1}{45}z^3-\frac{2}{945}z^5+\cdots & |z|<\pi \\
\operatorname{sech}(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{E_{2n}}{(2n)!}z^{2n} & = & 1-\frac{1}{2}z^2+\frac{5}{24}z^4-\frac{61}{720}z^6+\cdots & |z|<\frac{\pi}{2} \\
\operatorname{csch}(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{(2-2^{2n})B_{2n}}{(2n)!}z^{2n-1} & = & \frac{1}{z}-\frac{1}{6}z+\frac{7}{360}z^3-\frac{31}{15120}z^5+\cdots & |z|<\pi
\end{array}$$

接下来是反三角和反双曲函数，其中 $(x)^{(n)}=x(x+1)(x+2)\cdots(x+n-1)$ 是上升阶乘幂，是阶乘的推广. 以下给出两组，另外的三角或双曲函数的展开式可以根据恒等变换获得.

$$\begin{array}{rclcll}
\arcsin(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{(\frac{1}{2})^{(n)}}{2n+1}z^{2n+1} & = & z+\frac{1}{6}z^3+\frac{3}{40}z^5+\frac{5}{112}z^7+\cdots & |z|<1 \\
\arctan(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}z^{2n+1} & = & z-\frac{1}{3}z^3+\frac{1}{5}z^5-\frac{1}{7}z^7+\cdots & |z|<1\text{或}z=\pm1 \\
\operatorname{arsinh}(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{(-1)^n(\frac{1}{2})^{(n)}}{2n+1}z^{2n+1} & = & z-\frac{1}{6}z^3+\frac{3}{40}z^5-\frac{5}{112}z^7+\cdots & |z|<1 \\
\operatorname{artanh}(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{1}{2n+1}z^{2n+1} & = & z+\frac{1}{3}z^3+\frac{1}{5}z^5+\frac{1}{7}z^7+\cdots & |z|\le1\text{且}z\ne\pm1
\end{array}$$

接下来是一些特殊函数：朗博 W 函数的主值、归一化的误差函数、0 阶第一类贝塞尔函数和第一类完全椭圆积分.

$$\begin{array}{rclcll}
W_0(z) & = & \displaystyle\sum_{n=1}^{\infty}(-n)^{n-1}\frac{z^n}{n!} & = & z-z^2+\frac{3}{2}z^3-\frac{8}{3}z^4+\cdots & |z|<e^{-1} \\
\frac{\sqrt{\pi}}{2}\operatorname{erf}(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{(-1)^nz^{2n+1}}{n!(2n+1)} & = & z-\frac{1}{3}z^3+\frac{1}{10}z^5-\frac{1}{42}z^7+\cdots & |z|<1 \\
J_0(z) & = & \displaystyle\sum_{n=0}^{\infty}\frac{(-1)^n}{4^n(n!)^2}z^{2n} & = & 1-\frac{1}{4}z^2+\frac{1}{64}z^4-\frac{1}{2304}z^6+\cdots & |z|\in\mathbb{C} \\
K(z) & = & \displaystyle\frac{\pi}{2}\sum_{n=0}^{\infty}\left(\frac{(2n)!}{2^{2n}(n!)^2}\right)^2z^{2n} & = & \frac{\pi}{2}+\frac{\pi z}{8}+\frac{9\pi z^2}{128}+\frac{25\pi z^3}{512}+\cdots & |z|\le1
\end{array}$$

2.4 洛朗级数

带有极点的函数在极点处无法进行泰勒展开，例如 $y=e^{\frac{1}{x^2}}$，在 $x=0$ 处无法进行泰勒展开，但是如果把它强行代入泰勒级数中，可以发现该级数在定义域内都收敛，只不过 x 的次数会出现负数，因此可以大胆地改写泰勒级数的形式，得到双边幂级数(bilateral power series)：

$$f(z)=\sum_{n=0}^{\infty}a_n(z-z_0)\Longrightarrow f(z)\sim\sum_{n=-\infty}^{\infty}a_n(z-z_0)$$

这相当于把泰勒级数中 $z-z_0$ 的次数从 $\mathbb{N}$ 扩展到了 $\mathbb{Z}$，其中次数为负数的部分称为主部(principal part)，记为 $P(z)$，剩下的部分即为泰勒级数，姑且记作 $T(z)$，此时这个级数就可以写成：

$$f(z)\sim P(z)+T(z),\qquad P(z)=\sum_{n=-\infty}^{-1}a_n(z-z_0)^n,\qquad T(z)=\sum_{n=0}^{\infty}a_n(z-z_0)^n$$

这里使用了符号“$\sim$”而不是等于号，非常令人不满，因为把次数扩展到了负数之后，级数的收敛域可能就变了. 这好办，做代换，令 $\zeta=(z-z_0)^{-1}$，这样就有

$$P(z)=\sum_{n=1}^{\infty}a_n\zeta^n$$

并设其收敛域为 $|\zeta|<1/r$，这样就得到了 $P(z)$ 的收敛域 $|z-z_0|>r$，如果 $T(z)$ 的收敛域为 $|z-z_0|<R, R>r$，那么整个级数的收敛域就为 $r<|z-z_0|<R$，此时就可以欢快地把“$\sim$”换成等号了.

在收敛域内任取一点 s，并做两个半径不等的圆环 $C_1:|z-z_0|=r_1$ 和 $C_2:|z-z_0|=r_2$ 夹住点 s，其中 $r_1<r_2$，根据复连通域的柯西积分公式[8]，可得

$$f(s)=\frac{1}{2\pi i}\oint_{C_1^-}\frac{f(z)}{z-s}\mathrm{d}z+\frac{1}{2\pi i}\oint_{C_2^+}\frac{f(z)}{z-s}\mathrm{d}z$$

其中的第一个积分（沿着内环负方向的积分）

$$\frac{1}{2\pi i}\oint_{\color{red}{C_1^-}}\frac{f(z)}{z-s}\mathrm{d}z=\frac{1}{2\pi i}\oint_{\color{red}{C_1^+}}\frac{f(z)}{(s-z_0)-(z-z_0)}\mathrm{d}z=\frac{1}{2\pi i}\oint_{C_1^+}\frac{f(z)}{s-z_0}\frac{\mathrm{d}z}{1-\frac{z-z_0}{s-z_0}}$$

由于 z 在内环上，s 在内外环之间，因此 $s>z$，即 $\dfrac{z-z_0}{s-z_0}<1$，所以可以使用泰勒展开

$$\frac{1}{1-\frac{z-z_0}{s-z_0}}=\sum_{n=0}^{\infty}\left(\frac{z-z_0}{s-z_0}\right)^n=\sum_{n=1}^{\infty}\left(\frac{z-z_0}{s-z_0}\right)^{n-1}$$

回代，得到

$$\frac{1}{2\pi i}\oint_{C_1^-}\frac{f(z)}{z-s}\mathrm{d}z=\frac{1}{2\pi i}\oint_{C_1^+}\frac{f(z)}{s-z_0}\sum_{n=1}^{\infty}\left(\frac{z-z_0}{s-z_0}\right)^{n-1}\mathrm{d}z=\sum_{n=1}^{\infty}\color{red}{\left[\frac{1}{2\pi i}\oint_{C_1^+}\frac{f(z)}{(z-z_0)^{1-n}}\mathrm{d}z\right]}(s-z_0)^{-n}=P(z)$$

然后对于第二个积分（沿着外环的积分），也用同样的方法讨论

$$\frac{1}{2\pi i}\oint_{C_2^+}\frac{f(z)}{z-s}\mathrm{d}z=\frac{1}{2\pi i}\oint_{C_2^+}\frac{f(z)}{z-z_0}\frac{\mathrm{d}z}{1-\frac{z-z_0}{s-z_0}}=\sum_{n=0}^{\infty}\color{red}{\left[\frac{1}{2\pi i}\oint_{C_2^+}\frac{f(z)}{(z-z_0)^{1+n}}\mathrm{d}z\right]}(s-z_0)^n=T(z)$$

若令 C_1,C_2 圆环重叠于 s 点，则 $T(z)$ 和 $P(z)$ 可以合并，就得到了

[8]复连通域的柯西积分公式：若 Ω 是一个复连通区域，其边界 $\partial\Omega=C^++C_1^-+C_2^-+\cdots+C_n^-$ 是分段光滑闭曲线，函数 $f(z)$ 在 Ω 内解析，则

$$\int_{\partial\Omega}f(z)\mathrm{d}z=0$$

定理 洛朗级数 (Laurent series)

如果函数 $f(z)$ 在圆环域 $r < |z-z_0| < R$ 内解析，则该函数可以展开为双边幂级数

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z-z_0)^n$$

其中

$$a_n = \frac{1}{2\pi i} \oint_{C^+} \frac{f(z)}{(z-z_0)^{n+1}} \mathrm{d}z$$

C 为圆环内绕 z_0 的任意闭曲线.

这个公式很少用到，但它为我们之前对与泰勒公式的大胆套用提供了坚实的理论基础.

2.5 反函数的幂级数展开

前几节探索泰勒公式和洛朗级数时，很快就推出了泰勒定理和洛朗级数表达式，之后好几节都是它们的应用. 这一节则不同，首先需要引入较多相关概念，比如卷绕数和对数留数等，然后再将其推广，之后才能导出最终的拉格朗日反演定理.

2.5.1 卷绕数和对数留数

通俗理解，用笔在纸上随意画一个首尾相接的曲线（一笔画完），画出的曲线可能会绕某个点转好几圈，这个“圈数”就称为该曲线对于这个点的卷绕数(winding number).

现在用数学的语言把卷绕数的概念重新叙述一遍. 假设你画线的那张纸是复平面，记该简单闭曲线为 C，记你开始画线的时间为 $t=0$，画完的时间 $t=1$，这个曲线围绕点 z_0 转了若干圈（可以一圈也不转），画成的曲线方程便可以按极坐标的思路写作

$$C(t) = z_0 + r(t)e^{i\theta(t)}, \qquad 0 \le t \le 1$$

通俗理解，$r(t)$ 是 t 时刻笔尖离 z_0 的距离，$\theta(t)$ 则是笔尖相对 z_0 的辐角. 由于辐角每变化 2π，就表明曲线绕 z_0 转了一圈，所以曲线的卷绕数为

$$\mathrm{Ind}_C(z_0) = \frac{\theta(1)-\theta(0)}{2\pi}$$

改写解析式，分离出 $\theta(t)$，得到 $\theta(t) = \ln(C(t)-z_0) - \ln r(t)$，回代

$$\begin{aligned}\mathrm{Ind}_C(z_0) &= \frac{1}{2\pi i}\left[\left[\ln(C(1)-z_0) - \ln r(1)\right] - \left[\ln(C(0)-z_0) - \ln r(0)\right]\right] \\ &\xlongequal{r(1)=r(0)} \frac{1}{2\pi i}\left[\ln(C(1)-z_0) - \ln(C(0)-z_0)\right] \\ &= \int_0^1 \frac{C'(t)}{C(t)-z_0}\mathrm{d}t = \oint_C \frac{\mathrm{d}z}{z-z_0}\end{aligned}$$

这就得到了卷绕数的定义

定义 卷绕数

连续的闭曲线 C 对点 z_0 的卷绕数(winding number) 定义为

$$\mathrm{Ind}_C(z_0) = \oint_C \frac{\mathrm{d}z}{z-z_0}$$

来套一层娃，之前在定义卷绕数时，自变量是时间 t，随着 t 的变化笔尖在纸上画出一个闭曲线 C；现在假设自变量是简单闭曲线 C，围绕的点是原点，C 上的点经过半纯函数 $f(z)$ 映射为另一个简单闭曲线，那么 $f(z)$ 对于 C 的卷绕数为

$$\mathrm{Ind}=\frac{1}{2\pi}\Delta\arg[f(z)]=\frac{1}{2\pi i}\oint_C\frac{\mathrm{d}f(z)}{f(z)}=\frac{1}{2\pi i}\oint_C\frac{f'(z)}{f(z)}\mathrm{d}z$$

设 $f(z)$ 在简单闭区域 D 内为半纯函数，零点为 $z_1,z_2,\cdots,z_m$，重数分别为 m_i；极点为 $p_1,p_2,\cdots,p_n$，重数分别为 n_i，且都不是本性奇点. 则存在 D 内的全纯函数，使得

$$f(z)=g(z)\prod_{i=1}^{m}(z-z_i)^{m_i}\prod_{i=1}^{n}(z-p_i)^{-n_i}$$

所以

$$\frac{f'(z)}{f(z)}\mathrm{d}z=\mathrm{d}\ln f(z)=\frac{g'(z)}{g(z)}\mathrm{d}z+\sum_{i=1}^{m}\frac{m_i}{z-z_i}\mathrm{d}z-\sum_{i=1}^{n}\frac{n_i}{z-p_i}\mathrm{d}z$$

注意等号右边这三项，根据柯西-古萨定理，$g'(z)/g(z)$ 全纯所以积分为 0；而根据柯西积分公式，$m_i/(z-z_i)$ 积分为 m_i，因此

$$\mathrm{Ind}=\frac{1}{2\pi}\Delta\arg[f(z)]=\frac{1}{2\pi i}\oint_C\frac{\mathrm{d}f(z)}{f(z)}=\sum_{i=1}^{m}m_i-\sum_{i=1}^{n}n_i$$

这就推出了辐角原理(argument principle).

定理　辐角原理

若 $f(z)$ 在简单闭区域 D 内除了有限个极点以外解析，并有至多可数无穷多个零点，则该函数在 D 的边界曲线 ∂D 上的卷绕数等于零点个数减去极点个数（零点和极点按照重数和阶数来算），即

$$\mathrm{Ind}=\frac{1}{2\pi i}\oint_{\partial D}\mathrm{d}\ln f(z)=N-P$$

其中 N 表示 D 内零点的个数，P 表示 D 内极点的个数，注意这里的零点和极点要按照阶数来算，例如 n 级零点要看做 n 个零点.

2.5.2　一些定理的推广

本节来推广 3 个定理：辐角原理、柯西积分公式和最终的泰勒公式.

首先来推广辐角原理. 接上节，$f(z)$ 在 D 上的若干零点和极点在 $\ln f(z)$ 面前都变成了极点. 根据复连通域的柯西积分公式，可以在 $\ln\psi(z)$ 的每个极点周围作小圆周，使得每个圆周只包含一个极点，并把包含 p_i 的圆周记为 (p_i)，z_i 类似，这样就可以把 ∂D 的积分转到 (p_i) 和 (z_i) 上. 选取一个在 D 上的全纯函数 $\varphi(z)$，然后

$$\frac{1}{2\pi i}\oint_{\partial D}\varphi(z)\mathrm{d}\ln f(z)=\frac{1}{2\pi i}\left[\sum_i\oint_{(p_i)}\varphi(z)\mathrm{d}\ln f(z)+\sum_i\oint_{(z_i)}\varphi(z)\mathrm{d}\ln f(z)\right]$$

其中对于 m_i 重零点 z_i，存在一个解析函数 $g_i(z)$，使得 $\psi(z_i)=(z-z_i)^{m_i}g_i(z_i)$，其中 $g_i(z_i)\neq 0$. 对 $\ln\psi(z)$ 求导，得

$$\frac{\mathrm{d}}{\mathrm{d}z}\ln f(z)=\frac{f'(z)}{f(z)}=\frac{n_i}{z-z_i}+\frac{g_i'(z)}{g_i(z)}$$

由于 $g_i(z)$ 解析，所以只要 (z_i) 的半径足够小，$\frac{g_i'(z)}{g_i(z)}$ 在小圆周内也是解析的，根据柯西-古萨定理，沿着小圆周上对其积分得到 0，然后由留数定理[9]

$$\frac{1}{2\pi i}\oint_{(z_i)}\varphi(z)\frac{f'(z)}{f(z)}\mathrm{d}z=\frac{1}{2\pi i}\oint_{(z_i)}\varphi(z)\left(\frac{n_i}{z-z_i}+\frac{g_i'(z)}{g_i(z)}\right)\mathrm{d}z=\frac{n_i}{2\pi i}\oint_{(z_i)}\frac{\varphi(z)}{z-z_i}\mathrm{d}z=n_i\varphi(z_i)$$

对极点 p_i 用同样的方法讨论，可得

$$\frac{1}{2\pi i}\oint_{(p_i)}\varphi(z)\frac{f'(z)}{f(z)}\mathrm{d}z=-p_i\varphi(p_i)$$

代回，就得到了

定理　对数留数定理 (logarithmic residue theorem)

若 $f(z)$ 在简单闭曲线 C 内除了 m 个极点 $p_1,p_2,\cdots,p_m$ 以外处处解析，并有至多可数无穷多个零点 $z_1,z_2,z_3,\cdots$，函数 $\varphi(z)$ 在 C 内处处解析，则

$$\frac{1}{2\pi i}\oint_C\varphi(z)\frac{f'(z)}{f(z)}\mathrm{d}z=\sum_i n_i\varphi(z_i)-\sum_j n_j\varphi(p_j)$$

如果在区域 C 内 $f(z)$ 处处解析，且只有 1 个一级零点 z_0，则可以令 $f(z)=\psi(z)-\psi(z_0)$，易知 $\psi(z)$ 在 C 内解析且无零点. 此时

$$\frac{\mathrm{d}}{\mathrm{d}z}\ln f(z)=\frac{f'(z)}{f(z)}=\frac{\psi'(z)}{\psi(z)-\psi(z_0)}$$

代入对数留数定理，得到

$$\varphi(z_0)=\frac{1}{2\pi i}\oint_C\varphi(z)\frac{f'(z)}{f(z)}\mathrm{d}z=\frac{1}{2\pi i}\oint_C\frac{\varphi(z)\psi'(z)}{\psi(z)-\psi(z_0)}\mathrm{d}z$$

此即为

定理　扩展的柯西积分公式

设 $f(z)$ 在区域 C 内解析且没有零点，$\varphi(z)$ 在 C 内解析，则

$$f(z)=\frac{1}{2\pi i}\oint_C\frac{f(\xi)\varphi'(\xi)}{\varphi(\xi)-\varphi(z)}\mathrm{d}\xi$$

令 $\psi(z)=z$ 得到的就是常见的柯西积分公式.

泰勒公式和柯西积分公式告诉我们解析函数总可以展成次数为正的幂级数：

$$f(z)=\sum_{n=0}^{\infty}a_nz^n=a_0+a_1z+a_2z^2+a_3z^3\cdots$$

之前推广的对数留数定理和柯西积分公式都有一个共同的特征，那就是把一个具体的函数，比如 $\varphi(z)\equiv 1$ 或 $\varphi(z)=z$，推广到了任意的 $\varphi(z)$，接下来推广的目标也是这样，即把泰勒级数推广成这个形式.

$$f(z)=\sum_{n=0}^{\infty}a_n\varphi^n(z)=a_0+a_1\varphi(z)+a_2\varphi^2(z)+a_3\varphi^3(z)+\cdots$$

[9]留数定理：若简单闭曲线 C 包含 $f(z)$ 的极点 $p_1,p_2,\cdots,p_n$，且 $f(z)$ 在 C 内其他地方处处解析，则有

$$\oint_C f(z)\mathrm{d}z=2\pi i\sum_{k=1}^{n}\mathrm{Res}[f(z),z_k]$$

回忆用柯西积分公式导出解析函数的幂级数的过程，现在我们用推广的柯西积分公式来导出推广的幂级数. 首先改写推广的柯西积分公式.

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\oint_C \frac{f(\xi)\varphi'(\xi)}{\varphi(\xi)-\varphi(z)}\mathrm{d}\xi = \frac{1}{2\pi i}\oint_C \frac{f(\xi)\varphi'(\xi)}{\varphi(\xi)}\cdot\frac{1}{1-\frac{\varphi(z)}{\varphi(\xi)}}\mathrm{d}\xi \\ &= \frac{1}{2\pi i}\oint_C \frac{f(\xi)\varphi'(\xi)}{\varphi(\xi)}\left(\frac{\varphi(z)}{\varphi(\xi)}\right)^n\mathrm{d}\xi = \left[\frac{1}{2\pi i}\oint_C \frac{f(\xi)\varphi'(\xi)}{\varphi^{n+1}(\xi)}\mathrm{d}\xi\right]\varphi^n(z) \end{aligned}$$

因此系数就出来了，然后根据分部积分法和留数定理

$$a_n = \frac{1}{2\pi i}\oint_C \frac{f(\xi)\varphi'(\xi)}{\varphi^{n+1}(\xi)}\mathrm{d}\xi = \frac{1}{2n\pi i}\oint_C \frac{f'(\xi)}{\varphi^n(\xi)}\mathrm{d}\xi = \frac{1}{n!}\lim_{z\to z_0}\frac{\mathrm{d}^{n-1}}{\mathrm{d}z^{n-1}}\left[f'(z)\left(\frac{z-z_0}{\varphi(z)}\right)^n\right]$$

以上便是拉格朗日-布尔曼公式(Lagrange-Bürmann formula). 当 $\varphi(z)=z$ 时，公式退化为泰勒级数.

定理　拉格朗日-布尔曼公式

设 $f(z)$ 和 $\varphi(z)$ 在 $B(0,\delta)$ 内解析，且 $f'(0)\neq 0, \varphi(\xi_0)=0$，则

$$f(z)=\sum_{n=1}^{\infty}\left[\lim_{\xi\to\xi_0}\frac{\mathrm{d}^{n-1}}{\mathrm{d}\xi^{n-1}}\left(f'(\xi)\left(\frac{\xi-\xi_0}{\varphi(\xi)}\right)^n\right)\right]\frac{\varphi^n(z)}{n!}$$

特别地，当 $\xi_0=0$ 时，有

$$f(z)=\sum_{n=1}^{\infty}\left[\lim_{\xi\to 0}\frac{\mathrm{d}^{n-1}}{\mathrm{d}\xi^{n-1}}\frac{\xi^n f'(\xi)}{\varphi^n(\xi)}\right]\frac{\varphi^n(z)}{n!}$$

令 $f(z)=z$，带入拉格朗日-布尔曼公式的特殊情况，得到

$$z=\sum_{n=1}^{\infty}\left[\lim_{\xi\to 0}\frac{\mathrm{d}^{n-1}}{\mathrm{d}\xi^{n-1}}\left(\frac{\xi}{\varphi(\xi)}\right)^n\right]\frac{\varphi^n(z)}{n!}$$

如果恰好有 $\varphi^{-1}(w)=z$，则 $\varphi(z)=w$，此时

$$\varphi^{-1}(w)=\sum_{n=1}^{\infty}\left[\lim_{\xi\to 0}\frac{\mathrm{d}^{n-1}}{\mathrm{d}\xi^{n-1}}\left(\frac{\xi}{\varphi(\xi)}\right)^n\right]\frac{w^n}{n!}$$

这就是反函数的幂级数展开公式.

推论　反函数的幂级数展开公式

若 $f(z)$ 在 $B(0,\delta)$ 内解析且有反函数且 $f(0)=0$，则 $f(z)$ 和 $f^{-1}(z)$ 可分别做幂级数展开:

$$f(z)=\sum_{n=1}^{\infty}a_n z^n = a_1 z + a_2 z^2 + a_3 z^3\cdots$$
$$f^{-1}(z)=\sum_{n=1}^{\infty}b_n z^n = b_1 z + b_2 z^2 + b_3 z^3\cdots$$

其中 $b_n=\dfrac{1}{n!}\lim\limits_{\xi\to 0}\dfrac{\mathrm{d}^{n-1}}{\mathrm{d}\xi^{n-1}}\left(\dfrac{\xi}{f(\xi)}\right)^n$

2.5.3 拉格朗日级数

设目标函数 $f(z)$ 和一个辅助函数 $g(z)$ 在 C 内解析，然后取 C 内一点 s，并令 $\psi(z)=z-s-tg(z)$，其中 t 是参数，由辐角原理可得

$$N-P=\frac{1}{2\pi i}\oint_C\frac{1-tg'(z)}{z-s-tg(z)}\mathrm{d}z=\frac{1}{2\pi i}\oint_C\frac{1-tg'(z)}{z-s}\frac{\mathrm{d}z}{1-\frac{tg(z)}{z-s}}$$

如果再令 $|tg(z)|<|z-s|$，那么又可以作泰勒展开

$$N-P=\frac{1}{2\pi i}\oint_C\frac{1-tg'(z)}{z-s}\sum_{n=0}^{\infty}\left(\frac{tg(z)}{z-s}\right)^n\mathrm{d}z=\frac{t^n}{2\pi i}\sum_{n=0}^{\infty}\oint_C\frac{[g(z)]^n}{(z-s)^{n+1}}\mathrm{d}z-\frac{t^{n+1}}{2\pi i}\sum_{n=0}^{\infty}\oint_C\frac{g'(z)[g(z)]^n}{(z-s)^{n+1}}$$

由复变函数的高阶导数公式[10]可得

$$\frac{t^n}{2\pi i}\oint_C\frac{[g(z)]^n}{(z-s)^{n+1}}\mathrm{d}z=\frac{t^n}{n!}\frac{\mathrm{d}^n}{\mathrm{d}z^n}[g(s)]^n,\qquad \frac{t^{n+1}}{2\pi i}\oint_C\frac{g'(z)[g(z)]^n}{(z-s)^{n+1}}\mathrm{d}z=\frac{t^{n+1}}{(n+1)!}\frac{\mathrm{d}^{n+1}}{\mathrm{d}z^{n+1}}[g(s)]^{n+1}$$

代回，得到

$$N-P=\sum_{n=0}^{\infty}\frac{t^n}{n!}\frac{\mathrm{d}^n}{\mathrm{d}z^n}[g(s)]^n-\sum_{n=0}^{\infty}\frac{t^{n+1}}{(n+1)!}\frac{\mathrm{d}^{n+1}}{\mathrm{d}z^{n+1}}[g(s)]^{n+1}$$

比较两端的 t^n 系数，可以得到

$$N-P=\frac{t^0}{0!}\frac{\mathrm{d}^0}{\mathrm{d}z^0}[g(s)]^0=1$$

由于 $\psi(x)$ 在 C 内解析，所以不存在极点，即 $P=0$，因此 $N=1$，这就说明了方程 $\psi(x)=0$ 在 C 内只有一个根，把它记作 z，由对数留数定理得

$$f(z)=\frac{1}{2\pi i}\oint_C f(z)\frac{\psi'(z)}{\psi(z)}\mathrm{d}z=\frac{1}{2\pi i}\oint_C f(z)\frac{1-tg'(z)}{z-s-tg(z)}\mathrm{d}z$$

仿照上面推导的过程，就得到了

定理　拉格朗日展开公式

若 $f(z)$ 和辅助函数 $g(z)$ 在 C 上及 C 内解析，C 内一点 s、C 上一点 z 和参数 t 满足 $|tg(z)|<|z-s|$，则

(1) 方程 $z=s+tg(z)$ 在 C 内有且仅有一根；

(2) 该根可以展开为

$$f(z)=f(s)+\sum_{n=1}^{\infty}\frac{t^n}{n!}\frac{d^{n-1}}{dz^{n-1}}\{f'(s)[g(s)]^n\}$$

2.6 傅里叶级数

移至第四章

[10]复变函数的高阶导数公式：若函数 $f(z)$ 在简单闭曲线及其内部解析，z_0 是 C 内一点，则

$$\oint_C\frac{f(z)}{(z-z_0)^{n+1}}=\frac{2\pi i}{n!}\frac{\mathrm{d}}{\mathrm{d}z}f(z_0)$$

2.7 极点展开

来捋一捋奇点和极点的关系. 首先极点肯定是奇点，因为它没有定义，更不用说解析了. 若 z_0 为 $f(z)$ 的极点，则在 z_0 的某个邻域内可以找到一个解析函数 $g(z)$，使得 $f(z)=g(z)/(z-z_0)^n$，而 n 表示极点的等级. 可以这样理解，$f(z)$ 在 z_0 不解析，是因为出现了 $1/(z-z_0)^n$ 这个因子，在 $z\to z_0$ 的过程中，$f(z)$ 趋于无穷就是这个因子导致的. 由于 $g(z)$ 是解析的，在 z_0 的某个邻域内是有界量，所以 $f(z)$ 趋近于无穷的"速度"完全由极点的等级决定.

在某区域内解析的函数称为全纯函数(holomorphic function)，存在极点但除去极点以外处处解析的函数称为半纯函数(meromorphic function)，例如 $\cot(z)$ 就是一个半纯函数，它有无穷多个极点，而且都是一级极点，像这样的函数可以通过它的极点来刻画.

全纯函数 $g(z)$ 可以展开成泰勒级数 $g(z)=a_0+a_1z+a_2z^2+a_3z^3+\cdots$，此时具有 n 个一阶极点 $p_1,p_2,\cdots,p_n$ 半纯函数 $f(z)$ 就可以改写为

$$f(z)=\frac{a_0+a_1z+a_2z^2+a_3z^3+\cdots}{(z-p_0)(z-p_1)\cdots(z-p_n)}$$

如果 $f(z)$ 有无穷多个极点，则也许可以大胆地写成

$$f(z)=\frac{a_0+a_1z+a_2z^2+a_3z^3+\cdots}{(z-p_1)(z-p_2)(z-p_3)\cdots}$$

如果我们此时大胆地认为分子分母次数相同（都是无穷多次），就可以大胆地使用有理分式展开，即

$$f(z)=k_0+\frac{k_1}{z-p_1}+\frac{k_2}{z-p_2}+\frac{k_3}{z-p_3}+\cdots$$

欧拉发现 $\cot(z)$ 存在这种形式的展开，之后米塔格-累夫勒证明了一个定理保证这种展开是可行的.

定理 米塔格-累夫勒定理 (Mittag-Leffler's theorem, 1876)

设 $\{p_n\}_{n=1}^{\infty}$ 为没有聚点的复数数列，$\{P_n(z)\}_{n=1}^{\infty}$ 为一系列常数项为零的多项式，则存在一个半纯函数 $f(z)$，$p_1,p_2,p_3,\cdots$ 为其极点，且在以 p_n 为圆心的某个去心圆域内，$f(z)$ 洛朗展开式的主部为 $P_n\left(\dfrac{1}{z-p_n}\right)$.

如果极点全为一级极点，则其洛朗展开式即为有理分式的形式. 但该定理只给有理分式形式的展开提供了可能性，而并没有给出确切的展开式，这就是接下来要做的事.

给出确切展开式需要做较多的铺垫和限制. 把极点 p_i 重新编号，使得 $0<|p_1|\le|p_2|\le|p_2|\le|p_3|\le\cdots$，然后作出一系列闭曲线 $\{C_m\}_{m=1}^{\infty}$，使得闭曲线 C_m 围住 $p_1,p_2,\cdots,p_m$，接下来记 (z_0) 表示围住点 z_0 足够小的闭曲线.

在 C_m 取一点 z，根据复合闭路定理和柯西积分公式，有

$$\begin{aligned}\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi&=\frac{1}{2\pi i}\oint_{(z)}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi+\sum_{n=1}^{m}\frac{1}{2\pi i}\int_{(p_n)}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi\\&=f(z)+\sum_{n=1}^{m}\frac{1}{2\pi i}\int_{(p_n)}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi\end{aligned}\tag{1}$$

接下来分别整理这个式子的左边和右边，先整理右边. 假设 $f(\xi)$ 在点 p_n 的洛朗级数为 $P(\xi-p_n)+T(\xi-p_n)$，其中 $P(\xi-p_n)$ 表示主部，则蓝色部分即为

$$\sum_{n=1}^{m}\frac{1}{2\pi i}\oint_{(p_n)}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi=\sum_{n=1}^{m}\left[\frac{1}{2\pi i}\oint_{(p_n)}\frac{P(\xi-p_n)}{\xi-z}\mathrm{d}\xi+\frac{1}{2\pi i}\oint_{(p_n)}\frac{T(\xi-p_n)}{\xi-z}\mathrm{d}\xi\right]\tag{2}$$

因为 $\dfrac{T(\xi-p_n)}{\xi-z}$ 在 p_n 的邻域内解析，根据柯西积分公式，可以知道棕色部分为零. 现在作封闭曲线 C_R，是以 ξ 为圆心、半径为 R 的圆，R 足够大以至于包含 z 和 p_n 两点，根据复合闭路定理

$$\frac{1}{2\pi i}\oint_{(p_n)}\frac{P(\xi-p_n)}{\xi-z}\mathrm{d}\xi=\frac{1}{2\pi i}\oint_{C_R}\frac{P(\xi-p_n)}{\xi-z}\mathrm{d}\xi-\frac{1}{2\pi i}\oint_{(z)}\frac{P(\xi-p_n)}{\xi-z}\mathrm{d}\xi \tag{3}$$

注意到当 $R\to\infty$ 时，$\dfrac{P(\xi-p_n)}{\xi-z}=o(\xi^{-2})$，此时该积分趋于 0，这样就得到了

$$\frac{1}{2\pi i}\oint_{(p_n)}\frac{P(\xi-p_n)}{\xi-z}\mathrm{d}\xi=-\frac{1}{2\pi i}\oint_{(z)}\frac{P(\xi-p_n)}{\xi-z}\mathrm{d}\xi=-P(z-p_n) \tag{4}$$

代回 (1) 式，得到

$$\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi=f(z)-\sum_{n=1}^{m}P(z-p_n) \tag{5}$$

接下来整理 (1) 式左边，这是一个技巧性很强的地方，由等比数列求和公式可知

$$\sum_{n=0}^{p}\left(\frac{z}{\xi}\right)^n=\frac{1-(z/\xi)^p}{1-z/\xi}\Longrightarrow\frac{1}{1-z/\xi}=\frac{1-(z/\xi)^p+(z/\xi)^p}{1-z/\xi}=\sum_{n=0}^{p}\left(\frac{z}{\xi}\right)+\frac{(z/\xi)^p}{1-z/\xi}$$

因此 (1) 式左边可以化为

$$\begin{aligned}\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi&=\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi}\frac{1}{1-z/\xi}\mathrm{d}\xi\\&=\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi}\left[\sum_{n=0}^{p}\left(\frac{z}{\xi}\right)^n+\frac{(z/\xi)^p}{1-z/\xi}\right]\mathrm{d}\xi\\&=\sum_{n=0}^{p}\frac{z^n}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi^{n+1}}\mathrm{d}\xi+\frac{z^{p+1}}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi^{p+1}(\xi-z)}\mathrm{d}\xi\end{aligned} \tag{6}$$

对蓝色部分像 (1)(2) 式那样操作，先用复合闭路定理，再用高阶导数公式和洛朗级数展开，其中展开式 $T(\xi-p_k)$ 的部分为零.

$$\begin{aligned}\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi^{n+1}}\mathrm{d}\xi&=\frac{1}{2\pi i}\oint_{(z)}\frac{f(\xi)}{\xi^{n+1}}\mathrm{d}\xi+\sum_{k=1}^{m}\frac{1}{2\pi i}\int_{(p_k)}\frac{f(\xi)}{\xi^{n+1}}\mathrm{d}\xi\\&=\frac{f^{(n)}(0)}{n!}+\sum_{k=1}^{m}\frac{1}{2\pi i}\int_{(p_k)}\frac{P(\xi-p_k)}{\xi^{n+1}}\mathrm{d}\xi\end{aligned} \tag{7}$$

像之前一样，作圆 $C_R:|z|=R$，R 足够大使得极点 p_k 能含在内，这时候根据复合闭路定理和高阶导数公式

$$\begin{aligned}\frac{1}{2\pi i}\oint_{(p_k)}\frac{P(\xi-p_k)}{\xi^{n+1}}\mathrm{d}\xi&=\frac{1}{2\pi i}\oint_{C_R}\frac{P(\xi-p_k)}{\xi^{n+1}}\mathrm{d}\xi-\frac{1}{2\pi i}\oint_{(0)}\frac{P(\xi-p_k)}{\xi^{n+1}}\mathrm{d}\xi\\&=\frac{1}{2\pi i}\oint_{C_R}\frac{P(\xi-p_k)}{\xi^{n+1}}\mathrm{d}\xi-\frac{1}{n!}\frac{\mathrm{d}^n}{\mathrm{d}z^n}P(\xi-p_k)\bigg|_{\xi=0}\end{aligned} \tag{8}$$

像之前那样讨论，当 $R\to\infty$ 时，该积分趋于零，这样就得到了

$$\frac{1}{2\pi i}\oint_{(p_k)}\frac{P(\xi-p_n)}{\xi^{n+1}}\mathrm{d}\xi=-\frac{1}{k!}\frac{\mathrm{d}^k}{\mathrm{d}z^k}P(\xi-p_k)\bigg|_{\xi=0} \tag{9}$$

代回 (7) 式，得到

$$\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi^{n+1}}\mathrm{d}\xi=\frac{f^{(n)}(0)}{n!}-\sum_{k=1}^{m}\frac{1}{n!}\frac{\mathrm{d}^n}{\mathrm{d}z^n}P(\xi-p_k)\bigg|_{\xi=0} \tag{10}$$

代回 (6) 式，得到

$$\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi=\sum_{n=0}^{p}\frac{z^n}{n!}\left[f^{(n)}(0)-\sum_{k=1}^{m}\frac{\mathrm{d}^n}{\mathrm{d}z^n}P(\xi-p_k)\bigg|_{\xi=0}\right]+\frac{z^{p+1}}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi^{p+1}(\xi-z)}\mathrm{d}\xi \tag{11}$$

若记 $\varphi_{kp}(z)=\sum_{n=0}^{p}\frac{z^n}{n!}\frac{\mathrm{d}^n}{\mathrm{d}z^n}P(\xi-p_k)\Big|_{\xi=0}$ 则 (11) 式变为

$$\frac{1}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi-z}\mathrm{d}\xi=\sum_{n=0}^{p}\frac{z^n}{n!}f^{(n)}(0)-\sum_{k=1}^{m}\varphi_{kp}(z)+\frac{z^{p+1}}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi^{p+1}(\xi-z)}\mathrm{d}\xi \tag{12}$$

比较 (12) 式和 (5) 式，得到

$$f(z)=\sum_{n=0}^{p}\frac{z^n}{n!}f^{(n)}(0)+\sum_{n=1}^{m}[P(z-p_n)-\varphi_{np}(z)]+\frac{z^{p+1}}{2\pi i}\oint_{C_m}\frac{f(\xi)}{\xi^{p+1}(\xi-z)}\mathrm{d}\xi \tag{13}$$

现在来讨论蓝色这一节，稍微放缩一下

$$\begin{aligned}\oint_{C_m}\frac{f(\xi)}{\xi^{p+1}(\xi-z)}\mathrm{d}\xi&\le\oint_{C_m}\frac{|f(\xi)|}{|\xi^{p+1}||\xi-z|}|\mathrm{d}\xi|\quad(\text{积分放缩不等式})\\&\le\oint_{C_m}\frac{|\xi^{-p}f(\xi)|}{R_m|R_m-|z||}|\mathrm{d}\xi|\quad(\text{因为}R_m<|\xi|)\\&\le\frac{Ml_m}{R_m|R_m-|z||}\quad(\text{因为}|z^{-p}f(z)|<M)\end{aligned}$$

如果进一步假设闭曲线的周长 l_m 和闭曲线上的点离原点的最近距离 R_m 的比值 $|l_m/R_m|$ 有界，则当 $m\to\infty$ 时，$\oint_{C_m}\frac{f(\xi)}{\xi^{p+1}(\xi^{n+1})}\mathrm{d}\xi=0$，而无穷多个极点正是被展开的函数应当具有的性质，所以把该结果带入 (13) 式，最终就得到了

定理 米塔格-累夫勒定理 (Mittag-Leffler's theorem, 1884)

设有半纯函数 $f(z)$ 和一组围线序列 $\{C_m\}_{m=1}^{\infty}$ 满足

(1) $f(z)$ 有可数无穷多个极点，按绝对值大小排列为 $0<|a_1|\le|a_2|\le|a_3|\le\cdots$；

(2) 围线 C_m 围住 $p_1,p_2,\cdots,p_m$；

(3) 当 $m\to\infty$ 时围线周长 $l_m\to\infty$，围线上的点到原点的最小距离 $R_m\to\infty$，但 l_m/R_m 有界；

(4) C_m 上存在一点 z_m，使得存在最小的自然数 p 能让 $|z_m^{-p}f(z_m)|<M$，M 与 m 无关.

则

$$f(z)=\sum_{n=0}^{p}\frac{z^n}{n!}f^{(n)}(0)+\sum_{n=1}^{\infty}[P(z-p_n)-\varphi_{np}(z)]$$

其中 $P(z-z_n)$ 表示 $f(z)$ 在极点 p_n) 处洛朗展开的主部，$\varphi_{kp}(z)=\sum_{n=0}^{p}\frac{z^n}{n!}\frac{\mathrm{d}^n}{\mathrm{d}z^n}P(\xi-p_k)\Big|_{\xi=0}$.

这个展开式非常复杂，可以添加几个条件来简化. 若函数的所有极点均为 1 级极点，则 $P(z-p_n)=\frac{k_n}{z-p_n}$，其中 k_n 是常数[11]；若 $f(z)$ 在 C_m 上有界，则 $p=0$，此时展开式可以化简为

$$f(z)=f(0)+\sum_{n=0}^{\infty}k_n\left(\frac{1}{z-p_n}+\frac{1}{p_n}\right)$$

[11]极点：若 $f(z)$ 在 p_n 处的洛朗展开式有有限多个负幂项，而且负幂项的最高次为 m，则 p_n 为 f 的 m 级极点，因此在 1 级极点的洛朗展开式主部仅有一项 $c_{-1}(z-p_n)$.

其中 $k_n=\dfrac{1}{2\pi i}\displaystyle\oint_{(p_n)} f(z)\mathrm{d}z$.

米塔格-累夫勒定理适合用于展开含有 1 级极点的周期函数，以下给出几个例子. 作为对比，麦克劳林展开需要强调收敛域，极点展开则在整个定义域内都成立.

$$
\begin{aligned}
\tan(z) &= \sum_{n=-\infty}^{\infty}\frac{-1}{z-\left(n+\frac{1}{2}\right)\pi} &&= -\frac{1}{z-\frac{1}{2}\pi}-\frac{1}{z+\frac{3}{2}\pi}-\frac{1}{z-\frac{3}{2}\pi}-\cdots\\
\cot(z) &= \sum_{n=-\infty}^{\infty}\frac{1}{z-n\pi} &&= \frac{1}{z}+\frac{1}{z+\pi}+\frac{1}{z-\pi}+\frac{1}{z-2\pi}+\frac{1}{z-2\pi}+\cdots\\
\sec(z) &= \sum_{n=-\infty}^{\infty}\frac{(-1)^{n-1}}{z-\left(n+\frac{1}{2}\right)\pi} &&= \frac{1}{z+\frac{1}{2}\pi}-\frac{1}{z-\frac{1}{2}\pi}+\frac{1}{z+\frac{3}{2}\pi}-\frac{1}{z-\frac{3}{2}\pi}+\cdots\\
\csc(z) &= \sum_{n=-\infty}^{\infty}\frac{(-1)^{n}}{z-n\pi} &&= \frac{1}{z}-\frac{1}{z+\pi}-\frac{1}{z-\pi}+\frac{1}{z-2\pi}+\frac{1}{z-2\pi}+\cdots
\end{aligned}
$$

2.8 零点分解

零点分解和之前的极点展开充分体现了对偶性，这里说的是“分解”而不是“展开”其区别就在于“分解”是把函数写成无穷乘积的形式，有别于“展开”的无穷级数. 代数基本定理告诉我们，任意一个多项式都可以在复数范围内分解成多个一次因子. 要分解

$$P(z)=a_nz^n+a_{n-1}z^{n-1}+\cdots+a_2z^2+a_1z+a_0$$

只要令 $P(z)=0$，解出它的 n 个根 $z_1,z_2,\cdots,z_n$，就可以作分解

$$P(z)=(z-z_1)(z-z_2)\cdots(z-z_{n-1})(z-z_n)$$

这就把多项式函数和它的零点关联起来了.

顺着这个想法，魏尔斯特拉斯思考到：全纯函数能被高次多项式拟合，那么这种能“看出”函数零点的分解是否能扩展到任意全纯函数上呢？更进一步地，任意给定一个递增且无界的复数数列 $\{z_n\}_{n=1}^{\infty}$，是否能找到一个函数正好以 $z_1,z_2,z_3,\cdots$ 为零点呢？

2.8.1 基本因式

无论以上想法是否合理，被零点所确定的函数一定不可能是 e^z 或随意一个 $e^{f(z)}$ 形式的函数，因为这样的函数没有零点，因此在无穷乘积分解某个函数 $f(z)$ 的时候，还需要考虑到这一类函数对分解的影响，办法就是把它们作为因子提出来，得到 $f(z)=e^{g(z)}h(z)$ 的形式. 但实际上 $h(z)$ 也没有办法直接分解成一次因式，但可以分解为类似于一次因式的基本因式.

定义 基本因式

函数

$$E_p(z):=\begin{cases}(1-z) & p=0,\\ (1-z)e^{z+\frac{z^2}{2}+\frac{z^3}{3}+\cdots+\frac{z^p}{p}} & p>0\end{cases}$$

称为基本因式(elementary factor).

魏尔斯特拉斯的第二个想法由以下引理证实.

引理　无界的零点序列可以确定一个整函数

设 $\{z_n\}_{n=1}^{\infty}$ 为复数数列，满足任意 $z_n \neq 0$ 且 $\lim\limits_{n\to\infty}|z_n| = +\infty$；$\{p_n\}_{n=1}^{\infty}$ 为自然数数列，若对于任意 $r > 0$，级数 $\sum\limits_{n=1}^{\infty}\left(\dfrac{r}{|z_n|}\right)^{1+p_n}$ 收敛，那么函数 $h(z) = \prod\limits_{n=1}^{\infty} E_{p_n}\left(\dfrac{z}{z_n}\right)$ 是一个整函数，且零点恰好为所有 z_n.

证明：分两步，第一步证明当 $|z| < 1$ 时 $|1 - E_p(z)| \leq |z|^{p+1}$. 当 $p = 0$ 时显然成立；当 $p > 0$ 时，求导然后用等比数列求和公式

$$E_p'(z) = [(1-z)(1+z+z^2+\cdots+z^{p-1}) - 1]e^{z+\frac{z^2}{2}+\frac{z^3}{3}+\cdots+\frac{z^p}{p}} = -z^p e^{z+\frac{z^2}{2}+\frac{z^3}{3}+\cdots+\frac{z^p}{p}}$$

所以 $-E_p'(z)$ 在 $z = 0$ 处有 p 级零点，且洛朗展开的系数都是非负实系数. 由于

$$1 - E_p(z) = \int_0^z E_p(\xi)\mathrm{d}\xi$$

所以 $1-E_p(z)$ 在 $z = 0$ 处有 $p+1$ 级零点，且洛朗展开的系数都是非负实系数. 这时只要设 $\varphi(z) = \dfrac{1-E_p(z)}{z^{p+1}}$，则 $\varphi(z)$ 的洛朗展开系数也是非负实系数，所以当 $|z| \leq 1$ 时有 $|\varphi(z)| \leq \varphi(1) = 1$，因此 $|1-E_p(z)| \leq |z|^{p+1}$.

第二步即证明该命题. 因为 $\lim\limits_{n\to\infty}|z_n| = +\infty$，所以除开有限的元素外，$\{|z_n|\}$ 中其他无限个元素都大于 r. 将 r 固定，若 $|z| \leq r$ 且 $|z_n| > r$，则

$$\left|1 - E_{p_n}\left(\frac{z}{z_n}\right)\right| \leq \left|\frac{z}{z_n}\right|^{1+p_n} \leq \left(\frac{r}{|z_n|}\right)^{1+p_n}$$

由比较判别法可知级数 $\sum\limits_{n=1}^{\infty}|1 - E_p(z)|$ 在复平面内一致收敛，因此无穷乘积所确定的函数为整函数.

数列 $\{p_n\}$ 的选取有很大的任意性，就令 $p_n = n$，得到的数列也符合要求.

2.8.2　魏尔斯特拉斯分解定理

已知 $h(z)$ 是以上由零点确定的整函数，再乘以 $e^{g(z)}$ 这类没有零点的函数，就可以得到任意的整函数的分解定理.

定理　魏尔斯特拉斯分解定理 (Weierstrass factorization theorem)

设 $f(z)$ 为整函数且 $f(0) \neq 0$，复数数列 $\{z_n\}_{n=1}^{\infty}$ 按照重数枚举除 $z = 0$ 以外的零点（m 级零点按照 m 个零点计算），则存在一个自然数数列 p_n 和整函数 $g(z)$，使得 $f(z)$ 可以被分解为

$$f(z) = e^{g(z)}\prod_{n=1}^{\infty} E_{p_n}\left(\frac{z}{z_n}\right)$$

如果 $f(z)$ 在 $z = 0$ 处有 m 阶零点，则

$$f(z) = z^m e^{g(z)}\prod_{n=1}^{\infty} E_{p_n}\left(\frac{z}{z_n}\right)$$

我们希望不确定的对象尽可能减少，现在已经得出了 $\{p_n\}$ 的范例，就差 $g(z)$ 了，接下来用米塔格-累夫勒定理来导出更精确的表达式.

令 $g(z) = f'(z)/f(z)$ 满足简化版本米塔格-累夫勒定理的使用条件，那么

$$\frac{f'(z)}{f(z)} = \frac{f'(0)}{f(0)} + \sum_{n=0}^{\infty} k_n\left(\frac{1}{z-p_n} + \frac{1}{p_n}\right)$$

根据对数留数定理

$$k_n = \frac{1}{2\pi i}\oint_{(p_n)}\frac{f'(\xi)}{f(\xi)}\mathrm{d}\xi = 1$$

两边求积分得

$$\begin{aligned}\int_0^{2\pi}\frac{f'(z)}{f(z)} &= \int_0^{2\pi}\frac{f'(0)}{f(0)}\mathrm{d}z + \int_0^{2\pi}\sum_{n=0}^{\infty}k_n\left(\frac{1}{z-p_n}+\frac{1}{p_n}\right)\mathrm{d}z\\ \Longleftrightarrow \ln\frac{f(z)}{f(0)} &= \frac{f'(0)}{f(0)}z + \sum_{n=0}^{\infty}\left(\frac{1}{z-p_n}+\frac{1}{p_n}\right)\mathrm{d}z\end{aligned}$$

两边求指数得

$$\frac{f(z)}{f(0)} = e^{\frac{f'(0)}{f(0)}z} + \prod_{n=1}^{\infty}\left(1-\frac{z}{z_n}\right)e^{\frac{z}{z_n}}$$

稍微整理一下就是

推论　零点分解

设 $f(z)$ 为整函数，具有不为零的零点 $\{z_n\}_{n=1}^{\infty}$，且存在围线序列 $\{C_m\}_{n=1}^{\infty}$，函数 $\left|\frac{f'(z)}{f(z)}\Big|_{z\in C_m}\right| < M$，且 M 与 m 无关，则

$$\boxed{f(z) = f(0)e^{\frac{f'(0)}{f(0)}z}\prod_{n=1}^{\infty}\left(1-\frac{z}{z_n}\right)e^{\frac{z}{z_n}}}$$

2.8.3 哈达玛分解定理

哈达玛在魏尔斯特拉斯的基础上更进了一步，指出 $g(z)$ 是一个增长“慢于”某个上限的多项式，为严格定义“增长速度”，引进“阶数”的概念.

定义　阶数

对于函数 $f(z)$，若存在 $\rho, R > 0$，使得 $|z| > R$ 时，有 $|f(z)| < e^{|z|^{\rho}}$，则 ρ 的最小可能值称为 $f(z)$ 的阶数.

举几个例子，e^z 的阶数为 1，$\sin(z)$ 的阶数为 1，$\cos\sqrt{z}$ 的阶数为 1/2，e^{z^2} 的阶数为 2，e^{e^z} 的阶数为 $+\infty$.

定理　哈达玛分解定理 (Hadamard factorization theorem)

设 $f(z)$ 为阶数为 $\rho < +\infty$ 整函数且 $f(0) \neq 0$，复数数列 $\{z_n\}_{n=1}^{\infty}$ 按照重数枚举除 $z = 0$ 以外的零点（m 级零点按照 m 个零点计算），则 $f(z)$ 可以被分解为

$$f(z) = e^{g(z)}\prod_{n=1}^{\infty}E_p\left(\frac{z}{z_n}\right)$$

如果 $f(z)$ 在 $z = 0$ 处有 m 阶零点，则

$$f(z) = z^m e^{g(z)}\prod_{n=1}^{\infty}E_p\left(\frac{z}{z_n}\right)$$

其中 $g(z)$ 为阶数不大于 ρ 的多项式，$p = \lfloor\rho\rfloor$.

$$
\begin{aligned}
\sin(z) &= z\prod_{n\in\mathbb{Z}^*}\left(1-\frac{z}{\pi n}\right)e^{\frac{z}{\pi n}} &&= z\left(1-\frac{z}{\pi}\right)\left(1+\frac{z}{\pi}\right)\left(1-\frac{z}{2\pi}\right)\left(1+\frac{z}{2\pi}\right)\cdots \\
\cos(z) &= \sum_{n\mathrm{odd}\in\mathbb{Z}}\left(1-\frac{2z}{\pi n}\right)e^{\frac{2z}{\pi n}} &&= \left(1-\frac{2z}{\pi}\right)\left(1+\frac{2z}{\pi}\right)\left(1-\frac{2z}{3\pi}\right)\left(1+\frac{2z}{3\pi}\right)\cdots
\end{aligned}
$$

2.9 皮瑟级数

已知函数 $f(z^k)$ 的洛朗级数为

$$f(z)=\sum_{n=-\infty}^{\infty}c_n z^n$$

令 $w=z^k$，得到

$$f(w)=\sum_{n=-\infty}^{\infty}c_n w^{n/k}$$

像这种次数带有分数的幂级数，称为皮瑟级数(Puiseux series).

例如 $f(z)=\sqrt{z^2-1}$，在 $z=0$ 点它的洛朗级数可以通过配凑的过程来求得：

$$
\begin{aligned}
\sqrt{z^2-1} &= i\sqrt{1-z^2}=i\left(1+(iz)^2\right)^{1/2} \\
&= i\left(1-\frac{z^2}{2}-\frac{z^4}{8}-\frac{z^6}{16}-\cdots\right) \\
&= 1-\frac{iz^2}{2}-\frac{iz^4}{8}-\frac{iz^6}{16}-\cdots
\end{aligned}
$$

但是在 $z=1$ 点就没法配凑展开了，只能套定义式算 c_n，而皮瑟级数仍然可以配凑出来.

$$\sqrt{z^2-1}=\sqrt{(z+1)(z-1)}=\sqrt{2(z-1)}\sqrt{1+\frac{z-1}{2}}$$

令 $w=z-1$，则

$$
\begin{aligned}
\sqrt{z^2-1} &= \sqrt{2w}\sqrt{1+\frac{w}{2}} \\
&= \sqrt{2w}\left(1+\frac{w}{4}-\frac{w^2}{32}+\cdots\right) \\
&= \sqrt{2w}+\frac{\sqrt{2}}{4}w^{3/2}-\frac{\sqrt{2}}{32}w^{5/2}+\cdots
\end{aligned}
$$

第四部分　傅里叶分析

这一部分的符号有些混乱，请注意区分 $f(t), f[n], \hat{f}(\omega), \hat{f}(s), \hat{f}[z], \tilde{f}(t), \mathcal{F}[f(t)](\omega), \mathcal{F}[f[n]](\omega), \mathcal{F}[f[n]][k]$, $\mathcal{L}[f(t)](s), \mathcal{Z}[f(t)](z)$，接下来把这些符号梳理一遍.

- $f(t)$ 表示一般的（时域上的）连续函数，该函数在整个 $\mathbb{R}$ 上都有定义，自变量是 t，并赋予其时间的物理意义；
- $f[n]$ 表示一般的（时域上的）离散函数（即数列），该函数在整个 $\mathbb{Z}$ 上都有定义，自变量使用 n 表示，并赋予其采样点序号的物理意义；

- $\tilde{f}(\cdot)$ 中的波浪上标可以加在任意一个域中的任意一个函数上，强调该函数是周期函数；
- $\hat{f}(\cdot)$ 中的尖角上标 ^ 表示该函数是变换域中的函数，自变量字母的选取表示该函数处在不同变换域；
- $\mathcal{F}[f(t)](\omega)$，表示一个自变量是 ω 的连续频域函数，由连续时域函数 $f(t)$ 经过傅里叶变换得到，赋予 ω 角频率的物理意义，赋予 $\xi=\dfrac{\omega}{2\pi}$ 频率的物理意义（字母 f 为函数专用，所以频率改用 ξ，这是傅里叶分析的书籍上惯用的字母）.
- $\mathcal{F}[f[n]](\omega)$，表示一个自变量是 ω 的连续频域函数，由离散序列 $f[n]$ 经过离散时间傅里叶变换得到，赋予 ω 角频率的物理意义.
- $\mathcal{F}[f[n]][k]$，表示一个自变量是 k 的离散函数，由离散序列 $f[n]$ 经过离散傅里叶变换得到，赋予 k 离散角频率的物理意义.
- $\mathcal{L}[f(t)](s)$，表示一个自变量是 s 的连续函数，由连续函数 $f(t)$ 经过双边拉普拉斯变换得到，赋予 s 复频率的物理意义.
- $\mathcal{Z}[f[n]][z]$，表示一个自变量是 z 的离散函数，由离散序列 $f[n]$ 经过双边 Z 变换得到，赋予 z 离散复频率的物理意义.

注：以上这些变换本质上是函数到另一个函数的映射，凡是性质足够好的函数都可以应用这些变换，例如 Schwartz 空间及其对偶空间中的函数都可以进行傅里叶变换，变换后的函数仍属于该空间. 所以在此层面上并不存在所谓“时域”、“变换域”之类的说法，但为了贴合实际应用需要，我们往往通过其自变量使用的字母来区分这些域，同时也区分 $\mathcal{F}[\cdot]$ 是哪一种傅里叶变换，例如我们认为 $f(t)=e^{-2t}$ 是时域函数，$\hat{f}(\omega)=e^{-2\omega}$ 是连续频域内的函数，尽管这两个函数的对应法则是相同的，在 λ 演算层面都可以表示为 $\lambda x.e^{-2x}$.

1 正交展开

正交展开不是某种特殊的展开定理，而是一种重要的展开思想，首先回顾一下线性代数的知识.

> n 维线性向量空间中只要确定了一组完备的基底 $\mathcal{A}=\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n\}$，则空间中的所有向量都可以用基底的线性组合表示出来：
>
> $$\boldsymbol{\alpha}=k_1\boldsymbol{e}_1+k_2\boldsymbol{e}_2+\cdots+k_n\boldsymbol{e}_n$$
>
> 其中 k_i 是数，可以是实数、复数或其他无向的量.
> 如果空间中还定义了**内积**，则称该向量空间为**内积空间**，内积是一种运算，其结果为一个数量（实数或复数)，两个向量的内积记为 $\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle$，满足
>
> (1) $\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle=\overline{\langle\boldsymbol{\beta},\boldsymbol{\alpha}\rangle}$，上划线表示取共轭，如果内积运算定义在实数域上，则有 $\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle=\langle\boldsymbol{\beta},\boldsymbol{\alpha}\rangle$.
>
> (2) $\langle m\boldsymbol{\alpha}+n\boldsymbol{\beta},\boldsymbol{\gamma}\rangle=m\langle\boldsymbol{\alpha},\boldsymbol{\gamma}\rangle+n\langle\boldsymbol{\beta},\boldsymbol{\gamma}\rangle$，其中 m,n 是数量.
>
> (3) $\langle\boldsymbol{\alpha},\boldsymbol{\alpha}\rangle\geq 0$，当且仅当 $\boldsymbol{\alpha}=\mathbf{0}$ 时等号成立.
>
> 定义了内积之后，向量的**模长**便可定义为 $\|\boldsymbol{\alpha}\|=\sqrt{\langle\boldsymbol{\alpha},\boldsymbol{\alpha}\rangle}$，易知 $\|\boldsymbol{\alpha}\|\geq 0$；
> 线性组合的系数可表示为 $k_i=\dfrac{\langle\boldsymbol{\alpha},\boldsymbol{e}_i\rangle}{\langle\boldsymbol{e}_i,\boldsymbol{e}_i\rangle}$，成为向量 $\boldsymbol{\alpha}$ 在向量 $\boldsymbol{e}_i$ 上的**投影**.
> 向量的**正交**则表示 $\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle=0$ 这种情况，可简记为 $\boldsymbol{\alpha}\perp\boldsymbol{\beta}$.

> 假设 $\mathcal{A}=\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n\}$ 是自然基,即两两正交且 $\|\boldsymbol{e}_i\|=1$,则立即可以确定线性组合的系数:$k_i=\langle\boldsymbol{\alpha},\boldsymbol{e}_i\rangle$,而模也可以立即确定为 $\|\boldsymbol{\alpha}\|=\sqrt{k_1^2+k_2^2+\cdots+k_n^2}=\sqrt{\langle\boldsymbol{\alpha},\boldsymbol{e}_1\rangle^2+\langle\boldsymbol{\alpha},\boldsymbol{e}_2\rangle^2+\cdots+\langle\boldsymbol{\alpha},\boldsymbol{e}_n\rangle^2}$

正交展开的核心就是这些，但以上只给了抽象的框架，具体操作起来大有可为.

仔细思考线性组合的式子：$\boldsymbol{\alpha}=k_1\boldsymbol{e}_1+k_2\boldsymbol{e}_2+\cdots+k_n\boldsymbol{e}_n$，它想表达的是 $\boldsymbol{\alpha}$ 可以由向量 $e_1,e_2,\cdots,e_n$ 各自贡献一点组合而成，投影即系数 k_i 定量描述了向量 $\boldsymbol{e}_i$ 贡献的大小. 如果为 $k_i=0$ 表示向量 $\boldsymbol{e}_i$ 没有贡献，此时得到的 $\boldsymbol{\alpha}$ 和 $\boldsymbol{e}_i$ 无关，或称正交.

在函数空间中，函数就是向量，它的内积有一套独特的定义，但大多数教材并没有解释为什么要这么定义，以下是一种可能的理解.

线性代数告诉我们，实区间 $[a,b]$ 上的连续实变函数构成了一个线性空间，但函数空间不像向量空间那么直观，它的内积并没有那么容易定义出来。假设函数 f 在实区间 $[a,b]$ 上等距抽样，例如间隔 $\dfrac{1}{n}$ 取值，然后用一个向量 $\vec{\boldsymbol{u}}_f$ 来记录这些值：

$$\vec{\boldsymbol{u}}_f=\left[f(a),f(a+\frac{b-a}{n}),f(a+\frac{2(b-a)}{n}),f(a+\frac{3(b-a)}{n}),\cdots,f(a+\frac{(n-1)(b-a)}{n}),f(b)\right]$$

对于另一个在区间 $[a,b]$ 上连续的函数 $g(x)$，也采用相同的操作：

$$\vec{\boldsymbol{u}}_g=\left[g(a),g(a+\frac{b-a}{n}),g(a+\frac{2(b-a)}{n}),g(a+\frac{3(b-a)}{n}),\cdots,g(a+\frac{(n-1)(b-a)}{n}),g(b)\right]$$

此时对两个向量求内积：

$$\langle\vec{\boldsymbol{u}}_f,\vec{\boldsymbol{u}}_g\rangle=f(a)g(a)+f(a+\frac{b-a}{n})g(a+\frac{b-a}{n})+f(a+\frac{2(b-a)}{n})g(a+\frac{2(b-a)}{n})+\cdots+f(b)g(b)$$

现在令 $n\to\infty$，意味着抽样间隔 $\dfrac{1}{n}\to 0$，由于两个函数都是连续函数，不会发生跳跃，因此可以认为 $\vec{\boldsymbol{u}}_f$ 和 $\vec{\boldsymbol{u}}_g$ 这两个向量记录了函数的所有信息. 但是这样的内积往往会趋近于无穷大. 由于是等距抽样，每个样本的权重相等，所以在每个单项式上乘以权重 $1/n$，然后奇迹般地凑出了积分的定义式

$$\lim_{n\to\infty}\frac{1}{n}\langle\vec{\boldsymbol{u}}_f,\vec{\boldsymbol{u}}_g\rangle=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}f(a+\frac{k(b-a)}{n})g(a+\frac{k(b-a)}{n}):=\int_a^b f(x)g(x)\mathrm{d}x$$

以上是实变函数的情况，在这种情况下，才满足 $[f(x)]^2\equiv|f(x)|^2$，根据定积分的保号性，这即可满足线性空间中 $\|f\|\geq 0$ 的要求. 如果扩展到复变函数，由于两个复数的乘积还是复数，复变函数作定积分以后也是复数，不一定能与 0 比大小，所以为了符合模长的定义，需要将被积函数由 $[f(z)]^2$ 改为 $|f(z)|^2$，而后者等价于 $f(z)\overline{f(z)}$，这就得到了复变函数空间的内积.

定义 函数空间的内积

对于在复连通域 D 上定义的连续函数 $f(z), g(z)$，规定其内积(inner product) 为

$$\langle f, g\rangle := \int_D f(z)\overline{g(z)}\mathrm{d}z$$

若 $\langle f, g\rangle = 0$，则称这两个函数是正交的(orthonormal). 函数的模定义为

$$\|f\| := \sqrt{\int_D f(z)\overline{f(z)}\mathrm{d}z}$$

当复连通域定退化为区间 $[a, b]$ 时，复变函数退化为实变函数时，其内积便退化为

$$\langle f, g\rangle := \int_a^b f(x)g(x)\mathrm{d}x$$

模长退化为

$$\|f\| := \sqrt{\int_a^b [f(x)]^2\mathrm{d}x}$$

定义好函数空间的内积之后，便可以尝试在空间中寻找一组函数作为基底 $\mathcal{A} = \{\varphi_1, \varphi_2, \varphi_3, \cdots\}$，然后把空间中的任意都写成它们的线性组合的形式，

$$f(z) = c_1\varphi_1(z) + c_2\varphi_2(z) + c_3\varphi_3(z) + \cdots$$

如果基底 $\mathcal{A}$ 是正交基底，则可以进一步得到

$$c_n = \frac{\langle f, \varphi_n\rangle}{\langle \varphi_n, \varphi_n\rangle} = \frac{\langle f, \varphi_n\rangle}{\|\varphi_n\|^2} = \frac{\displaystyle\int_D f(z)\overline{\varphi_n(z)}\mathrm{d}z}{\displaystyle\int_D \varphi_n(z)\overline{\varphi_n(z)}\mathrm{d}z}$$

这便是正交展开的基本思路，由于这个思路最早应用在傅里叶级数上，所以又称这种展开方式为广义傅里叶展开.

以上有个细节需要注意，函数空间的基底写成 $\mathcal{A} = \{\varphi_1, \varphi_2, \varphi_3, \cdots\}$ 而不是 $\mathcal{A} = \{\varphi_1, \varphi_2, \varphi_3, \cdots, \varphi_n\}$，是因为基向量的个数可能是无限的，甚至可能是不可数无穷多的. 如果是实数多个的基向量 $\mathcal{A} = \{\varphi_s\}_{s\in I}$，那么每个基向量的分量 c_s 中的下标取值也不再可数，此时 $f(z)$ 就不能简单地展开成基向量线性组合的形式，而是像上面那样间隔取样，缩小间隔，最后凑出积分的形式，即

$$f(z) = \int_I c_s\varphi_s(z)\mathrm{d}s$$

剧透一下，这实际上就是傅里叶反变换公式的意义.

最后简单总结一下什么是正交展开. 如果向量 $\boldsymbol{\alpha}$ 属于线性空间 S，那么总是可以找到一组完备的正交基[12]，使得 $\boldsymbol{\alpha}$ 可以被这组基底线性表出，该线性表出式就是向量 $\boldsymbol{\alpha}$ 的正交展开.

2 平行地导出傅里叶级数和傅里叶变换

有了正交展开的思想，我们就可以摧枯拉朽般的导出傅里叶级数和连续傅里叶变换，流畅自然的程度令带积分余项的泰勒公式也自愧不如，而不用像很多教材那样再拿三角函数的积分倒来倒去，这也是连续傅里叶变换和傅里叶级数能写在一起的原因.

[12] “所有线性空间，无论维数有穷还是无穷，都能找到一组完备的正交基底”这个命题依赖于选择公理

- **导出 $[t_0, t_0+T]$ 上三角函数形式的傅里叶级数：基底 $\mathcal{A}=\{\sin(n\omega t),\cos(n\omega t)\}_{n=0}^{\infty}$（$\sin(0\omega t)$ 除外）**

需要注意，这里的 $\omega=\dfrac{2\pi}{T}$，有角频率的物理意义，这也限定了 $\sin(n\omega t)$ 和 $cos(n\omega t)$ 的周期都是 T。首先验证这些基底确实是正交的，别忘了三个积化和差公式，这里贴心地把它们写出来：

$$\begin{aligned}
\sin\alpha\cos\beta &= +\frac{1}{2}\left[\sin(\alpha+\beta)+\sin(\alpha-\beta)\right]\\
\sin\alpha\sin\beta &= -\frac{1}{2}\left[\cos(\alpha+\beta)-\cos(\alpha-\beta)\right]\\
\cos\alpha\cos\beta &= +\frac{1}{2}\left[\cos(\alpha+\beta)+\cos(\alpha-\beta)\right]
\end{aligned}$$

带入计算内积：

$$\begin{aligned}
&\forall m,n\in\mathbb{N},\quad m\neq n:\\
&\langle\sin(m\omega t),\cos(n\omega t)\rangle := \int_{t_0}^{t_0+T}\sin(m\omega t)\cos(n\omega t)\mathrm{d}t = -\left[\frac{\cos((m-n)\omega t)}{2(m-n)\omega}+\frac{\cos((m+n)\omega t)}{2(m+n)\omega}\right]_{t_0}^{t_0+T}=0\\
&\langle\sin(m\omega t),\sin(n\omega t)\rangle := \int_{t_0}^{t_0+T}\sin(m\omega t)\sin(n\omega t)\mathrm{d}t = \left[\frac{\sin((m-n)\omega t)}{2(m-n)\omega}-\frac{\sin((m+n)\omega t)}{2(m+n)\omega}\right]_{t_0}^{t_0+T}=0\\
&\langle\cos(m\omega t),\cos(n\omega t)\rangle := \int_{t_0}^{t_0+T}\cos(m\omega t)\cos(n\omega t)\mathrm{d}t = \left[\frac{\sin((m-n)\omega t)}{2(m-n)\omega}+\frac{\sin((m+n)\omega t)}{2(m+n)\omega}\right]_{t_0}^{t_0+T}=0
\end{aligned}$$

内积都是 0，所以正交. 接下来求出它们的模长：

$$\begin{aligned}
\|1\| &= \sqrt{\int_{t_0}^{t_0+T}\mathrm{d}t}=\sqrt{T}\\
\|\cos(n\omega t)\| &= \sqrt{\int_{t_0}^{t_0+T}\cos^2(n\omega t)\mathrm{d}t}=\sqrt{\frac{T}{2}},\quad n\in\mathbb{N}^*\\
\|\sin(n\omega t)\| &= \sqrt{\int_{t_0}^{t_0+T}\sin^2(n\omega t)\mathrm{d}t}=\sqrt{\frac{T}{2}},\quad n\in\mathbb{N}^*
\end{aligned}$$

这时候满足某些条件的函数就可以做正交展开：

$$f(t)=\frac{\langle f(t),1\rangle}{\|1\|^2}+\sum_{n=1}^{\infty}\left[\frac{\langle f(t),\sin(n\omega t)\rangle}{\|\sin(n\omega t)\|^2}\sin(n\omega t)+\frac{\langle f(t),\cos(n\omega t)\rangle}{\|\cos(n\omega t)\|^2}\cos(n\omega t)\right]$$

这就是周期为 T 的傅里叶级数，整理一下就好.

定义　三角函数形式的傅里叶级数

定义在区间 $[t_0, t_0+T]$ 上的函数 $f(t)$，将

$$S_\infty[f](t)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left[a_n\cos(n\omega t)+b_n\sin(n\omega t)\right]$$

称为 $f(t)$ 的三角函数形式的傅里叶级数(Fourier series)，其中

$$\begin{aligned}
a_n &= \frac{2}{T}\int_{t_0}^{t_0+T}f(t)\cos(n\omega t)\mathrm{d}t\\
b_n &= \frac{2}{T}\int_{t_0}^{t_0+T}f(t)\sin(n\omega t)\mathrm{d}t\\
\omega &= \frac{2\pi}{T}
\end{aligned}$$

嘿！上面已经说好了 $f(t)$ 等于这一大串级数，为什么这里又要搞出个 $S_\infty[f](t)$ 出来呢？先卖个关子，等讨论到收敛性的时候就会发现 $f(t)$ 需要满足一定的条件才能让 $S_\infty[f](t)$ 收敛到 $f(t)$ 上，而且收敛的方式也值得探究.

- **导出 $[t_0, t_0+T]$ 上极坐标形式的傅里叶级数：基底 $\mathcal{A}=\{A_n\sin(n\omega t+\varphi_n)\}_{n=0}^\infty$**

你以为我又会验证一遍正交性，再计算模长，然后还要求出每个 A_n 和 φ_n 的表达式？辅助角公式它不香吗？

$$A\sin(t)+B\cos(t)=\sqrt{A^2+B^2}\sin(t+\arctan\frac{B}{A})=\sqrt{A^2+B^2}\cos(t+\operatorname{arccot}\frac{B}{A})$$

把正弦项和余弦项合并，就得到了

定义　极坐标形式的傅里叶级数

定义在区间 $[t_0, t_0+T]$ 上的函数 $f(t)$，将

$$S_\infty[f](t)=\frac{a_0}{2}+\sum_{k=1}^\infty A_k\sin(k\omega t+\varphi_k)=\frac{a_0}{2}+\sum_{k=1}^\infty A_k\cos(k\omega t+\psi_k)$$

称为 $f(t)$ 的极坐标形式的傅里叶级数，其中

$$\begin{aligned}A_k&=\sqrt{a_k^2+b_k^2},\quad \omega=\frac{2\pi}{T}\\ \varphi_k&=\arctan\frac{b_k}{a_k},\quad \psi_k=\operatorname{arccot}\frac{b_k}{a_k}\end{aligned}$$

极坐标形式的傅里叶级数用得最少，但它证实了傅里叶的大胆猜测：每一个周期函数都可以表示为多个正弦函数的叠加. 此时的参数 A_k 和 φ_k 就具有了物理意义，前者表示振幅，后者表示初相；而最前面的 $a_0/2$ 就表示一个周期内函数的均值，用来修正上下平移.

- **导出 $[t_0, t_0+T]$ 上复数形式的傅里叶级数：基底 $\mathcal{A}=\{e^{in\omega t}\}_{n=-\infty}^\infty$**

根据欧拉公式替换三角函数：

$$\cos(t)=\frac{e^{it}+e^{-it}}{2},\qquad \sin(t)=-i\frac{e^{it}-e^{-it}}{2}$$

然后整理三角函数项的傅里叶级数表达式

$$\begin{aligned}S_\infty[f](t)&=\frac{a_0}{2}+\sum_{k=1}^\infty[a_k\sin(k\omega t)+b_k\sin(k\omega t)]\\&=\frac{a_0}{2}+\sum_{k=0}^\infty\left(a_k\frac{e^{ik\omega t}+e^{-ik\omega t}}{2}-ib_k\frac{e^{ik\omega t}-e^{ik\omega t}}{2}\right)\\&=\frac{a_0}{2}+\sum_{k=1}^\infty\left(\frac{a_k-ib_k}{2}e^{ik\omega t}+\frac{a_k+ib_k}{2}e^{-ik\omega t}\right)\end{aligned}$$

其中

$$\frac{a_k\pm ib_k}{2}=\frac{1}{T}\int_{t_0}^{t_0+T}f(t)\left[\cos(k\omega t)\pm i\sin(k\omega t)\right]\mathrm{d}t=\frac{1}{T}\int_{t_0}^{t_0+T}f(t)e^{\pm ik\omega t}\mathrm{d}t$$

回代，得到

$$f(t)=\frac{a_0}{2}+\sum_{k=0}^\infty\left[\frac{1}{T}\int_{t_0}^{t_0+T}f(t)e^{-ik\omega t}\mathrm{d}te^{ik\omega t}+\frac{1}{T}\int_{t_0}^{t_0+T}f(t)e^{ik\omega t}\mathrm{d}te^{-ik\omega t}\right]$$

若令 $c_k = \dfrac{1}{T}\displaystyle\int_{t_0}^{t_0+T} f(t)e^{-ik\omega t}\mathrm{d}t$，则易验证 $\dfrac{a_0}{2} = c_0$，此时就有

定义 复数项形式的傅里叶级数

定义在区间 $[t_0, t_0+T]$ 上的函数 $f(t)$，将

$$S_\infty[f](t) = \sum_{k=-\infty}^{+\infty} c_k e^{ik\omega t}$$

称为 $f(t)$ 的复数项形式的傅里叶级数，其中

$$c_k = \frac{1}{T}\int_{t_0}^{t_0+T} f(t)e^{-ik\omega t}\mathrm{d}t, \qquad \omega = \frac{2\pi}{T}$$

复数项的傅里叶级数，看起来比三角函数项的傅里叶级数简洁. 由推导公式可以看出，复数项级数的系数 $c_k = \dfrac{a_k - ib_k}{2}$，反过来有 $a_k = 2\mathrm{Re}[c_k]$，$b_k = -2\mathrm{Im}[c_k]$.

- **导出连续傅里叶级数：基底** $\mathcal{A} = \{e^{i\omega t}\}_{\omega\in\mathbb{R}}$

由于基底有不可数无穷多个，所以应该把求和换成积分；由于基底是复变函数，所以作内积的时候记得取共轭，即原函数在基底 $e^{i\omega t}$ 上的投影：

$$c_\omega = \langle f(t), e^{i\omega t}\rangle = \int_{\mathbb{R}} f(t)e^{-i\omega t}\mathrm{d}t$$

而原函数：

$$f(t) = \int_{\mathbb{R}} c_\omega e^{i\omega t}\mathrm{d}\omega$$

这其实就是连续傅里叶变换.

定义 傅里叶变换

对于函数 $f(t)$，

$$\hat{f}(\omega) = \mathcal{F}[f(t)](\omega) := \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\mathrm{d}t$$

如果存在，则称为 $f(t)$ 的傅里叶变换(Fourier transform)，其反变换为

$$f(t) = \mathcal{F}^{-1}[\hat{f}(\omega)](t) := \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega t}\mathrm{d}\omega$$

$f(t)$ 的定义域一般称为时域(time domain)，$\hat{f}(\omega)$ 的定义域一般称为频域(frequency domain).

太快啦，一下子推出了三种傅里叶级数，现在把脚步放慢，这里，我们需要验证一下基底的正交性，计算基底的模长.

2.0.1 傅里叶级数作为原函数的延拓

以上导出傅里叶级数的时候，特别强调了区间是 $[t_0, t_0+T]$，尽管在本章最开始时规定了 $f(t)$ 必须在整个 $\mathbb{R}$ 上有定义，但回顾刚才导出的过程，$f(t)$ 即使在该区间以外没有定义也不影响该区间内傅里叶级

数的存在. 与此同时，$S_\infty[f](t)$ 作为一系列三角函数的和，在区间外仍然有定义，并且考虑到这些三角函数的角频率是 $\omega, 2\omega, 3\omega, \cdots$、最小正周期是 $T, T/2, T/3, \cdots$，因此 T 是 $S_\infty[f](t)$ 是的最小正周期. 这意味着 $S_\infty[f](t)$ 会在整个 $\mathbb{R}$ 上不断重复着 $[t_0, t_0+T]$ 的故事，这就把一个可能只在原区间内连续的函数 $f(t)$ 伸展成了在 $\mathbb{R}$ 上定义的、以 T 为周期的函数 $S(t)$，即称为延拓(continuation).

若 $f(t)$ 是奇函数或偶函数，那么其定义区间关于原点对称，此时 $\{a_n\}$ 和 $\{b_n\}$ 有一串全是 0，此时傅里叶级数只剩下正弦（奇函数）或余弦项（偶函数），此时就称其为正弦级数或余弦级数.

如果被展开的函数 $f(t)$ 只在区间 $[0, T]$ 上有定义，用傅里叶级数可以给予其两种不同的延拓，一种是奇延拓，即构造奇函数 $g(t)$

$$g(t) = \begin{cases} f(t) & t \in [0, T] \\ -f(-t) & t \in [-T, 0] \end{cases}$$

另一种延拓方式是偶延拓

$$g(t) = \begin{cases} f(t) & t \in [0, T] \\ -f(-t) & t \in [-T, 0] \end{cases}$$

然后对 $g(t)$ 进行傅里叶展开，就可以得到正弦级数和余弦级数.

以下是一些常用周期函数的傅里叶展开

(1) 梯形波（奇函数）

$$f(t) \sim \frac{4A}{\pi\omega d} \sum_{n\,\mathrm{odd}} \sin\left(\frac{n\omega d}{n^2}\right) \sin(n\omega t)$$

(2) 矩形脉冲（偶函数）

$$f(t) \sim \alpha A + \frac{2A}{\pi} \sum_{n=1}^{\infty} \frac{\sin(\alpha n\pi)}{n} \cos(n\omega t)$$

(3) 方波（奇函数）

$$f(t) \sim \frac{4A}{\pi} \sum_{n\,\mathrm{odd}} \frac{\sin(n\omega t)}{n}$$

(4) 三角波（奇函数）

$$f(t) \sim \frac{8A}{\pi^2} \sum_{n\,\mathrm{odd}} \frac{i^{n-1} \sin(n\omega t)}{n^2}$$

(5) 锯齿波（非奇非偶）

$$f(t) \sim \frac{A}{2} - \frac{A}{\pi} \sum_{n=1}^{\infty} \frac{\sin(n\omega t)}{n}$$

2.1 傅里叶级和傅里叶变换的联系

刚才平行地导出了傅里叶级数和傅里叶变换，很快也很自然，但傅里叶级数和傅里叶变换的联系却被忽略了，这一节就来补上这部分的内容。

联系向量相加的几何意义，向量的基分解公式告诉我们，空间中任意一个向量 $\vec{\alpha}$ 都可以由每一个 $\vec{\varepsilon}_n$ 伸长或缩短为为原来的 a_n 倍之后首尾相接得到. 由于 $e^{in\omega t}$ 在复平面上表现为一个以角速度 $n\omega$ 旋转的单位向量. 所以复数项的傅里叶级数告诉我们，$f(t)$（更准确地说是起点位于原点、终点在实轴上 $f(t)$ 处的向量）可以由多个正在分别以角速度 $\{\cdots, -2\omega, -\omega, 0, \omega, 2\omega, \cdots\}$ 旋转、长度分别为 $\{\cdots, c_{-2}, c_{-1}, c_0, c_1, c_2, \cdots\}$ 的向量首尾相接得来（负的角速度表示方向相反），同时还能保证这些向量首尾相接之后终点恰好在实轴上.

然而这些旋转向量的角速度是分立的，存在角速度为 ω 的旋转向量，也存在角速度为 ω 的旋转向量，却不存在角速度为 1.5ω 的旋转向量. 角速度值最小间隔是 $\Delta\omega = \omega = \dfrac{2\pi}{T}$. 如果将 T 增大，那么 $\Delta\omega$ 就会

减小，当 $T\to\infty$ 时，$\Delta\omega\to 0$，这时可以认为 $f(t)$ 展开的基底是 $\mathcal{A}=\{e^{i\omega t}\}_{\omega\in\mathbb{R}}$，根据正交展开的思想，这时应该改写为

$$f(t)=\int_{\mathbb{R}} c_\omega e^{i\omega t}\mathrm{d}t$$

其中系数

$$c_\omega=\frac{\langle f(t),e^{i\omega t}\rangle}{\langle e^{i\omega t},e^{i\omega t}\rangle}=\int_{\mathbb{R}} f(t)\overline{e^{i\omega t}}\mathrm{d}t=\int_{\mathbb{R}} f(t)e^{-i\omega t}\mathrm{d}t$$

这就是傅里叶变换及其反变换，但你会发现 $\langle e^{i\omega t},e^{i\omega t}\rangle$ 似乎算不出来，所以为了严谨起见，以下仍通过定积分导出. 关于这个内积到底怎么算，接下来会通过严格定义某些东西把它彻底解决.

设 $\Delta\omega=\Omega_n-\Omega_{n-1}=\dfrac{2\pi}{T}$，把区间移回 $[-\dfrac{T}{2},\dfrac{T}{2}]$，然后复数项的傅里叶级数就可以写为

$$\begin{aligned}
f(t)&=\lim_{T\to\infty}\sum_{k=-\infty}^{+\infty} c_k e^{-ik\omega t}\\
&=\lim_{T\to\infty}\sum_{k=-\infty}^{+\infty}\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)e^{-ik\omega t}\mathrm{d}t\cdot e^{ik\omega t}\\
&=\lim_{\Delta\omega\to 0}\sum_{k=-\infty}^{+\infty}\left[\int_{-\frac{\pi}{\Delta\omega}}^{\frac{\pi}{\Delta\omega}} f(t)e^{-ik\omega t}\mathrm{d}t\right]e^{ik\omega t}\frac{\Delta\omega}{2\pi}
\end{aligned}$$

这个式子很长，但这恰好是定积分的定义，最终得到

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)e^{-i\omega t}\mathrm{d}t\right]e^{i\omega t}\mathrm{d}\omega$$

称为傅里叶积分公式. 现在可以给出定义了.

定义　傅里叶变换

对于函数 $f(t)$，

$$\hat{f}(\omega)=\mathcal{F}[f(t)](\omega):=\int_{-\infty}^{\infty} f(t)e^{-i\omega t}\mathrm{d}t$$

如果存在，则称为 $f(x)$ 的傅里叶变换(Fourier transform)，其中 $\mathcal{F}$ 表示傅里叶变换算子，其反变换为

$$f(t)=\mathcal{F}^{-1}[\hat{f}(\omega)](t):=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega t}\mathrm{d}\omega$$

$f(t)$ 的定义域一般称为时域(time domain)，$\hat{f}(\omega)$ 的定义域一般称为频域(frequency domain).

$\mathcal{F}$ 是一个定义在函数空间上的函数，即所谓的“函数的函数”，用类型论的语言来表示，由于 $f:\mathbb{R}\to\mathbb{R}$，$\hat{f}:\mathbb{R}\to\mathbb{R}$，所以 $\mathcal{F}:(\mathbb{R}\to\mathbb{R})\to(\mathbb{R}\to\mathbb{R})$.

仔细思考傅里叶变换的意义，为什么称 $\hat{f}(\omega)$ 的定义域为 “频域” 呢？回到复数项形式的傅里叶级数：

$$f(x)=\sum_{n=-\infty}^{\infty} c_n e^{in\omega x}$$

以上表明复数项的傅里叶级数的本质是函数在正交基底上的分解公式，那么傅里叶变换呢？在单位正交基底中，某个向量 $\vec{\alpha}$ 在某个基底向量上 $\vec{\varepsilon}_n$ 的投影为 $\vec{\alpha}_n=\langle\vec{\alpha},\vec{\varepsilon}_n\rangle$. 代入函数空间 $\mathbb{R}$ 中两个函数的内积

定义（至于为什么这样定义，参见 3.1 节），函数 $f(t)$ 在函数 $\psi_\xi(t)$ 上的投影：

$$f(\xi)=\langle f(t),\psi_\xi(t)\rangle=\int_{-\infty}^{\infty}f(t)\psi_\xi(t)\mathrm{d}t$$

同时看一看傅里叶变换的定义：

$$\hat{f}(\omega)=\int_{-\infty}^{\infty}f(t)e^{-i\omega t}\mathrm{d}t$$

两个公式的结构很相似，只要定义 $\psi_\xi(t)=e^{-i\xi t}$ 就在形式上完全相同了，不过想把两者真正联系起来，还有两点需要补充.

- $\psi_\xi(t)=e^{-i\xi t}$ 是单位向量，即模长为 1.
  证明：这是显然的.

- 当参数 $\xi\neq\rho$ 时，函数 $\psi_\xi(t)$ 和 $\psi_\rho(t)$ 是正交的.
  证明：证明 $\psi_\xi(t)$ 和 $\psi_\rho(t)$ 内积为 0 即可.

$$\langle\psi_\xi(t),\psi_\rho(t)\rangle=\int_{-\infty}^{\infty}\overline{\psi_\xi(t)}\psi_\rho(t)\mathrm{d}t=\int_{-\infty}^{\infty}e^{-i(\xi-\rho)t}\mathrm{d}t$$

  下一节会提到这个积分实际上等效于 δ 函数.

$$\langle\psi_\xi(t),\psi_\rho(t)\rangle\sim 2\pi\delta(\xi-\rho)$$

  因此只要 $\rho\neq\xi$ 内积就可认为是 0.

联系概率论的知识，当随机变量的取值变的稠密甚至连续的时候，就不适合用概率函数 $P(X)$ 而应该用概率密度函数 $p(x)$ 了，此处的思想也相同，当频率可取的值连续的时候，$\hat{f}(\omega)$ 表示的是频率的“密度”.

2.1.1 傅里叶变换的性质

傅里叶变换有以下性质 (记 $\hat{f}(\omega)=\mathcal{F}[f(t)](\omega)$)：

(1) 线性

$$\mathcal{F}\left[a_1f_1(t)+a_2f_2(t)\right](\omega)=a_1\hat{f}_1(\omega)+a_2\hat{f}_2(\omega)$$

证明：直接套定义，利用积分的线性性质

$$\begin{aligned}\mathcal{F}\left[a_1f_1(t)+a_2f_2(t)\right](\omega)&=\int_{\infty}^{\infty}\left[a_1f_1(t)+a_2f_2(t)\right]e^{-i\omega t}\mathrm{d}t\\&=a_1\int_{\infty}^{\infty}f_1(t)e^{-i\omega t}\mathrm{d}t+a_2\int_{\infty}^{\infty}f_2(t)e^{-i\omega t}\mathrm{d}t\\&=a_1\hat{f}_1(\omega)+a_2\hat{f}_2(\omega)\end{aligned}$$

(2) 共轭对称性

$$\mathcal{F}\left[\overline{f(t)}\right](\omega)=\overline{\hat{f}(-\omega)}$$

证明：由于 $f(t)=|f(t)|\exp(i\mathrm{Arg}[f(t)])$，所以

$$\mathcal{F}\left[\overline{f(t)}\right](\omega)=\int_{-\infty}^{\infty}|f(t)|e^{-i\mathrm{Arg}[f(t)]}e^{-i\omega t}\mathrm{d}t=\int_{-\infty}^{\infty}\overline{|f(t)|e^{i\mathrm{Arg}[f(t)]}e^{i\omega t}}\mathrm{d}t=\overline{\hat{f}(-\omega)}$$

(3) 时域-频域对偶

$$\mathcal{F}\left[\hat{f}(\omega)\right](t) = 2\pi f(-t)$$

证明：直接套定义

$$\mathcal{F}\left[\hat{f}(t)\right](\omega) = \int_{-\infty}^{\infty} \hat{f}(t) e^{i(-\omega)t} \mathrm{d}t = 2\pi \mathcal{F}^{-1}\left[\hat{f}(t)\right](-\omega) = 2\pi f(-\omega)$$

(4) 平移

$$\mathcal{F}\left[f(t-t_0)\right](\omega) = \hat{f}(\omega) e^{-i\omega t_0}$$
$$\mathcal{F}\left[\hat{f}(\omega-\omega_0)\right](t) = 2\pi f(t) e^{-i\omega_0 t}$$

证明：还是直接套定义

$$\mathcal{F}\left[f(t-t_0)\right](\omega) = \int_{-\infty}^{\infty} f(t-t_0) e^{-i\omega t} \mathrm{d}t$$

令 $\tau = t - t_0$，得

$$\mathcal{F}\left[f(t-t_0)\right](\omega) = \int_{-\infty}^{\infty} f(\tau) e^{-i\omega(\tau+t_0)} \mathrm{d}t = e^{-i\omega t_0} \int_{-\infty}^{\infty} f(\tau) e^{-i\omega(\tau)} \mathrm{d}t = e^{-i\omega t_0} \hat{f}(\omega)$$

(5) 缩放

$$\mathcal{F}\left[f(at)\right](\omega) = \frac{1}{|a|} \hat{f}\left(\frac{\omega}{a}\right)$$

(6) 全域积分值

$$\int_{-\infty}^{+\infty} f(t) \mathrm{d}t = \hat{f}(0)$$
$$\int_{-\infty}^{+\infty} \hat{f}(\omega) \mathrm{d}\omega = 2\pi f(0)$$

证明：在定义式中令 $\omega = 0$ 或 $t = 0$ 即可

(7) 微积分性质

$$\mathcal{F}\left[f^{(n)}(t)\right](\omega) = (i\omega)^n \hat{f}(\omega)$$
$$\mathcal{F}\left[(-it)^n f(t)\right](\omega) = \hat{f}^{(n)}(\omega)$$
$$\mathcal{F}\left[\int_{-\infty}^{t} f(t) \mathrm{d}t\right](\omega) = \pi \hat{f}(\omega) \delta(\omega) + \frac{\hat{f}(\omega)}{i\omega}$$
$$\mathcal{F}\left[-\frac{f(t)}{it} + \pi f(0) \delta(t)\right](\omega) = \int_{-\infty}^{\omega} \hat{f}(\omega) \mathrm{d}\omega$$

(8) 卷积定理

$$\mathcal{F}\left[f_1(t) * f_2(t)\right](\omega) = \hat{f}_1(\omega) \hat{f}_2(\omega)$$
$$\mathcal{F}\left[f_1(t) f_2(t)\right](\omega) = \frac{1}{2\pi}\left[\hat{f}_1(\omega) * \hat{f}_2(\omega)\right]$$

(9) **普朗歇尔定理 (Plancherel theorem)**

$$\int_{-\infty}^{\infty} f^2(t) \mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\hat{f}(\omega)\right|^2 \mathrm{d}\omega$$

2.2 傅里叶级数的收敛性讨论

这是最硬核的一节，需要做大量的数学推导．一般来说，当把函数展成级数之后，要讨论级数的收敛性，最初傅里叶提出这个级数以后，没有对其收敛性进行严格讨论，这些工作是由其学生狄利克雷完成的．尽管这部分工作并没有上文所讨论的傅里叶级数的提出与合理化那么具有开创性，但需要用到的数学工具却更加多而丰富．

2.2.1 吉布斯现象

对于有跳跃间断点的函数，观察它的傅里叶级数的逼近情况．以周期 $T=8$、高位值 $A=2$、高位占空 $\tau=4$ 的方波为例，一个周期内它的解析式可以写作

$$f(t)=A\left[\varepsilon\left(t+\frac{\tau}{2}\right)-\varepsilon\left(t+\frac{\tau}{2}\right)\right]=2\left[\varepsilon\left(t+2\right)-\varepsilon\left(t+2\right)\right]$$

其指数形式的傅里叶级数系数为

$$c_n=\frac{1}{T}\int_T f(t)e^{-in\omega t}\mathrm{d}t=\int_{-2}^{2}Ae^{-in\omega t}\mathrm{d}t=\frac{A}{in\omega T}\left[e^{-in\omega t}-e^{in\omega t}\right]=\mathrm{sinc}\frac{\pi n}{2}$$

因此

$$S_N[f](t)=\sum_{n=-N}^{N}\mathrm{sinc}\frac{\pi n}{2}e^{in\omega t}=1+2\sum_{n=0}^{N}\mathrm{sinc}\frac{\pi n}{2}\cos\frac{\pi t}{4}$$

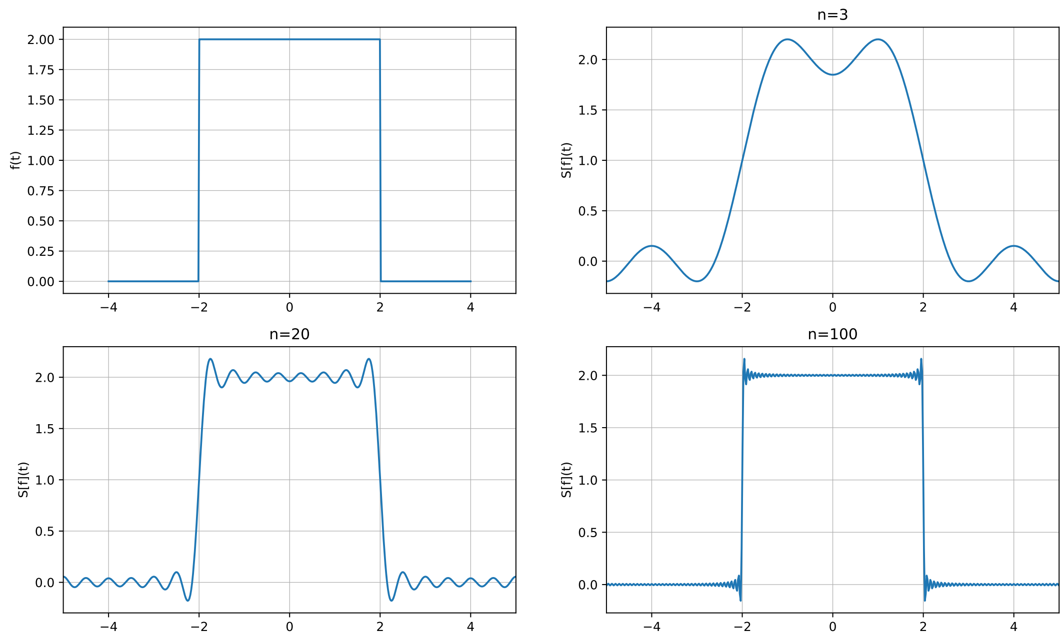

Figure 1: 吉布斯现象

把图像画出来，可以发现随着逼近程度的增加（即 n 不断增大），部分和在间断点两侧的值总是会出现峰值无法减小的震荡（这个峰值又称为超调量），超调量几乎与 n 无关，但是与跳跃的高度成正比．设跳跃的高度为 a，随着 n 的增加，超调量的值近似为 $\beta a=0.0895a$，即大约 9%．

用数学语言来描述，设被展开函数为 $f(x)$，其复数项傅里叶级数的部分和为 $S_N[f](x)$，即

$$S_N[f](x)=\sum_{n=-N}^{N}c_ke^{in\omega x}=\sum_{n=-N}^{N}\int_T f(t)e^{-in\omega t}\mathrm{d}te^{in\omega x}=\sum_{n=-N}^{N}\int_T f(t)e^{in\omega(x-t)}\mathrm{d}t$$

根据控制收敛定理，此时有 $\lim\limits_{n\to\infty}f_n(x)=f(x)$

定理　吉布斯现象 (Gibbs' phenomenon)

若 $f(x)$ 的周期为 T，跳跃间断点为 x_0，两端值之差 $f(x_0^+)-f(x_0^-)=a$，则有

$$\lim_{N\to\infty}S_N[f]\left(x_0+\frac{T}{2N}\right)=f(x_0^+)+\beta a$$

$$\lim_{N\to\infty}S_N[f]\left(x_0-\frac{T}{2N}\right)=f(x_0^-)-\beta a$$

其中 β 表示超调量相对于跳跃高度的比例，为

$$\beta=\frac{1}{\pi}\int_0^{\pi}\frac{\sin(x)}{x}\mathrm{d}x-\frac{1}{2}\approx 0.08948987\cdots$$

证明：设 $h(x)$ 为 $[0,T]$ 上的连续函数，且导函数也连续，在 $[0,T]$ 内随机选择两点 $a,b(b>a)$，记 $b-a=L$，然后构造一个以 T 为周期的函数 $f(x)$ 在 $[0,T]$ 内定义为

$$f(x)=\begin{cases}h(x), & 0<a\le x<b<T\\ 0, & \text{其他情况}\end{cases}$$

这时 $f(x)$ 就有了两个跳跃间断点 $x=a$ 和 $x=b$，接下来仅证明

$$\lim_{N\to\infty}S_N[f](a+\frac{T}{2N})=f(a^+)(1+\beta)$$

$f(a^-),f(b^+),f(b^-)$ 的情况类似.

$f(x)$ 的复数形式的傅里叶级数的系数

$$c_n=\frac{1}{T}\int_0^T f(t)e^{-in\omega t}\mathrm{d}t=\frac{1}{T}\int_a^b h(t)e^{-in\omega t}\mathrm{d}t$$

整理 $S_N[f](x)$，把求和项全变成正的

$$S_N[f](x)=c_0+\frac{1}{T}\sum_{n=1}^{N}\int_a^b h(t)\left(e^{in\omega(x-t)}+e^{-in\omega(x-t)}\right)\mathrm{d}t=c_0+\frac{2}{T}\sum_{n=1}^{N}\int_a^b h(t)\cos[n\omega(x-t)]\mathrm{d}t$$

对其使用分部积分

$$\begin{aligned}&\int_a^b h(t)\cos[n\omega(x-t)]\mathrm{d}t\\ &=\left[\frac{-1}{n\omega}h(t)\sin[n\omega(x-t)]\right]\Bigg|_a^b-\int_a^b h'(t)\left[\frac{-1}{n\omega}\sin[n\omega(x-t)]\mathrm{d}t\right]\\ &=\frac{1}{n\omega}h(a)\sin[n\omega(x-a)]-\frac{1}{n\omega}h(b)\sin[n\omega(x-b)]+\frac{1}{n\omega}\int_a^b h'(t)\sin[n\omega(x-t)]\mathrm{d}t\end{aligned}$$

再代回原式

$$S_N[f](x) = c_0 + \frac{h(a)}{\pi}\sum_{n=1}^{N}\frac{\sin[n\omega(x-a)]}{n} - \frac{h(b)}{\pi}\sum_{n=1}^{N}\frac{\sin[n\omega(x-b)]}{n} + \frac{1}{\pi}\sum_{n=1}^{N}\frac{1}{n}\int_a^b h'(t)\sin\left[n\omega(x-t)\right]\mathrm{d}t$$

于是

$$\begin{aligned}
S_N[f]\left(a+\frac{T}{2n}\right) &= c_0 + \frac{h(a)}{\pi}\sum_{n=1}^{N}\frac{1}{n}\sin\left(\frac{n\pi}{N}\right) - \frac{h(b)}{\pi}\sum_{n=1}^{N}\frac{1}{n}\sin\left(\frac{n\pi}{N}+n\omega L\right) \\
&\quad + \frac{1}{\pi}\sum_{n=1}^{N}\frac{1}{n}\int_a^b f'(t)\sin\left[\frac{n\pi}{N}+n\omega(t-a)\right]\mathrm{d}t \\
&= \frac{1}{T}\int_a^b h(t)\mathrm{d}t + \frac{h(a)}{\pi}I_1(N) + \frac{h(b)}{\pi}I_2(N) + \frac{1}{\pi}I_3(N)
\end{aligned}$$

接下来分别讨论这三段，第一段恰好可以凑出黎曼和

$$\begin{aligned}
I_1(\infty) &= \lim_{N\to\infty}\sum_{n=1}^{n}\frac{1}{n}\sin\left(\frac{n\pi}{N}\right) \\
&= \lim_{N\to\infty}\sum_{n=1}^{N}\frac{\pi}{N}\frac{\sin\left(\frac{n\pi}{N}\right)}{\frac{n\pi}{N}} = \pi\lim_{N\to\infty}\sum_{n=1}^{N}\frac{1}{N}\mathrm{sinc}\left(\frac{n\pi}{N}\right) \\
&= \pi\int_0^1\frac{\sin(\pi x)}{\pi x}\mathrm{d}x = \mathrm{Si}(\pi) = \pi\left(\frac{1}{2}+\beta\right)
\end{aligned}$$

至于第二段，需要给出一个引理

引理

$$\sum_{n=1}^{\infty}\frac{\sin(kn)}{n} = \frac{\pi-k}{2}$$

该引理的证明：构造 $(0,\pi]$ 上的函数

$$f(x) = \begin{cases} \dfrac{\pi-x}{2}, & x\in[0,\pi] \\ \dfrac{-\pi-x}{2}, & x\in(-\pi,0) \end{cases}$$

是奇函数，展成傅里叶级数之后只存在 sin 项

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin(nx)\mathrm{d}x = \frac{2}{\pi}\int_0^{\pi}\frac{\pi-x}{2}\sin(nx)\mathrm{d}x = \frac{1}{n}$$

因此在 $[0,\pi]$ 上 $\dfrac{\pi-x}{2} = \displaystyle\sum_{n=1}^{\infty}\frac{\sin(nx)}{n}$，引理证毕.

根据引理，有

$$\begin{aligned}
I_2(\infty) &= \lim_{N\to\infty}\sum_{n=0}^{N}\frac{1}{n}\sin\left(\frac{n\pi}{N}-n\omega L\right) = \lim_{N\to\infty}\sum_{n=0}^{N}\frac{1}{n}\sin\left[n\left(\frac{\pi}{N}-\omega L\right)\right] \\
&\sim \lim_{N\to\infty}\frac{\pi-\left(\frac{\pi}{N}-\omega L\right)}{2} = \pi\left(\frac{L}{T}-\frac{1}{2}\right)
\end{aligned}$$

第三段也可以根据引理得到

$$\begin{aligned}
I_3(\infty) &= \lim_{N\to\infty}\sum_{n=1}^{N}\frac{1}{n}\int_a^b f'(t)\sin\left[\frac{n\pi}{N}+n\omega(t-a)\right]\mathrm{d}t \\
&= \lim_{N\to\infty}\int_a^b h'(t)\sum_{n=1}^{N}\frac{1}{n}\sin\left[\frac{n\pi}{N}+n\omega(t-a)\right]\mathrm{d}t \\
(\text{根据引理}) &= \pi\int_a^b h'(t)\left(\frac{t-a}{T}-\frac{1}{2}\right)\mathrm{d}t \\
&= \frac{\pi}{T}\int_a^b h'(t)t\mathrm{d}t-\pi\left(\frac{a}{T}+\frac{1}{2}\right)\int_a^b h'(t)\mathrm{d}t \\
(\text{分部积分}) &= \frac{\pi}{T}\left[h(b)b-h(a)a-\int_a^b h(t)\mathrm{d}t\right]-\pi[h(b)-h(a)]\left(\frac{a}{T}+\frac{1}{2}\right)
\end{aligned}$$

回代，得到

$$\begin{aligned}
&\lim_{N\to\infty}S_N[f]\left(a+\frac{T}{2N}\right) \\
&=\frac{1}{T}\int_a^b h(t)\mathrm{d}t+h(a)\left(\frac{1}{2}+\beta\right)-h(b)\left(\frac{L}{T}-\frac{1}{2}\right) \\
&\quad+\frac{1}{T}\left[h(b)b-h(a)a-\int_a^b h(t)\mathrm{d}t\right]-[h(b)-h(a)]\left(\frac{a}{T}+\frac{1}{2}\right) \\
&=h(a)\left(\frac{1}{2}+\beta-\frac{a}{T}+\frac{a}{T}+\frac{1}{2}\right)+h(b)\left(\frac{L}{T}-\frac{1}{2}+\frac{b}{T}-\frac{a}{T}+\frac{1}{2}\right) \\
&=h(a)(1+\beta)
\end{aligned}$$

这就定量求出了吉布斯现象的超调量为 $\beta=\dfrac{\mathrm{Si}(\pi)}{\pi}-\dfrac{1}{2}\approx 0.08948987\cdots$.

2.2.2 狄利克雷核

狄利克雷核是一系列函数的统称，引入狄利克雷核可以更方便地讨论收敛性.

定义 狄利克雷核

函数

$$D_n(x)=\sum_{k=-n}^{n}e^{ikx}$$

称为 n 阶狄利克雷核(Dirichlet kernel).

狄利克雷核具有如下性质:

(1) 狄利克雷核可以写成闭合表达式

$$D_n(x)=\frac{\sin\left(\frac{2n+1}{2}x\right)}{\sin\left(\frac{1}{2}x\right)}$$

证明：根据欧拉公式

$$D_n(x)=\sum_{k=-n}^{n}e^{ikx}=1+\sum_{k=1}^{n}(e^{ikx}+e^{-ikx})=1+2\sum_{k=1}^{n}\cos(nx)$$

两边同乘 $\sin(x/2)$，用积化和差

$$
\begin{aligned}
\sin\left(\frac{x}{2}\right)D_n(x) &= \sin\left(\frac{x}{2}\right)+\sum_{k=1}^{n}2\sin\left(\frac{x}{2}\right)\cos(nx)\\
&= \sin\left(\frac{x}{2}\right)+\sum_{k=1}^{n}\left[\sin\left(\frac{2n+1}{2}x\right)-\sin\left(\frac{2n-1}{2}x\right)\right]\\
&= \cancel{\sin\frac{x}{2}}+\cancel{\sin\frac{3x}{2}}-\cancel{\sin\frac{x}{2}}+\cancel{\sin\frac{5x}{2}}-\cancel{\sin\frac{3x}{2}}+\cdots+\sin\frac{(2n+1)x}{2}\\
&= \sin\frac{(2n+1)x}{2}
\end{aligned}
$$

因此 $D_n(x)=\dfrac{\sin\left(\frac{2n+1}{2}x\right)}{\sin\left(\frac{1}{2}x\right)}$

(2) 傅里叶级数的部分和可以写成函数本身与 n 阶狄利克雷核的周期卷积

$$S_n[f](x)=(f*D_n)(x)$$

2.2.3 在原点处的间断点收敛性讨论

定义 利普希茨条件

若函数 $f(x)$ 在某个邻域 $B(x_0,\delta)$ 内满足

$$|f(x)-f(x_0)|\le C|x-x_0|^{\alpha}$$

则称函数 $f(x)$ 在 x_0 处满足 α 阶利普希茨条件(Lipschitz condition)；
如果 $f(x)$ 在定义域内处处满足利普希茨条件，则称 $f(x)$ 满足一致的(uniform) 利普希茨条件；
如果将邻域 $B(x_0,\delta)$ 改为单边邻域 $[x_0,x_0+\delta)$，则称 $f(x)$ 满足单边的(unilateral) 利普希茨条件.

试想 $\alpha>1$，则有 $\dfrac{|f(x)-f(x_0)|}{|x-x_0|}\le C|x-x_0|^{\alpha-1}$，再令 $x\to x_0$ 即得到 $f'(x)=0$，则限制了 $f(x)$ 必须为常值函数，这不是我们想要的结果，所以一般要求 $0<\alpha<1$. 接下来令函数 $f(x)$ 在 x_0 的左右两侧都满足单边利普希茨条件.

接下来详细证明 $f(x)$ 在 $x=0$ 处出现跳跃间断点（函数在 $x=0$ 处不连续但 $f(0^{\pm})$ 存在）时，$S_\infty[f](0)$ 会等于什么（剧透警告：会等于 $f(0^{\pm})$ 的算术平均）.

根据狄利克雷核的两个性质：$S_N[f](x)=\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi}f(x-t)D_N(t)\mathrm{d}t$ 和 $\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi}D_N(t)\mathrm{d}t=1$，可以得到下面这个式子

$$
\begin{aligned}
S_N[f](x)-f(x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}f(x-t)D_N(t)\mathrm{d}t-\frac{1}{2\pi}\int_{-\pi}^{\pi}f(x)D_N(t)\mathrm{d}t\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}[f(x-t)-f(x)]D_N(t)\mathrm{d}t
\end{aligned}
$$

令 $\overline{f}(x_0) = \dfrac{f(x_0^+) + f(x_0^-)}{2}$，根据狄利克雷核的奇偶性将式子整理一下

$$
\begin{aligned}
2\pi\left(S_N[f](x_0) - \overline{f}(x_0)\right) &= \int_{-\pi}^{\pi}\left[f(x_0 - t) - \frac{f(x_0+)}{2} - \frac{f(x_0^-)}{2}\right] D_N(t)\mathrm{d}t \\
&= \left(\int_{-\pi}^{0} + \int_{0}^{\pi}\right) f(x_0 - t) D_N(t)\mathrm{d}t - \left[\frac{f(x_0^+)}{2} + \frac{f(x_0^-)}{2}\right]\left(\int_{-\pi}^{0} + \int_{0}^{\pi}\right) D_N(t)\mathrm{d}t \\
(\text{奇偶性}) &= \int_{0}^{\pi}[f(x_0 - t) - f(x_0^-)]D_N(t)\mathrm{d}t + \int_{-\pi}^{0}[f(x_0 - t) - f(x_0^+)]D_N(t)\mathrm{d}t \\
&\quad + \cancel{\left[\frac{f(x_0^-)}{2} - \frac{f(x_0^+)}{2}\right]\int_{0}^{\pi} D_N(t)\mathrm{d}t} + \cancel{\left[\frac{f(x_0^+)}{2} - \frac{f(x_0^-)}{2}\right]\int_{-\pi}^{0} D_N(t)\mathrm{d}t} \\
&= \int_{0}^{\pi}[f(x_0 - t) - f(x_0^-)]D_N(t)\mathrm{d}t + \int_{-\pi}^{0}[f(x_0 - t) - f(x_0^+)]D_N(t)\mathrm{d}t \\
(\text{拆成 4 段}) &= \int_{0}^{\delta}[f(x_0 - t) - f(x_0^-)]D_N(t)\mathrm{d}t + \int_{\delta}^{\pi}[f(x_0 - t) - f(x_0^-)]D_N(t)\mathrm{d}t \\
&\quad + \int_{-\pi}^{-\delta}[f(x_0 - t) - f(x_0^+)]D_N(t)\mathrm{d}t + \int_{-\delta}^{0}[f(x_0 - t) - f(x_0^+)]D_N(t)\mathrm{d}t \\
&= I_1 + I_2 + I_3 + I_4
\end{aligned}
$$

接下来就分别研究这四段，首先进一步整理 I_3，注意到

$$
D_N(t) = \frac{\sin[(N+\frac{1}{2})t]}{\sin\frac{t}{2}} = \frac{\sin(Nt)\cos(\frac{t}{2}) + \cos(Nt)\sin(\frac{t}{2})}{\sin\frac{t}{2}} = \sin(Nt)\cot\frac{t}{2} + \cos(Nt)
$$

所以将 I_3 拆成两部分

$$
\begin{aligned}
I_3 &= \int_{-\pi}^{-\delta}[f(x_0 - t) - f(x_0^+)]D_N(t)\mathrm{d}t \\
&= \int_{-\pi}^{\delta}[f(x_0 - t) - f(x_0^+)]\cot\frac{t}{2}\sin(Nt)\mathrm{d}t + \int_{-\pi}^{\delta}[f(x_0 - t) - f(x_0^+)]\cos(Nt)\mathrm{d}t \\
&= \int_{-\pi}^{\pi}[f(x_0 - t) - f(x_0^+)]\cot\frac{t}{2}\chi_{[-\pi,-\delta]}(t)\sin(Nt)\mathrm{d}t + \int_{-\pi}^{\pi}[f(x_0 - t) - f(x_0^+)]\chi_{[-\pi,-\delta]}(t)\cos(Nt)\mathrm{d}t
\end{aligned}
$$

其中 $\chi_S(x)$ 是示性函数，定义为

$$
\chi_S(x) = \begin{cases} 1, & x \in S \\ 0, & x \notin S \end{cases}
$$

一般地，将傅里叶系数分成实部和虚部两部分，即

$$
\int_{-\pi}^{\pi} f(t)e^{iNt}\mathrm{d}t = \int_{-\pi}^{\pi} f(t)\cos(Nt)\mathrm{d}t + i\int_{-\pi}^{\pi} f(t)\sin(Nt)\mathrm{d}t
$$

而根据黎曼-勒贝格引理，当 $N \to \infty$ 时傅里叶系数趋于 0，其实部和虚部都要趋于 0，因此 $\lim\limits_{N\to\infty} I_3 \to 0$，同理 $\lim\limits_{N\to\infty} I_2 \to 0$. 接下来讨论 I_1 和 I_4，首先对 I_4 放缩一下

$$
\begin{aligned}
|I_4| &= \left|\int_{-\delta}^{0}[f(x_0 - t) - f(x_0^+)]D_N(t)\mathrm{d}t\right| = \left|\int_{-\delta}^{0}\frac{f(x_0 - t) - f(x_0^+)}{t}tD_N(t)\mathrm{d}t\right| \\
&\le \int_{-\delta}^{0}\left|\frac{f(x_0 - t) - f(x_0^+)}{t}uD_N(t)\right|\mathrm{d}t
\end{aligned}
$$

2.2.4 在任意处间断点的收敛性讨论

TBD

跨越千山万水，经过了好几页的数学推导，我们终于得到了傅里叶级数收敛的定理.

定理　傅里叶级数收敛定理/狄利克雷定理

若函数 $f(x)$ 的周期为 T，或仅在一个长度为 T 的区间上有定义，同时满足以下三个条件：

(1) $f(x)$ 在一个周期区间上绝对可积，即 $\displaystyle\int_T |f(x)|\mathrm{d}x < +\infty$.

(2) 在任意有界的区间内，$f(x)$ 只能存在有限个极值点.

(3) 在任意有界的区间内，$f(x)$ 只能存在有限个第一类间断点.

以上三者称为狄利克雷条件(Dirichlet conditions)，此时该函数可以延拓成的**一致收敛的**傅里叶级数 $S_\infty[f](x)$，满足：

(1) $S_\infty[f](x)$ 的定义域是 $\mathbb{R}$.

(2) 若 x 是 $f(x)$ 的连续点，那么 $S_\infty[f](x) = f(x)$.

(3) 若 x 是 $f(x)$ 的间断点，那么 $S_\infty[f](x) = \displaystyle\lim_{h\to 0} \frac{f(x+h)+f(x-h)}{2}$.

2.3 时域-频域的不确定性原理

TBD 傅里叶带宽定理、Gabor 极限，可以同海森堡不确定性原理相联系.

2.4 单位冲激函数（狄拉克 delta 函数）

讨论傅里叶变换怎能不提到它？毕竟它最初就随着傅里叶变换被首先提出来的（所以提出者不是狄拉克，但首先由狄拉克下了较为严格的定义）. 单位冲激函数(unit impulse function) 是一个非常特殊的函数，很难给出显式的定义，但其性质又决定了它具有广泛的用途.

2.4.1 几种描述和近似

最常见地，它由以下两个等式隐式地定义出来：

$$\begin{cases} \displaystyle\int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1 \\ \forall x \neq 0 : \delta(x) = 0. \end{cases}$$

根据积分的几何意义容易发现，它在 $x=0$ 处的值只能为正无穷大，这在黎曼积分的意义下是一种不可积的函数；即使 $x=0$ 看成无穷间断点，那么黎曼积分为 0，而不可能是 1. 黎曼积分不允许它存在，这充分说明了 δ 函数的特殊性.

至于这个函数是怎么被提出来的，最初傅里叶在《热分析理论》(*Théorie analytique de la chaleur*) 一书中考虑了如下积分，把它用现在的形式写下来：

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}\omega e^{i\omega x} \int_{-\infty}^{\infty} \mathrm{d}\alpha f(\alpha) e^{-i\omega x}$$

柯西把它改成了另一种更好看的形式

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{i\omega(x-\alpha)}\mathrm{d}\omega\right] f(\alpha)\mathrm{d}\alpha$$

可以看出，反常积分

$$\int_{-\infty}^{\infty} e^{i\omega(x-\alpha)}\mathrm{d}\omega$$

被单独提了出来，起到了筛选的作用，乘上适当的系数，命名为一个函数，就是现在的 δ 函数:

$$\delta(x-\alpha)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i\omega(x-\alpha)}\mathrm{d}\omega$$

把它代入傅里叶变换，得到

$$\mathcal{F}\left[\delta(x-\alpha)\right](\omega)=\int_{-\infty}^{\infty}\delta(x-\alpha)e^{-i\omega x}\mathrm{d}x=e^{-i\omega\alpha}$$

当 $\alpha=0$ 时积分值为 1，因此 δ 函数的傅里叶变换为常函数 $\hat{f}(\omega)\equiv 1$.

尽管 δ 函数的性质比较特殊，但可以利用某些“正常”的函数，通过调整某些参数，来逼近 δ 函数. 严格地说，我们希望找到“正常”函数组成的序列 $\{f_n(x)\}_{n=1}^{\infty}$，满足以下几条类 δ 函数的性质:

(1) 最值在 $x=0$ 处取到:

$$f_n(x)_{\max}=f_n(0)$$

(2) 全域积分为 1:

$$\int_{-\infty}^{\infty}f_n(x)\mathrm{d}x=1$$

(3) 逼近性质:

$$\lim_{n\to\infty}f_n(0)=+\infty$$

随着 n 的增大，$f_n(x)$ 的形态就越来越接近 δ 函数，以下列举几个例子:

(1) 采样函数

$$f_n(x)=\frac{\sin(nx)}{\pi x}$$

(2) 正态分布钟形曲线

$$f_n(x)=\frac{1}{\sqrt{\pi n}}\exp\left(-\frac{x^2}{n}\right)$$

(3) 泊松核

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

(4) 柯西主值（其实等价于采样函数）

$$f_n(x)=\int_{-n}^{n}e^{i\omega x}\mathrm{d}\omega$$

2.4.2 关于积分定义的问题

以上提到，δ 函数是 $f_n(x)=\dfrac{1}{2\pi}\displaystyle\int_{-n}^{n}e^{i\omega x}\mathrm{d}\omega$ 在 $n\to\infty$ 时的形式极限，即可以说:

$$\delta(x)\sim\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{i\omega x}\mathrm{d}\omega$$

现在来仔细计算这个积分. 当 $x=0$ 时，$e^{i\omega x}\equiv 1$，因此积分值为正无穷，符合 δ 函数的特征. 但当 $x\neq 0$ 时，有

$$\int_{-\infty}^{+\infty}e^{i\omega x}\mathrm{d}\omega=\lim_{a\to+\infty}\int_{-a}^{+a}e^{i\omega x}\mathrm{d}\omega=\lim_{a\to+\infty}\left.\frac{1}{ix}e^{i\omega x}\right|_{-a}^{a}=\lim_{a\to+\infty}\frac{e^{iax}-e^{-iax}}{2ix}=\lim_{a\to+\infty}\frac{2}{x}\sin(ax)$$

这个极限是不存在的，因此利用积分直接定义 δ 函数其实是有问题的，当然也有如下狡辩:

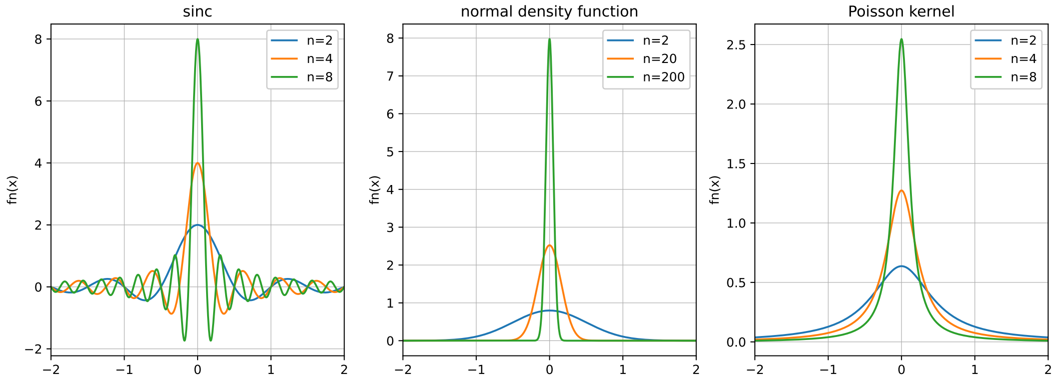

Figure 2: δ 函数的逼近

- **狡辩 1** 可改写以上极限，

$$\lim_{a\to+\infty}\frac{2}{x}\sin(ax)=\lim_{a\to+\infty}2\pi\frac{\sin(ax)}{\pi x}$$

其中 $\lim\limits_{a\to+\infty}\dfrac{\sin(ax)}{\pi x}\sim\delta(x)$，因此原极限收敛于 $2\pi\delta(x)$.

- **狡辩 2** 可改写以上积分，

$$\begin{aligned}\int_{-\infty}^{+\infty}e^{i\omega x}\mathrm{d}\omega&=\lim_{\varepsilon\to0}\int_{-\infty}^{+\infty}e^{i\omega x}e^{-\varepsilon\omega^2}\mathrm{d}\omega=\lim_{\varepsilon\to0}\exp\left(-\frac{x^2}{4\varepsilon}\right)\int_{-\infty}^{+\infty}\exp\left[-\varepsilon\left(\omega-\frac{ix}{2\varepsilon}\right)^2\right]\mathrm{d}\omega\\&=\lim_{\varepsilon\to0}\sqrt{\frac{\pi}{\varepsilon}}\exp\left(-\frac{x^2}{4\varepsilon}\right)\sim2\pi\delta(x)\end{aligned}$$

其中最后一个等号来源于后面要介绍的高斯积分.

利用该积分等效于 δ 函数是可行的，但是单独将其拿出来就不严谨了.

2.4.3 冲激函数的严格定义

严格定义 δ 函数是一件困难的事情，因为它需要较多的代数和拓扑的知识储备，这一节将把这些储备抖出一部分. 首先定义拓扑的概念.

- **定义 (拓扑)**

令集合 $\mathcal{T}$ 为 S 的子集构成的集合，同时满足以下三个条件：

(1) $\emptyset\in\mathcal{T},\quad S\in\mathcal{T}$

(2) 若 $x,y\in\mathcal{T}$，则 $x\cap y\in\mathcal{T}$

(3) 对于任意 $i\in I$，若 $x_i\in\mathcal{T}$，则 $\bigcup x_i\in\mathcal{T}$.

$\mathcal{T}$ 称为 S 的拓扑(topology)，$\mathcal{T}$ 中的元素称为开集(open set)，其补集称为闭集(closed set)，S 连同拓扑 $\mathcal{T}$ 称为拓扑空间(topological space).

看起来很抽象，但只要考虑一个特例 $\mathbb{R}$，将开集类比为开区间，逐条讨论起来就直观多了.

首先“形式上”有 $(a,a)=\{x:a<x<a\}=\emptyset$，因此空集可以看做是一个开区间. $\mathbb{R}$ 可以表示为 $(-\infty,+\infty)$，因此实数集本身也是一个开区间，第一条说的就是这两个特殊的对象是开集. 但需要注意的是，这两个家伙也是闭集.

有限个开区间的交集一定是开区间，但无限个开区间的交集可能是闭区间，比如 $\{(a-\frac{1}{n},b+\frac{1}{n})\}_{n=1}^{\infty}$ 这无穷个开区间的交集即为闭区间 $[a,b]$，第二条说的就是有限个开集的交集也是开集. 不信你可以在 $[a,b]$ 之外任取一个数，例如 $b+\varepsilon$，总存在 n 使得 $b<b+\frac{1}{n}<b+\varepsilon$，因此这个 $b+\varepsilon$ 就不属于这些集合的交集.

并集则不同，即使无限多个并集，无论是可数无限还是不可数无限，它们的并集的边界都是“开”的，例如并集 $(1,2)\cup(3,4)\cup(5,6)$，尽管不连通，但它们的边界都是“开”的，都是小括号，因此这是一个开集，又例如区间 $\{(0,n)\}_{n=1}^{\infty}$，取了并集就是 $(0,+\infty)$，也是个开集. 第三条把这些边界为“开”的区间纳入了开集的范围内.

由上所述，实数集以及全体边界为“开”的区间组成的集合组成了一个拓扑空间，但以实数集为基础的拓扑不止这一个，闭区间也可以包含在内，实数集的全体子集也可以组成一个拓扑，因此哪怕把范围限制在 $\mathbb{R}^n$ 上，拓扑和开集也是一个非常抽象而广泛的概念，为了接下来的讨论更为方便，现在来明确 $\mathcal{T}$ 的内容.

- **定义（标准拓扑）**
  拓扑空间 $\langle S,\mathcal{T}\rangle$ 的结构如下：
  (1) 拓扑空间 $S=\mathbb{R}^n$，n 为任意正整数.
  (2) 对于 $\mathbb{R}^n$ 中的任意两个元素 $x=(x_1,x_2,\cdots,x_n)$ 和 $y=(y_1,y_2,\cdots,y_n)$，度量规定为
  $$d(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2}$$
  (3) 对于 $\mathbb{R}^n$ 中的任意元素，开球定义为
  $$B_r(x_0)=\{x:d(x,x_0)<r\}$$
  (4) 对于任意 $\mathbb{R}$ 的子集 U，若存在开球 $B_r(x_0)\subseteq U$，则 $U\in\mathcal{T}$.

  由此得来的 $\langle S,\mathcal{T}\rangle$ 称为标准拓扑(standard topology).

标准拓扑中的开集限定得比刚才所说的“边界为‘开’的区间”更严格，除了空集和 $\mathbb{R}^n$ 本身，对于 $\langle\mathbb{R},\mathcal{T}\rangle$，开区间就是开集；对于 $\langle\mathbb{R}^2,\mathcal{T}\rangle$，单连通域去除其围线后的区域才能称为开集；对于 $\langle\mathbb{R}^3,\mathcal{T}\rangle$，封闭几何体去除其表面后的东西才能称为开集. 此时“开集”的概念已经非常直观了.

- **定义（$\mathbb{R}^n$ 中的覆盖和紧集）**
  在拓扑空间 $\langle\mathbb{R}^n,\mathcal{T}\rangle$ 中，设 S 是集合，$\{S_n\}_{n\in I}$ 是一串开集，若 $S\subseteq\bigcup S_i$，则称 $\{S_n\}_{n\in I}$ 是集合 S 的一个开覆盖.
  若闭集 S 的的每一个开覆盖都有有限子覆盖，则称 S 是紧集(compact set).

覆盖的定义很形象，这一串开集就像一块块布去遮挡集合 S，每一块布料的“面积”可以有限也可以无限，每块布之间可以重叠也可以不重叠，如果能完全遮住，则这一堆布料就称为 S) 的开覆盖. 如果无论这些布料遮盖的方式如何，我都可以抽走某些布料，使得剩下有限块布料也能完全覆盖 S，那么这个 S 就被称为紧集. 可见，在 $\mathbb{R}^n$ 中，紧集同有界闭集等价，因为如果集合是无界的，而每块布料都是“有限大”的，那么就不可能使用有限块布料覆盖该集合.

- **定义（连续映射）**
  设有两个拓扑空间 $\langle X,\mathcal{T}_1\rangle$ 和 $\langle Y,\mathcal{T}_2\rangle$，以及映射 $f:X\to Y$，若对于任意 $V\in\mathcal{T}_2$，其原象
  $$\{x:f(x)\in V\}\in\mathcal{T}_1$$
  则称该映射 f 在拓扑意义上是连续的(continuous).

- **定义（标准拓扑上的连续映射）**
  设映射 $f:\mathbb{R}^m\to\mathbb{R}^n$，若对于任意开集 $U\subseteq\mathbb{R}^n$，若其原象 $f^{-1}(U)$ 也是开集，则称该映射 f 在拓扑意义上是连续的(continuous).

其实标准拓扑上的连续和“一致连续”这两个概念是等价的，证明略. 如果一个映射建立在向量拓扑空间中，而且对于该向量空间是线性的，对于该拓扑空间是连续的，则称该映射为“**连续线性映射**”.

- **定义（对偶空间和 weak* 拓扑）**
  若 $\langle S,\mathcal{T}\rangle$ 是向量拓扑空间，则所有 $S\to\mathbb{R}$ 的连续线性映射组成的集合称为 S 的对偶空间(dual space)，记作 X^*
  在 X^* 上建立拓扑，并定义映射 $f:X^*\to\mathbb{R}$，使得 f 是连续函数的包含开集最少的拓扑称为 X^* 的 waek* 拓扑.

weak* 拓扑的定义有点抽象.

基础知识已经铺好，接下来开始正式定义 δ 函数. 由于 δ 函数在 $\mathbb{R}^n$ 上起作用，所以只用考虑其上的拓扑结构和函数空间.

- **定义（函数的极限）**
  将 $\mathbb{R}^n$ 上全体具有紧致支撑集的任意次可微函数的集合记作 $C_c^\infty(\mathbb{R}^n)$，对于其中的函数序列 $\{\phi\}_{n=1}^\infty$，若存在 $\phi\in C_c^\infty(\mathbb{R}^n)$，使得

  (1) 存在紧致集 K，使得 ϕ_n 的支撑集都是 K 的子集.

  (2) 随着 n 的增加，ϕ_n 以及它们的任意阶导数均收敛于 ϕ 及其对应阶导数.

  则称函数序列 $\{\phi_n\}_{n=1}^\infty$ 收敛于 ϕ，记作 $\lim\limits_{n\to\infty}\phi_n=\phi$. 并规定 $C_c^\infty(\mathbb{R}^n)$ 上的拓扑结构为满足此收敛方式且开集最少的拓扑.

- **定义（$\mathbb{R}^n$ 中的支撑集）**
  对于函数 $f:\mathbb{R}^n\to\mathbb{R}$，使得 $f(x)\neq 0$ 的全体 x 的集合称为该函数的支撑集(support set).

2.5 采样和插值

2.5.1 采样定理

采样定理是从连续到离散迈出的第一步，这个定理告诉我们，只要 $f(t)$ 存在最高频率，那么可以等间距地记录 $f(t)$ 的值，这些记录下来的离散的值可以完全确定连续的 $f(t)$.

(以下推导过程是香农的原始思路)

假设 $f(t)$ 的最高角频率是 ω_H，意思是当 $\omega>\omega_H$ 时，均有 $\hat{f}(\omega)=0$（不一定无定义），那么在傅里叶逆变换时可以使用 ω_H 来代替其积分上下界，即

$$f(t)=\mathcal{F}^{-1}[\hat{f}(\omega)](t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega t}\mathrm{d}\omega=\frac{1}{2\pi}\int_{-\omega_H}^{\omega_H}\hat{f}(\omega)e^{i\omega t}\mathrm{d}\omega$$

若令 $B=2\omega_H$ 表示双边带宽，则

$$\frac{2\pi}{B}f\left(-\frac{2\pi k}{B}\right)=\frac{1}{B}\int_{-B/2}^{B/2}\hat{f}(\omega)e^{\frac{i2\pi k\omega}{B}}\mathrm{d}\omega$$

将 $\hat{f}(\omega)$ 延拓成复数项形式的傅里叶级数（周期自然是 B，另外这里 ω 是哑变量，所以在展开式中直接写成 $2\pi/B$），接着可以发现系数 c_k 能被换掉

$$\hat{\hat{f}}(\omega) = \sum_{k=-\infty}^{+\infty} c_k e^{\frac{i2\pi k\omega}{B}} = \sum_{k=-\infty}^{+\infty} \left[\frac{1}{B}\int_B \hat{f}(\omega)e^{-\frac{i2\pi k\omega}{B}}\mathrm{d}\omega\right] e^{\frac{i2\pi k\omega}{B}} = \sum_{k=-\infty}^{+\infty} \frac{2\pi}{B} f\left(-\frac{2\pi k}{B}\right) e^{\frac{i2\pi k\omega}{B}}$$

仔细思考这个等式的含义，左边是 $f(t)$ 的频谱函数，它经过反傅里叶变换即可得到完整的 $f(t)$，右边是个无穷级数，参与到该无穷级数运算的有关于 $f(t)$ 的东西只有 $f(0), f(\pm 2\pi/\omega_B), f(\pm 4\pi/\omega_B), f(\pm 6\pi/\omega_B), \cdots$ 这些离散的值. 这就意味着，只要知道了这些采样值，就可以知道 $f(t)$ 每一点的值，而且是精确值.
在给出最终定理之前，再把式子整理一下. 考虑到 B 表示角频率（的差值），那么 $\frac{2\pi}{B}$ 就具有了时间间隔的含义. 如果令 $T_s = \frac{2\pi}{B}$，并设 $f[n] = f(nT_s)$，再做个简单换元，就得到了时域采样定理.

定理 奈奎斯特-香农采样定理 (Nyquist-Shannon sampling theorem)/低通采样定理

如果一个函数 $f(t)$ 的角频率不超过 ω_H，则 $f(t)$ 可以被一系列间距为 $T_s = \frac{\pi}{\omega_H}$ 的采样点确定. 此时 $f(t)$ 的双边频带宽度 $B = 2\omega_H$ 也称为奈奎斯特抽样率(Nyquist sampling rate)，T_s 称为奈奎斯特抽样间隔(Nyquist sampling interval). 更进一步地，该函数对应的频域表达式为

$$\hat{f}(\omega) = T_s \sum_{n=-\infty}^{\infty} f[n]e^{-iT_s n\omega}$$

换个角度想，为什么离散的采样值能
对偶地，有频域上的函数也有采样定理

推论 频域采样定理

如果时域函数 $f(t)$ 仅在长度为 T 的区间内有定义，记 $\omega_0 = \frac{2\pi}{T}$ 为频域采样间隔，$\hat{f}[n] = \hat{f}(n\omega_0)$ 为频域采样序列，则根据这个序列可确定延拓出完整的时域函数

$$\tilde{f}(t) = \frac{1}{T}\sum_{n=-\infty}^{+\infty} \hat{f}[n]e^{in\omega_0 t}$$

其中 $\tilde{f}(t)$ 是 $f(t)$ 的以 T 为周期的延拓.

接下来这条推论非常重要，它非常宏观地刻画了时域和频域的对偶性：时域和频域只要有一个域是离散的，那么另一个域就是连续且周期的的，反之成立. 这同时也否定了后面出现的离散傅里叶变换精确刻画原函数的可能性.

推论 周期性对偶

以 T_s 为采样间隔的序列所对应的频域函数具有周期 $\frac{2\pi}{T_s}$，以 ω_0 为频域采样间隔的序列对应的时域函数具有周期 $\frac{2\pi}{\omega_0}$.

证明：直接代入即可

$$\begin{aligned}\hat{f}\left(\omega+\frac{2\pi}{T_s}\right)&=T_s\sum_{n=-\infty}^{\infty}f[n]\exp\left(-iT_sn(\omega+\frac{2\pi}{T_s})\right)=T_s\sum_{n=-\infty}^{\infty}f[n]\exp\left(-iT_sn\omega+i2n\pi\right)\\&=T_s\sum_{n=-\infty}^{\infty}f[n]\exp(-iT_sn\omega)=\hat{f}(\omega)\end{aligned}$$

频域的证明类似.

2.5.2 插值公式：从采样点恢复原信号

采样定理好是好，但想从采样点还原完整的 $f(t)$，要先计算一波无穷级数，然后再反傅里叶变换才行. 为何不现在就完成这个过程呢？上式两边同时求傅里叶反变换，得到

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega t}\mathrm{d}\omega=\frac{1}{2\pi}\int_{-\omega_H}^{\omega_H}\hat{f}(\omega)e^{i\omega t}\mathrm{d}\omega=\int_{-\infty}^{\infty}\sum_{k=-\infty}^{+\infty}\frac{1}{\omega_B}f\left(-\frac{2\pi k}{\omega_B}\right)e^{\frac{i2\pi k\omega}{\omega_B}+i\omega t}\mathrm{d}\omega$$

根据控制收敛定理[13]，交换积分与求和的顺序，同时提出与被积变量无关的部分

$$f(t)=\frac{1}{\omega_B}\sum_{k=-\infty}^{\infty}f\left(\frac{-2k\pi}{\omega_B}\right)\int_{-\omega_H}^{\omega_H}e^{\frac{i2\pi k\omega}{\omega_B}+i\omega t}\mathrm{d}\omega$$

现在来看其中的积分项，恰好可以化成 $\operatorname{sinc}(x)$ 的形式

$$\begin{aligned}\int_{-\omega_H}^{\omega_H}e^{\frac{i2\pi k\omega}{\omega_B}+i\omega t}\mathrm{d}\omega&=\int_{-\omega_H}^{\omega_H}\exp\frac{i\omega(2k\pi+\omega_Bt)}{\omega_B}\mathrm{d}\omega=\frac{\omega_B}{i(2k\pi+\omega_Bt)}\exp\frac{i\omega(2k\pi+\omega_Bt)}{\omega_B}\Bigg|_{-\omega_H}^{\omega_H}\\&=\omega_B\cdot\frac{\exp\left(i\frac{2k\pi+\omega_Bt}{2}\right)-\exp\left(-i\frac{2k\pi+\omega_Bt}{2}\right)}{\frac{2k\pi+\omega_Bt}{2}\cdot 2i}=\omega_B\operatorname{sinc}(k\pi+\omega_Ht)\end{aligned}$$

将其带回上式，就是采样定理的副产品：

定理 惠特克-香农插值公式 (Whittaker-Shannon interpolation formula)

如果一个函数 $f(t)$ 的角频率不超过 ω_H，令 $T_s=\dfrac{\pi}{\omega_H}$ 为奈奎斯特采样间隔，再令 $f[n]=f(nT_s)$ 作为采样序列，则

$$f(t)=\sum_{k=-\infty}^{\infty}f[n]\operatorname{sinc}\left(\omega_H(t-nT_s)\right)$$

之所以称为插值公式，连续函数 $f(t)$ 可以看做 $f[n]$ 的延拓，因为当 $T_s=1$ 时，有 $f(n)=f[n]$.

2.5.3 离散时间傅里叶变换

采样定理说明了，完整的频域表达式可以由时域上间隔为 T_s 的采样点确定. 采样点组成的序列相当于取值限定在 $nT_s, n\in\mathbb{Z}$ 的函数，而数列本身是取值限定在整数的函数. 若令 $T_s=1$，那么采样点序列也就成了数列. 这告诉我们，不仅是连续函数，数列也可以做傅里叶变换，为了有别于连续函数的傅里叶变换，时域采样序列的傅里叶变换称为“离散时间傅里叶变换”.

[13] $\mathbb{C}$ 上的控制收敛定理 (dominated convergence theorem)：设 $\{f_n(x)\}_{n=0}^{\infty}$ 是一系列可测的且逐点收敛于 $f(x)$ 的函数，若存在实函数 $g(x)$ 使得任意 n 均有 $|f_n(x)|<g(x)$，则有

$$\lim_{n\to\infty}\int f_n(x)\mathrm{d}x=\int f(x)\mathrm{d}x$$

定义　离散时间傅里叶变换

令 $\{f[n]\}_{n=-\infty}^{+\infty}$ 为双边无限长的数列，且绝对可和，则

$$\hat{f}(\omega)=\mathcal{F}\left[f[n]\right](\omega):=\sum_{n=-\infty}^{+\infty} f[n]e^{in\omega}$$

称为该数列的离散时间傅里叶变换(discrete-time Fourier transformation, DTFT)，其逆变换为

$$f[n]=\mathcal{F}^{-1}\left[\hat{f}(\omega)\right][n]:=\int_{-\pi}^{\pi}\hat{f}(\omega)e^{in\omega}\mathrm{d}\omega$$

2.6　离散傅里叶变换

众所周知，随机的连续函数承载无限的信息，无法用有限个字符描述出来. 哪怕经过采样定理的优化，时域和频域总有一个域是连续的，另一个虽然是离散的，但也是无限长且没有周期性的，仍然不能存储. 但假如给离散序列取一个有限长但长度很大的子序列，并抛弃其他部分，然后让这个子序列周期循环形成新的无限长序列，就可以在一定程度上模拟原来的无限长序列，同时还能保证在另一个域也是周期且离散的. 在周期且离散的序列之间做的时域-频域变换，就是离散傅里叶变换.

定义　离散傅里叶变换

对于任意一个有限长序列 $\{f[n]\}_{n=0}^{N}$，定义

$$\hat{f}[k]=\mathcal{F}[f[n]][k]=\sum_{n=0}^{N-1} f[n]\exp\left(-i\frac{2\pi kn}{N}\right),\qquad 0\geq k\geq N-1$$

称为序列 $f[n]$ 的离散傅里叶变换(discrete Fourier transform, DFT)，其逆变换为

$$f[n]=\mathcal{F}^{-1}\left[\hat{f}[k]\right][n]=\frac{1}{N}\sum_{k=0}^{N-1}\hat{f}[k]\exp\left(i\frac{2\pi kn}{N}\right),\qquad 0\geq n\geq N-1$$

其中记 $W_N=\exp\left(\frac{-2\pi i}{N}\right)$，称为旋转因子.

其中旋转因子 W_N，形象地理解，就是 N 个单位向量按角度平分了单位圆，以与实轴共线的单位向量为第 0 个，W_N^k 就是顺时针数的第 k 个单位向量. W_N 很大程度上决定了离散傅里叶变换的性质以及快速算法的可能性，接下来讨论它的性质，其中出现的变量 k,n,N 等默认为整数.

(1) 中心对称性：$\overline{W_N^n}=W_N^{N-n}$.

(2) 周期性：$W_N^n=W_N^{n+N}$.

(3) 可约性：$W_N^n=W_{kN}^{kn}$.

(4) 正交性：$\displaystyle\sum_{k=0}^{N-1}W_N^{kn}=\begin{cases}N, & k|N\\ 0, & k\nmid N\end{cases}$

证明：从几何的角度，当 $k|N$ 时，$W_N^{kn}\equiv 1$，此时的求和即为 $N\times 1=N$，当 $k\nmid N$ 时，就套入等比数列求和公式：

$$\sum_{k=0}^{N-1}W_N^{kn}=\sum_{k=0}^{N-1}\exp\left(\frac{-2\pi kni}{N}\right)=\frac{1-\exp\left(\frac{-2\pi \cancel{N}ni}{\cancel{N}}\right)}{1-\exp\left(\frac{-2\pi ni}{N}\right)}=0$$

因此离散傅里叶变换有以下性质，记 $\hat{f}[k]=\mathcal{F}[f[n]][k]$

(1) 线性

$$\mathcal{F}[f_1[n]+f_2[n]][k]=\hat{f}_1[k]+\hat{f}_2[k]$$

证明略.

(2) 可用正变换计算逆变换

$$\mathcal{F}^{-1}\left[\hat{f}[k]\right][n]=\frac{1}{N}\overline{\mathcal{F}\left[\overline{\hat{f}[k]}\right][n]}$$

(3) 时域-频域对偶性质

$$\frac{1}{N}\mathcal{F}\left[\hat{f}[n]\right][k]=f[-k]$$

(4) 取相反数

$$\mathcal{F}[f[-n]][k]=\hat{f}[-k]$$

(5) 域的求和

$$\sum_{n=0}^{N-1}f[n]=\hat{f}[0]$$

$$\sum_{k=0}^{N-1}\hat{f}[k]=Nf[0]$$

(6) 序列加长在 $f[n]$ 末端添加若干个 0 使其长度增加到 rN，即令

$$g[n]=\begin{cases}f[n], & 0\geq n\geq N-1\\ 0, N\geq n\geq rN-1\end{cases}$$

则

$$\mathcal{F}[g[n]]=\hat{f}\left[\frac{k}{r}\right]$$

(7) 圆周移位定理

$$\mathcal{F}\left[\tilde{f}[n+m]\right]=W_N^{-km}\hat{\tilde{f}}[k]$$

(8) 圆周卷积定理

(9) 圆周相关定理

(10) 帕斯瓦尔定理

(11) 圆周对称性

这一条性质内容非常丰富，首先讨论序列的共轭对称分解：

$$\begin{aligned}f[n]&=\frac{1}{2}(2f[n])=\frac{1}{2}\left(2f[n]+\overline{f[n]}-\overline{f[n]}\right)\\&=\frac{1}{2}\left(f[n]+\overline{f[-n]}\right)+\frac{1}{2}\left(f[n]-\overline{f[-n]}\right)\end{aligned}$$

其中红色的部分取共轭之后仍然等于它本身（称为共轭对称性），蓝色的部分取共轭之后等于它的相反数（称为共轭反对称性），于是定义

定义　序列的共轭对称分解

对于任意一个序列 $f[n]$ 均有 $f[n] = f_e[n] + f_o[n]$，其中

$$f_e[n] = \frac{1}{2}\left(f[n] + \overline{f[-n]}\right)$$

称为 $f[n]$ 的共轭对称分量，

$$f_o[n] = \frac{1}{2}\left(f[n] - \overline{f[-n]}\right)$$

称为 $f[n]$ 的共轭反对称分量.

2.7　快速（离散）傅里叶变换

2.7.1　基 2-FFT 算法

TBD

2.7.2　基 4-FFT 算法

TBD

2.7.3　库利-图基算法

TBD

2.7.4　Chirp-Z

TBD

2.8　拉普拉斯变换：给傅里叶变换打补丁

TBD

2.9　Z 变换：给离散傅里叶变换打补丁

TBD

2.10　设计低通滤波器

TBD

第五部分　特殊函数博物馆

1　伯努利数和伯努利多项式

1.1　伯努利数引入

伯努利数不是一个数，是一个数列 $\{B_n\}_{n=0}^{\infty}$，也算是函数。这个数列有很多种导出的方法，但还是伯努利自己的定义最好理解："数列的第 k 项是自然数前 n 项 k 次方和公式的一次项系数"，请观察：

$$
\begin{aligned}
1^0+2^0+3^0+\cdots+n^0 &= 1n\\
1^1+2^1+3^1+\cdots+n^1 &= \frac{1}{2}n^2+\frac{1}{2}n\\
1^2+2^2+3^2+\cdots+n^2 &= \frac{1}{3}n^3+\frac{1}{2}n^2+\frac{1}{6}n\\
1^3+2^3+3^3+\cdots+n^3 &= \frac{1}{4}n^4+\frac{1}{2}n^3+\frac{1}{4}n^2+0n\\
1^4+2^4+3^4+\cdots+n^4 &= \frac{1}{5}n^5+\frac{1}{2}n^4+\frac{1}{3}n^3+0n^2-\frac{1}{30}n\\
1^5+2^5+3^5+\cdots+n^5 &= \frac{1}{6}n^6+\frac{1}{2}n^5+\frac{5}{12}n^4+0n^3-\frac{1}{12}n^2+0n
\end{aligned}
$$

图中标红的就是这个数列的前几项：$B_0=1, B_1=\frac{1}{2}, B_2=\frac{1}{6}, B_3=0, B_4=-\frac{1}{30}, B_5=0$. 上面这些标红的数对于求出其他的数字很有帮助，从上往下看：

第一列的系数 $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \cdots$ 为调和级数；

第二列的系数 $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\cdots$ 全都是 $\frac{1}{2}$；

第三列的系数 $\frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{5}{12}, \cdots$ 为 $\frac{1}{12}$ 乘以幂的次数；

第四列的系数全为 0.

但是伯努利发现一个更神奇的地方，当引入杨辉三角以后，这些系数都可以用伯努利数表示：

$$
\begin{aligned}
1^0+2^0+3^0+\cdots+n^0 &= \frac{1}{1}(1B_0n)\\
1^1+2^1+3^1+\cdots+n^1 &= \frac{1}{2}(1B_0n^2+2B_1n)\\
1^2+2^2+3^2+\cdots+n^2 &= \frac{1}{3}(1B_0n^3+3B_1n^2+3B_2n)\\
1^3+2^3+3^3+\cdots+n^3 &= \frac{1}{4}(1B_0n^4+4B_1n^3+6B_2n^2+4B_3n)\\
1^4+2^4+3^4+\cdots+n^4 &= \frac{1}{5}(1B_0n^5+5B_1n^4+10B_2n^3+10B_3n^2+5B_4n)\\
1^5+2^5+3^5+\cdots+n^5 &= \frac{1}{6}(1B_0n^6+6B_1n^5+15B_2n^4+20B_3n^3+15B_4n^2+6B_5n)
\end{aligned}
$$

于是我们可以猜想

$$
1^p+2^p+3^p+\cdots+n^p=\frac{1}{p+1}\sum_{k=0}^{p}C_{p+1}^{n}B_kn^{p+1-k}
$$

这个关系式确实成立，后面将会证明.

需要说明这里得到的伯努利数第二类伯努利数，一般记作 $\{B_n^+\}_{n=0}^{\infty}$，而第一类伯努利数 $\{B_n^-\}_{n=0}^{\infty}$ 由自然数幂的前 $n-1$ 项和得到，两者唯一的不同就是 $B_1^+=\frac{1}{2}, B_1^-=-\frac{1}{2}$，接下来若没有特别标明，则 $\{B_n\}$ 指的是 $\{B_n^-\}$.

1.2 使用生成函数导出

回顾复变函数解析的定义：如果函数 $f(z)$ 在开区域 D 内的任意一点均能展成幂级数，则称 $f(z)$ 在 D 内解析. 而不同的幂级数的区别仅在于 z^n 的系数 $\{a_n\}$ 以及 n 的范围，这就意味着数列 $\{a_n\}$ 可以唯一地确定一个解析函数，换句话说，它包含了函数 $f(z)$ 的所有信息. 如果 $f(z)$ 被一个形式幂级数（具有幂结构的级数）所确定，数列 $\{a_n\}$ 的每一项决定该级数的系数，则称 $f(z)$ 是数列 $\{a_n\}$ 的生成函数 (generating function). 常用的生成函数有好几种：

(1) 普通型：$\mathrm{G}(a_n;x)=\sum\limits_{n=0}^{\infty}a_nx^n$

(2) 指数型：$\mathrm{EG}(a_n;x)=\sum\limits_{n=0}^{\infty}\dfrac{a_n}{n!}x^n$

(3) 泊松型：$\mathrm{PG}(a_n;x)=\sum\limits_{n=0}^{\infty}a_ne^{-x}\dfrac{1}{x^n}$

(4) 贝尔型：$\mathrm{BG_p}(a_n;x)=\sum\limits_{n=0}^{\infty}a_{p^n}x^n$

(5) 朗博型：$\mathrm{LG}(a_n;x)=\sum\limits_{n=1}^{\infty}a_n\dfrac{x^n}{1-x^n}$

有了生成函数，就可以给出伯努利数和伯努利多项式比较正式的定义了.

定义　伯努利数和伯努利多项式

伯努利多项式由指数型生成函数隐式地定义：

$$\frac{xe^{tx}}{e^x-1}:=\sum_{n=0}^{\infty}\frac{B_n^-(t)}{n!}x^n$$

第一类和第二类伯努利数分别定义为

$$B_n^-:=B_n(0)\qquad B_n^+:=B_n(1)$$

当省略正负号时，默认为第一类伯努利数.

伯努利数大概是这个风格，奇数项除了第一项以外都是 0，偶数项除了第零项以外是越来越复杂的有正有负的分数:

B_0	$B_1^{\pm}$	B_2	B_4	B_6	B_8	B_{10}	B_{12}	B_{14}	B_{16}	B_{18}	B_{20}	B_{22}	B_{24}
1	$\pm\frac{1}{2}$	$\frac{1}{6}$	$-\frac{1}{30}$	$\frac{1}{42}$	$-\frac{1}{30}$	$\frac{5}{66}$	$-\frac{691}{2730}$	$\frac{7}{6}$	$-\frac{3617}{510}$	$\frac{43867}{798}$	$-\frac{174611}{330}$	$\frac{854513}{138}$	$-\frac{236364091}{2730}$

伯努利多项式列举如下

$$
\begin{aligned}
&B_0(x)=1\\
&B_1(x)=x-\frac{1}{2}\\
&B_2(x)=x^2-x+\frac{1}{6}\\
&B_3(x)=x^3-\frac{3}{2}x^2+\frac{1}{2}x\\
&B_4(x)=x^4-2x^3+x^2-\frac{1}{30}\\
&B_5(x)=x^5-\frac{5}{2}x^4+\frac{5}{3}x^3-\frac{1}{6}x\\
&B_6(x)=x^6-3x^5+\frac{5}{2}x^4-\frac{1}{2}x^2+\frac{1}{42}
\end{aligned}
$$

1.3 性质

伯努利数和伯努利多项式有很多好玩的性质，其中伯努利数的性质很多都可以从伯努利多项式的性质导出.

(1) 两类伯努利数都有各自的生成函数.

$$\frac{\pm x}{e^{\pm x}-1}:=\sum_{n=0}^{\infty}\frac{B_n^{\mp}}{n!}x^n$$

(2) 从分析的角度，伯努利数有两种显化的方式

$$B_n=\lim_{x\to 0}\frac{\mathrm{d}}{\mathrm{d}x}\frac{x}{e^x-1}=\frac{n!}{2\pi i}\oint_C\frac{z}{e^z-1}\frac{\mathrm{d}z}{z^{n+1}}$$

其中 $C:|z|=r<2\pi$.

证明思路：对生成函数使用泰勒定理得到第一种显化方式，再使用留数定理得到第二种显化方式，需要 $r<2\pi$ 是为了避开生成函数的极点 $z=2\pi i$.

(3) 伯努利数的递归公式为

$$B_0=1\qquad\sum_{n=0}^{k-1}\frac{B_n}{n!(k-n)!}=0$$

伯努利多项式的递归公式为

$$B_0(x)=1\qquad B_n(x)=\sum_{k=0}^{n}C_n^kB_kx^{n-k}$$

证明：先证明伯努利数的递归公式. 根据伯努利数的生成函数，可以巧妙地使用 1：

$$1=\frac{x}{e^x-1}\cdot\frac{e^x-1}{x}=\sum_{k=0}^{\infty}\frac{x^k}{(k+1)!}\cdot\sum_{n=0}^{\infty}\frac{B_n}{n!}$$

然后使用柯西乘积[14]得到

$$\text{原式}=\sum_{k=0}^{\infty}\sum_{n=0}^{k}\frac{B_nx^n}{n!}\cdot\frac{x^{k-n}}{(k-n+1)!}=\sum_{k=0}^{\infty}x^k\sum_{n=0}^{k}\frac{B_n}{n!(k-n+1)!}=\sum_{k=0}^{\infty}\left(\sum_{n=0}^{k-1}\frac{B_n}{n!(k-n)!}\right)x^k$$

[14]柯西乘积：$\left(\sum\limits_{n=0}^{\infty}a_n\right)\left(\sum\limits_{n=0}^{\infty}b_n\right)=\sum\limits_{k=0}^{\infty}\sum\limits_{n=0}^{k}a_nb_{k-n}$.

比较两边 x^k 项的系数，可以推知

$$\sum_{n=0}^{k-1} \frac{B_n}{n!(k-n)!}\bigg|_{k=0} = 1, \quad \sum_{n=0}^{k-1} \frac{B_n}{n!(k-n)!}\bigg|_{k>0} = 0$$

接着证明多项式的递归公式. 方法同上，利用生成函数

$$\frac{xe^{tx}}{e^x-1} = \frac{x}{e^x-1} \cdot e^{tx} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n \cdot \sum_{n=0}^{\infty} \frac{t^n}{n!} x^n$$

再使用柯西乘积

$$\text{原式} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{B_k x^k}{k!} \frac{t^{n-k} x^{n-k}}{(n-k)!} = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} \frac{B_k t^{n-k}}{k!} (n-k)! \right) x^n$$

但是根据伯努利多项式的定义，比较 x^n 项的系数

$$\frac{B_n(t)}{n!} = \sum_{k=0}^{n} \frac{B_k t^{n-k}}{k!} (n-k)!$$

整理就得到了

$$B_0(x) = 1, \qquad B_n(x) = \sum_{k=0}^{n} C_n^k B_k x^{n-k}$$

伯努利数的定义很麻烦，它的计算方法也很对得起这份麻烦. 与我们之前遇到的数列递推公式不同，第 n 个伯努利数需要用前 $n-1$ 的伯努利数来导出，所以即使强如欧拉也只算出了前 30 个.

(4) 伯努利多项式的微积分关系

$$\frac{\mathrm{d}^m}{\mathrm{d}x^n} B_n(x) = A_n^m B_{n-m}(x) \qquad \int_a^b B_n(x) \mathrm{d}x = \frac{1}{n+1} \left[B_{n+1}(b) - B_{n+1}(a) \right]$$

(5) 伯努利多项式关于 x 的递推公式为

$$B_0(x+1) = B_0(x), \qquad B_n(x+1) = B_n(x) + nx^{n-1} \ (n \geq 1)$$

证明：根据伯努利多项式的生成函数

$$\frac{xe^{(t+1)x}}{e^x-1} := \sum_{n=0}^{\infty} \frac{B_n(t+1)}{n!} x^n$$

但是左侧可以改写为

$$\begin{aligned} \frac{xe^{(t+1)x}}{e^x-1} &= \frac{xe^{tx}e^x - xe^{tx} + xe^{tx}}{e^x-1} = xe^{tx} + \frac{xe^{tx}}{e^x-1} \\ &= \sum_{n=0}^{\infty} \frac{B_n(t)}{n!} x^n + \sum_{n=0}^{\infty} \frac{t^n}{n!} x^{n+1} = B_0(t) + \sum_{n=1}^{\infty} \frac{B_n(t) + nt^{n-1}}{n!} x^n \end{aligned}$$

对比两边 x^n 的系数，得到

$$B_0(x+1) = B_0(x), \qquad B_n(x+1) = B_n(x) + nx^{n-1} \ (n \geq 1)$$

(6) 伯努利多项式的加法公式为

$$B_n(a+b) = \sum_{k=0}^{n} C_n^k B_k(a) b^{n-k}$$

证明：利用伯努利数的生成函数

$$\frac{xe^{(a+b)x}}{e^x-1}:=\sum_{n=0}^{\infty}\frac{B_n(a+b)}{n!}x^n$$

但同时左侧可以改写为

$$\begin{aligned}\frac{xe^{(a+b)x}}{e^x-1}&=\frac{xe^{ax}}{e^x-1}\cdot e^{bx}=\sum_{n=0}^{\infty}\frac{B_n(a)}{n!}x^n\cdot\sum_{n=0}^{\infty}\frac{b^n}{n!}x^n\\&=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{B_k(a)}{k!}x^k\frac{b^{n-k}}{(n-k)!}x^{n-k}=\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n}\frac{B_k(a)}{k!}\frac{b^{n-k}}{(n-k)!}\right)x^n\end{aligned}$$

对比两侧 x^n 的系数，有

$$\frac{B_n(a+b)}{n!}=\sum_{k=0}^{n}\frac{B_k(a)}{k!}\frac{b^{n-k}}{(n-k)!}$$

整理就是

$$B_n(a+b)=\sum_{k=0}^{n}C_n^kB_k(a)b^{n-k}$$

尽管看起来复杂，伯努利数也有显式表达式

命题 伯努利数的显式表达式 (Louis Saalschütz, 1893)

$$B_n^+=\sum_{m=0}^{n}\sum_{k=0}^{m}(-1)^kC_m^k\frac{k^n}{m+1}\qquad B_n^-=\sum_{m=0}^{n}\sum_{k=0}^{m}(-1)^kC_m^k\frac{(k+1)^n}{m+1}$$

还可以利用解析延拓的 zeta 函数来计算伯努利数：

$$B_n^+=n\zeta(1-n)\quad\text{或}\quad B_{2n}=(-1)^{n+1}\frac{2(2n)!}{(2\pi)^{2n}}\zeta(2n),\quad n\geq 1$$

伯努利数有正有负，但其绝对值的总体趋势是增加的，当 n 特别大时，可以用上面的公式以及斯特林公式近似估计伯努利数的绝对值：

$$|B_{2n}|\approx 4\sqrt{\pi n}\left(\frac{n}{\pi e}\right)^{2n}$$

可见这个数列的偶数项的绝对值呈指数式地增长。

1.4 自然数幂和

自然数前 n 项的 $2\sim 5$ 次方和详见上面，都是高次多项式的形式，而且可以看出，自然数前 n 项的 k 次方和公式最高次项的次数为 $k+1$. 先设目标函数 $S(p,n)=\sum_{k=1}^{n}k^p$，接下来要求的就是目标函数的表达式。首先构造它的指数型生成函数，然后把目标函数定义式代入，再交换求和顺序，然后利用 e 的定义和等比数列求和公式：

$$F(x,n):=\sum_{p=0}^{\infty}\frac{S(p,n)}{p!}x^p=\sum_{p=0}^{\infty}\sum_{k=0}^{n}\frac{(kx)^p}{p!}=\sum_{k=1}^{n}\sum_{p=0}^{\infty}\frac{(kx)^p}{p!}=\sum_{k=1}^{n}e^{kx}=\frac{1-e^{-nx}}{e^{-x}-1}=\frac{e^{nx}-1}{x}\frac{-x}{e^{-x}-1}$$

前者很容易展成幂级数，后者是伯努利数的生成函数的变形。到这一步，已经可以比较快捷地求出目标函数的表达式了，既然 $F(x,n)=\dfrac{1-e^{nx}}{e^{-x}-1}$ 是目标函数的指数型生成函数，那么根据麦克劳林公式不断求导即可，例如

$$\left.\frac{\partial F}{\partial x}\right|_{x\to 0+}=\frac{n(n+1)}{2},\qquad\left.\frac{\partial^2 F}{\partial x^2}\right|_{x\to 0+}=\frac{n(n+1)(2n+1)}{6}$$

但我还是想求出 $S(p,n)$ 的显式表达式，于是把它们展成级数，再利用级数的柯西乘积.

$$\begin{aligned}F(x,n)&=\left(\sum_{k=1}^{\infty}\frac{n^k}{k!}x^{k-1}\right)\left(\sum_{k=0}^{\infty}\frac{B_k^+}{k!}x^k\right)=\left(\sum_{k=0}^{\infty}\frac{n^{k+1}}{(k+1)!}x^k\right)\left(\sum_{k=0}^{\infty}\frac{B_k^+}{k!}x^k\right)\\&=\sum_{k=0}^{\infty}\left(\sum_{m=0}^{k}\frac{B_m^+}{m!}x^m\frac{n^{k-m+1}}{(k-m+1)!}x^{k-m}\right)=\sum_{k=0}^{\infty}\left(\frac{1}{k+1}\sum_{m=0}^{k}C_{k+1}^{m}B_m^+n^{k-m+1}\right)\frac{x^k}{k!}\end{aligned}$$

最后就得到了著名的自然数幂和公式

定理　等幂求和公式 (Faulhaber's formula)

对于自然数 k，有

$$\begin{aligned}\sum_{m=1}^{n}m^k&=1^k+2^k+3^k+\cdots+n^k\\&=\lim_{x\to0}\frac{\mathrm{d}^n}{\mathrm{d}x^n}\frac{1-e^{nx}}{e^{-x}-1}\\&=\frac{1}{k+1}\sum_{m=0}^{k}C_{k+1}^{m}B_m^+n^{k-m+1}\end{aligned}$$

感觉比上面的求导更加难算，怎么办？还好有伯努利多项式. 根据递推公式

$$\begin{aligned}&B_n(x+1)=B_n(x)+nx^{n-1}\\\Longrightarrow&m^k=\frac{1}{k+1}\left[B_{k+1}(m+1)-B_{k+1}(m)\right]\\\Longrightarrow&\sum_{m=1}^{n}m^k=\frac{1}{m+1}\left[B_{k+1}(n+1)-B_{k+1}\right]\end{aligned}$$

1.5 更多函数的幂级数展开

某些函数的麦克劳林展开式/洛朗展开式系数看起来毫无规律

$$\begin{aligned}\tan(x)&=x+\frac{1}{3}x^3+\frac{2}{15}x^5+\frac{17}{315}x^7+\frac{62}{2835}x^9+\frac{1382}{155925}x^{11}+\cdots\\\cot(x)&=\frac{1}{x}-\frac{1}{3}x-\frac{1}{45}x^3-\frac{2}{945}x^5-\frac{1}{4725}x^7-\frac{2}{93555}x^9-\cdots\\\csc(x)&=\frac{1}{x}+\frac{1}{6}x+\frac{7}{360}x^3+\frac{31}{15120}x^5+\frac{127}{604800}x^7+\frac{73}{3421773}x^9+\cdots\end{aligned}$$

当然如果引入了复数，展开式就会变得非常简单（但那就不叫麦克劳林展开式了）.

例如 $\tan(z)$ 在复数域上，借助欧拉公式得到（此为 mathematica 的结果）

$$\tan(z)=-i\frac{e^{iz}-e^{-iz}}{e^{iz}+e^{-iz}}=i+2i\frac{e^{iz}}{e^{iz}+e^{-iz}}=i+2i\frac{1}{1+e^{-2iz}}=i+2i\sum_{n=1}^{\infty}(-1)^ne^{-2inz}$$

在实数域上，它们可以借助伯努利数表达出来。

首先 $\cos(x)=\dfrac{e^{ix}+e^{-ix}}{2}$，　$\sin(x)=\dfrac{e^{ix}-e^{-ix}}{2i}$，然后

$$\cot(x)=\frac{\cos(x)}{\sin(x)}=i\frac{e^{ix}+e^{-ix}}{e^{ix}-e^{-ix}}=i+\frac{2i}{e^{i2x}-1}=i+2i\sum_{k=0}^{\infty}\frac{B_k}{k!}(i2x)^{k-1}$$

再注意到除了 $B_1=-\frac{1}{2}$ 以外，伯努利数列的奇数项都是 0，另外 $i^{2k}=(-1)^k$，于是

$$\text{原式}=i+2i\left(\sum_{k=0}^{\infty}\frac{B_{2k}(i2x)^{2k-1}}{(2k)!}+\frac{B_1}{1!}(i2x)^0\right)=\frac{1}{x}\sum_{k=0}^{\infty}\frac{(-1)^kB_{2k}(2x)^{2k}}{(2k)!},\quad |x|<\pi$$

同理，利用三角恒等式 $\tan(x)=\cot(x)-2\cot(2x)$ 和 $\csc(x)=\cot\left(\frac{x}{2}\right)-\cot(x)$，即可得到

$$\tan(x)=\frac{1}{x}\sum_{k=0}^{\infty}\frac{(-1)^k(1-2^{2k})B_{2k}}{(2k)!}(2x)^{2k},\quad |x|<\frac{\pi}{2}$$

$$\csc(x)=\frac{1}{x}\sum_{k=0}^{\infty}\frac{(-1)^k(2-2^{2k})B_{2k}}{(2k)!}x^{2k},\quad |x|<\pi$$

看起来伯努利数计算起来也太麻烦了，可以返璞归真，使用多项式除法. 例如 $\tan(x)$ 的展开式，即可用正余弦的展开式相除：

$$\tan(x)=\frac{\sin(x)}{\cos(x)}=\frac{x-\frac{1}{3!}x^3+\frac{1}{5!}x^5-\frac{1}{7!}x^7+o(x^7)}{1-\frac{1}{2!}x^2+\frac{1}{4!}x^4-\frac{1}{6!}x^6+o(x^6)}$$

详细过程省略，直接给出答案，设

$$\tan(x)=a_1x+a_2x^3+a_3x^5+a_4x^7+a_5x^9+o(x^9)$$

则

$$\begin{aligned}
&a_1=1\\
&a_2=-\frac{1}{3!}+\frac{1}{2!}a_1=\frac{1}{3}\\
&a_3=\frac{1}{5!}-\frac{1}{4!}a_1+\frac{1}{2!}a_2=\frac{2}{15}\\
&a_4=-\frac{1}{7!}+\frac{1}{6!}a_1-\frac{1}{4!}a_2+\frac{1}{2!}a_3=\frac{17}{315}\\
&a_5=\frac{1}{9!}-\frac{1}{8!}a_1+\frac{1}{6!}a_2-\frac{1}{4!}a_3+\frac{1}{2!}a_4=\frac{62}{2835}
\end{aligned}$$

1.6 欧拉-麦克劳林公式：用导数来求定积分

回顾定积分（黎曼积分）的定义，我们知道函数 $f(x)$ 在区间 $[x_1,x_2]$ 上的积分

$$I=\int_{x_1}^{x_2}f(x)\mathrm{d}x$$

如果存在的话，可以这样求出来：记 $L=x_2-x_1$ 为区间长，把区间 $[x_1,x_2]$ **平均**切成 n 份，得到 n 个长度为 L/n 的**闭**区间：

$$[x_1,x_1+\frac{1}{n}L],[x_1+\frac{1}{n}L,x_1+\frac{2}{n}L],\cdots,[x_1+\frac{n-1}{n}L,x_2]$$

然后在每个区间上各取一点 ξ_n，作和式来逼近积分值

$$S(L,n):=\sum_{k=1}^{n}f(\xi_n)\frac{L}{n}\approx\int_{x_1}^{x_2}f(x)\mathrm{d}x$$

当 n 趋于无穷时，每个区间的长度趋于 0，$S(L,n)$ 的极限即为积分值：

$$\int_{x_1}^{x_2}f(x)\mathrm{d}x=\lim_{n\to\infty}S(L,n)=\sum_{k=1}^{\infty}f(\xi_n)\frac{L}{n}$$

如果令积分的上下界为整数 a, b，然后让每个小区间长恰好为 1，即 $n = L = b - a$，再令 ξ_n 取在小区间的右端点上，即 $\xi_n = a + n$

$$S(L, L) = f(a+1) + f(a+2) + f(a+3) + \cdots + f(b) \approx I$$

得到一个非常粗糙的积分近似值，那么 $S(L, L)$ 和 I 之间究竟差了什么？接下来的工作就是找到它.

在此之前，先证明一个引理.

引理 二重求和交换次序

$$\sum_{i=0}^{N}\sum_{j=0}^{i} a_{ij} = \sum_{j=0}^{N}\sum_{i=j}^{N} a_{ij}$$

证明；将求和式写成阶梯形

$$\begin{array}{ccccccccccc}
S & = & a_{00} & & & & & & & & \\
 & + & a_{10} & + & a_{11} & & & & & & \\
 & + & a_{20} & + & a_{21} & + & a_{22} & & & & \\
 & & \vdots & & \vdots & & \vdots & & & & \\
 & + & a_{N0} & + & a_{N1} & + & a_{N2} & + & \cdots & + & a_{NN}
\end{array}$$

按行相加即为 $\sum_{i=0}^{N}\sum_{j=0}^{i} a_{ij}$，按列相加即为 $\sum_{j=0}^{N}\sum_{i=j}^{N} a_{ij}$.

接下来正式开始，假设函数 $f(x)$ 在区间 $[0, n]$ 内解析，即可以作麦克劳林展开 $f(x) = \sum_{r=0}^{\infty} a_r x^r$，则

$$\sum_{k=1}^{n} f(k) = \sum_{k=1}^{n}\sum_{r=0}^{\infty} a_r k^r = \sum_{r=0}^{\infty} a_r \sum_{k=0}^{n} k^r = {\color{red}\sum_{r=0}^{\infty}} \frac{a_r}{r+1} {\color{red}\sum_{m=0}^{r}} C_{r+1}^{m} B_m^{+} n^{r-m+1}$$

注意组合恒等式

$$\frac{{\color{blue}C_{r+1}^{m}}}{{\color{blue}r+1}} = \frac{(r+1)!}{m!(r+1-m)!(r+1)} = \frac{r!}{m!(r+1-m)!} = \frac{{\color{blue}A_r^{m-1}}}{{\color{blue}m!}}$$

然后借助引理给红色部分的求和号交换次序，得到

$$\sum_{k=1}^{n} f(k) = {\color{red}\sum_{m=0}^{\infty}} B_m^{+} {\color{red}\sum_{r=m}^{\infty}} \frac{a_r}{{\color{blue}r+1}} {\color{blue}C_{r+1}^{m}} n^{r-m+1} = \sum_{m=0}^{\infty} \frac{B_m^{+}}{{\color{blue}m!}} \sum_{r=m}^{\infty} {\color{blue}A_r^{m-1}} a_r n^{r-m+1}$$

将 $m = 0$ 的情况单独拿出来

$$\sum_{k=0}^{n} f(k) = B_0 \sum_{r=0}^{\infty} \frac{a_r n^{r+1}}{r+1} + \sum_{\color{red}m=1}^{\infty} \frac{B_{\color{red}m}^{+}}{m!} \sum_{\color{red}r=m}^{\infty} A_r^{\color{red}m-1} = B_0 \sum_{r=0}^{\infty} \frac{a_r n^{r+1}}{r+1} + \sum_{\color{red}m=0}^{\infty} \frac{B_{\color{red}m+1}^{+}}{(m+1)!} \sum_{\color{red}r=m+1}^{\infty} A_r^{\color{red}m} a^r n^{r-m}$$

注意到两个一般人注意不到的地方：

$$\frac{n^{r+1}}{r+1} = \int_0^n x^r \mathrm{d}x, \qquad A_r^m n^{r-m} = \frac{\mathrm{d}^m}{\mathrm{d}x^m} x^r \bigg|_{x=n}$$

以及一个谁都能注意到的地方：$B_0=1$，代入上式，然后交换微积分的次序

$$
\begin{aligned}
\sum_{k=0}^{n} f(k) &= \sum_{r=0}^{\infty} a_r \int_0^n x^r \mathrm{d}x + \sum_{m=0}^{\infty} \frac{B_{m+1}^{+}}{(m+1)!} \sum_{r=m+1}^{\infty} a_r \left.\frac{\mathrm{d}^m}{\mathrm{d}x^m} x^r\right|_{x=n} \\
&= \int_0^n \left[\sum_{r=0}^{\infty} a_r x^r\right] \mathrm{d}x + \sum_{m=0}^{\infty} \frac{B_{m+1}^{+}}{(m+1)!} \frac{\mathrm{d}^m}{\mathrm{d}x^m} \left[\sum_{r=m+1}^{\infty} a_r x^r\right]_{x=n} \\
&= \int_0^n f(x) \mathrm{d}x + \sum_{m=0}^{\infty} \frac{B_{m+1}^{+}}{(m+1)!} \frac{\mathrm{d}^m}{\mathrm{d}x^m} \left[f(x) - \sum_{r=0}^{m} a_r x^r\right]_{x=n}
\end{aligned}
$$

这里又有一个一般人注意不到的地方，多项式求了 m 次导数之后，常数项为原来的 x^m 项的系数乘上 $m!$，这时再代入 $x=0$，就只剩下了常数项，体现在原式中，即

$$
\frac{\mathrm{d}^m}{\mathrm{d}x^m} \sum_{r=0}^{m} a_r x^r = \frac{\mathrm{d}^m}{\mathrm{d}x^m} a_m x^m = m! a_m = \left.\frac{\mathrm{d}^m}{\mathrm{d}x^m} f(x)\right|_{x=0}
$$

因此

$$
\sum_{k=0}^{n} f(k) = \int_0^n f(x) \mathrm{d}x + \sum_{m=0}^{\infty} \frac{B_{m+1}^{+}}{(m+1)!} \left(\left.\frac{\mathrm{d}^m}{\mathrm{d}x^m} f(x)\right|_{x=n} - \left.\frac{\mathrm{d}^m}{\mathrm{d}x^m} f(x)\right|_{x=0} \right)
$$

整理一下，将 $m=0$ 的特殊情况拿出来，即

$$
\sum_{k=0}^{n} f(k) = \int_0^n f(x) \mathrm{d}x + \frac{f(n)-f(0)}{2} + \sum_{m=0}^{\infty} \frac{B_{2m}^{+}}{(2m)!} \left[f^{(2m-1)}(n) - f^{(2m-1)}(0)\right]
$$

这就解答了最初的问题，$S(L,L)$ 和 I 之间相差的东西就是

$$
S(L,L) - I = \frac{f(n)-f(0)}{2} + \sum_{m=0}^{\infty} \frac{B_{2m}^{+}}{(2m)!} \left[f^{(2m-1)}(n) - f^{(2m-1)}(0)\right]
$$

如果将 $[0,n]$ 推广至一般的区间

定理　欧拉-麦克劳林公式 (Euler-Maclaurin formula)

对于任意在 R_+ 内解析的函数 $f(x)$，有

$$
\sum_{k=0}^{n} f(k) = \int_0^n f(x) \mathrm{d}x + \frac{f(n)+f(0)}{2} + \sum_{m=1}^{\infty} \frac{B_{2m}^{+}}{(2m)!} \left[f^{(2m-1)}(n) - f^{(2m-1)}(0)\right]
$$

1.7 拉马努金和

我们知道无穷级数的定义

$$
\sum_{n=1}^{\infty} a_n := \lim_{N\to\infty} \sum_{n=1}^{N} a_n = \lim_{n\to\infty} S_n
$$

这是一种非常自然的定义，称为柯西求和(Cauchy summation)，若这个极限存在，则称该级数在柯西求和意义下收敛。有些级数在柯西求和意义下发散，但把级数的定义改一改，却可以得到收敛的结果。除了柯西求和外，级数求和还有以下几种定义：

(1) 切萨罗求和 (Cesaro summation)：

$$
\sum_{n=0}^{\infty} a_n := \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} S_k
$$

可以看出切萨罗求和就是把原数列前 n 项取算术平均之后当做 b_n，然后用新的数列 $\{b_n\}$ 代替 $\{a_n\}$ 进行柯西求和. 所以该求和定义有时称为切萨罗平均(*Cesaro mean*)，用切萨罗求和代替柯西求和可以防止傅里叶级数中的吉布斯现象.

(2) 广义切萨罗求和/(C,m) 求和:

$$(C,m)-\sum_{n=0}^{\infty}a_n=\lim_{n\to\infty}\sum_{k=0}^{n}\frac{C_n^k}{C_{n+m}^k}a_k$$

可见 $(C,0)$ 求和就是柯西求和，$(C,1)$ 求和就是切萨罗求和．只要 (C,m) 求和收敛，则对于任意 $n>m$，(C,n) 求和也收敛．因此只要广义切萨罗求和收敛，则切萨罗和收敛，接着柯西求和也会收敛．

(3) 侯德尔求和 (Hölder summation)：令

$$S_n^1=\frac{1}{n}\sum_{k=0}^{n}a_k,\qquad S_n^{m+1}=\frac{1}{n}\sum_{k=0}^{n}S_k^m$$

则级数的侯德尔求和为

$$\sum_{n=0}^{\infty}a_n=\lim_{\substack{n\to\infty\\ m\to\infty}}S_n^m$$

侯德尔求和则是切萨罗求和的另一种推广，b_n 为 $\{a_n\}$ 的前 n 项和的算术平均，然后用 $\{b_n\}$ 代替 $\{a_n\}$ 求和．如果还不收敛，则再令 c_n 为 $\{b_n\}$ 的前 n 项和的算术平均，让 $\{c_n\}$ 代替 $\{b_n\}$ 求和．将上述过程重复至收敛为止，所得的结果即为 $\{a_n\}$ 的侯德尔求和．只要级数的侯德尔求和收敛，切萨罗求和就会收敛，接着柯西求和也会收敛．

(4) 阿贝尔和 (Abel summation)

$$\sum_{n=0}^{\infty}a_n:=\lim_{x\to1^-}\lim_{n\to\infty}\sum_{k=0}^{n}a_nx^n$$

阿贝尔求和比 (C,m) 求和与侯德尔求和更强，级数只要在 (C,m) 求或与侯德尔求和意义下收敛，就会在阿贝尔求和下收敛，但反之不成立，详见例子．

这几种不同意义的求和可以处理一些发散级数：

- 格兰迪级数 (Grandi series)：$\{a_n\}=\{1,-1,1,-1,\cdots\}=\{(-1)^n\}$ 在切萨罗求和意义下，

$$\{S_n\}=\{1,0,1,0,\cdots\}=\frac{(-1)^n}{2}$$

$$\{\frac{1}{n}\sum_{k=1}^{n}S_n\}=\{1,\frac{1}{2},\frac{2}{3},\frac{2}{4},\frac{3}{5},\frac{3}{6}\}=\{\frac{1}{n}\lceil\frac{n}{2}\rceil\}\le\{\frac{1}{n}\frac{n+1}{2}\}=\frac{1}{2}\{\frac{n+1}{n}\}$$

因此该数列的切萨罗求和为

$$(C,1)-\sum_{n=1}^{\infty}a_n=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}S_n=\frac{1}{2}$$

而阿贝尔求和为

$$\sum_{n=1}^{\infty}a_n=\lim_{x\to1^-}\sum_{n=0}^{\infty}(-1)^nx^n=\lim_{x\to1^-}\frac{1}{1+x}=\frac{1}{2}$$

- $\{a_n\}=\{1,0,1,0,\cdots\}=\{\frac{1+(-1)^n}{2}\}$

- 令 $\exp\frac{1}{1-x}=\sum_{n=0}^{\infty}a_nx^n$，则 $\{a_n\}$ 在阿贝尔求和意义下收敛，而在切萨罗求和意义下不收敛．

现在可以根据欧拉-麦克劳林公式引入拉马努金和.
令欧拉-麦克劳林公式左端的求和范围改为 1 到 n，与此同时右侧的积分上下界也变成了 1 到 n，将其分为 1 到 0 和 0 到 n 两段

$$\sum_{k=1}^{n} f(k)=\int_{1}^{0} f(x)\mathrm{d}x+\int_{0}^{n} f(x)\mathrm{d}x+\frac{f(n)+f(1)}{2}+\sum_{m=1}^{\infty}\frac{B_{2m}^{+}}{(2m)!}\left[f^{(2m-1)}(n)-f^{(2m-1)}(1)\right]$$

将所有含有 n 的部分移到等式左边，得到

$$\sum_{k=1}^{n} f(k)-\int_{0}^{n} f(x)\mathrm{d}x-\frac{f(n)}{2}-\sum_{m=1}^{\infty}\frac{B_{2m}^{+}}{(2m)!}f^{(2m-1)}(n)=\int_{1}^{0} f(x)\mathrm{d}x+\frac{f(1)}{2}-\sum_{m=1}^{\infty}\frac{B_{2m}^{+}}{(2m)!}f^{(2m-1)}(1)$$

如果当 $n\to\infty$ 时左端极限不存在，等式右侧即为拉马努金求和.

定义 拉马努金求和

对于在区间 $[0,+\infty)$ 上可积的函数，定义

$$\sum_{n=1}^{[\mathcal{R}]} f(n)=\int_{1}^{0} f(x)\mathrm{d}x+\frac{f(1)}{2}-\sum_{m=1}^{\infty}\frac{B_{2m}^{+}}{(2m)!}f^{(2m-1)}(1)$$

为级数 $\sum_{n=1}^{\infty} f(n)$ 的拉马努金求和(Ramanujan summation).

不过既然是对发散级数的求和，结果就不那么符合直觉了，最著名的结果莫过于

命题 全体自然数的和为 $-1/12$

$$\sum_{n=1}^{[\mathcal{R}]} n=-\frac{1}{12}$$

证明：令 $f(x)=x$，

$$\sum_{n=1}^{[\mathcal{R}]} n=\sum_{n=1}^{[\mathcal{R}]} f(n)=\int_{1}^{0} f(x)\mathrm{d}x+\frac{f(1)}{2}-\frac{B_2}{2}=-\frac{1}{2}+\frac{1}{2}-\frac{1}{12}=-\frac{1}{12}$$

拉马努金求和的思想是把发散级数 $\sum_{n=1}^{\infty} a_n$ 的部分和 S_n 分成发散和收敛两部分，发散的部分作为 S_n 的渐进表达式（即之前的推导中被移到等式左边的与 n 有关的部分），收敛的部分就是所谓的拉马努金和，与 n 无关，例如我们知道调和级数

$$\sum_{n=0}^{n}\frac{1}{k}\approx\ln(n)+\gamma$$

是发散的级数，但是其部分和表达式可以近似写为 $\ln(n)$（与 n 有关）与欧拉-马歇罗尼常数 γ（与 n 无关）之和，所以调和级数的拉马努金和为 γ.

2 欧拉数

欧拉有很多的数学成果，和“欧拉数”这三个字有关的东西就有 Euler numbers, Euler's number, Eulerian number, Euler's constant, Eulerian integers 等，这里的欧拉数指的是第一个，和上面的伯努利数一样是一个数列 $\{E_n\}_{n=0}^{\infty}$.

欧拉数由生成函数导出：

$$\frac{1}{\cosh(x)}=\frac{2}{e^x+e^{-x}}=\sum_{k=0}^{\infty}\frac{E_k}{k!}x^k$$

同伯努利数一样，欧拉数有正有负，绝对值总体趋势是增加的，但都是整数，奇数项都为 0.

E_0	E_2	E_4	E_6	E_8	E_{10}	E_{12}	E_{14}	E_{16}	E_{18}
1	-1	5	-61	1385	-50521	2702765	-199360981	19391512145	-2404879675441

有了欧拉数，我们可以给更多的函数找到通用的展开式了。

$$\sec(x)=1+\frac{1}{2}x^2+\frac{5}{24}x^4+\frac{61}{720}x^6+\frac{277}{8064}x^8+\frac{50521}{3628800}x^{10}+\cdots$$

首先注意到 $\sec(x)=\dfrac{1}{\cos(x)}=\dfrac{2}{e^{ix}+e^{-ix}}$ 立即代入欧拉数的导出函数中，并略去奇数项，得到

$$\sec(x)=\sum_{k=0}^{\infty}\frac{E_k}{k!}(ix)^k=\sum_{k=0}^{\infty}\frac{(-1)^kE_{2k}}{(2k)!}x^{2k},\quad |x|<\frac{\pi}{2}$$

与伯努利数相似，偶数项欧拉数的绝对值呈指数式增长

$$|E_{2n}|\approx 8\sqrt{\frac{n}{\pi}}\left(\frac{4n}{\pi e}\right)^{2n}$$

增长快慢来说，若记 $A\prec B$ 表示 A 是 B 的高阶无穷大，那么对于正整数 n，当 $n\to+\infty$ 时，有 $\psi(2n)\prec 2n\prec e^{2n}\prec|B_{2n}|\prec|E_{2n}|\prec\Gamma(2n)\prec n^{2n}$

3 欧拉积分系列

3.1 伽马函数

Γ 函数也称为第二类欧拉积分，它是阶乘的扩展，阶乘定义在自然数上，Γ 函数则定义在复数集上，它有好几种不同的定义，但最常用的还是伯努利的积分定义：

$$\Gamma(z):=\int_0^{+\infty}t^{z-1}e^{-t}\mathrm{d}t,\quad \mathrm{Re}(z)>0$$

经过简单的换元，可以得到不同的结果.

(1) 令 $t=u^2$

$$\Gamma(z)=2\int_0^{\infty}e^{-u^2}u^{2z-1}\mathrm{d}u$$

(2) 令 $e^{-t}=u$

$$\Gamma(z)=2\int_0^{1}\left(\ln\frac{1}{u}\right)^{z-1}\mathrm{d}u$$

定义 Γ 函数的初衷是让阶乘在实数域上有定义，但是 $\Gamma(n+1)=n!$，为了和阶乘统一起来，又定义了 $\Pi(n):=\Gamma(n+1)=n!$，但是接下来不使用它.

Γ 函数有很多性质.

(1) 递推公式

$$\Gamma(z+1)=z\Gamma(z)$$

Γ 函数的递推公式就是根据阶乘得来的，不过假如不知道 Γ 和阶乘的关系，也可以利用分部积分法证明.

$$\Gamma(z+1)=\int_0^{+\infty}t^ze^{-t}\mathrm{d}t=-\int_0^{+\infty}t^z\mathrm{d}e^{-t}=-t^ze^{-t}\Big|_0^{+\infty}+\int_0^{+\infty}e^{-t}zt^{z-1}=z\Gamma(z)$$

又因为

$$\Gamma(1)=\int_0^{\infty}e^{-t}\mathrm{d}t=1$$

所以可以确立关系 $n!=\Gamma(n+1)$，这也从一方面说明了为什么要求 $\mathrm{Re}(z)>0$，试想 $\Gamma(0)=(-1)\times(-2)\times(-3)\times\cdots$，这是个无穷递降乘积，不但不收敛，还会以越来越大的幅度震荡，永远也求不出它的值.

(2) 欧拉的无穷乘积定义

$$\Gamma(z)=\frac{1}{z}\prod_{k=1}^{\infty}\frac{(1+\frac{1}{k})^z}{1+\frac{k}{z}},\quad z\neq 0,-1,-2,\cdots$$

欧拉提出无穷乘积定义是从这个极限开始的：$\lim\limits_{n\to\infty}\dfrac{n!(n+1)^m}{(n+m)!}=1$. 这个极限证明起来比较直观，把阶乘展开再化简就好了：

$$\lim_{n\to\infty}\frac{n!(n+1)^m}{(n+m)!}=\lim_{n\to\infty}\frac{(n+1)(n+1)\cdots(n+1)}{(n+2)(n+3)\cdots(n+m)}=\lim_{n\to\infty}\frac{n+1}{n+2}\cdot\lim_{n\to\infty}\frac{n+1}{n+3}\cdots\lim_{n\to\infty}\frac{n+1}{n+m}=1$$

两边同时乘 $m!$，并展开，就可以得到：

$$m!=\lim_{n\to\infty}\frac{m!n!(n+1)^m}{(n+m)!}=\lim_{n\to\infty}\frac{n!(n+1)^m}{A_{m+n}^n}=\lim_{n\to\infty}\frac{\left(\frac{2}{1}\frac{3}{2}\cdots\frac{n+1}{n}\right)^m}{(1+\frac{m}{1})(1+\frac{m}{2})\cdots(1+\frac{m}{n})}$$

写成连乘积的形式：

$$m!=\lim_{n\to\infty}\frac{\prod_{k=1}^n\left(1+\frac{1}{k}\right)^m}{\prod_{k=1}^n(1+\frac{m}{k})}=\prod_{k=1}^{\infty}\frac{(1+\frac{1}{k})^m}{1+\frac{m}{k}}$$

然后把这个极限扩展到复数域上，并根据关系式 $z!=\Gamma(z+1)=z\Gamma(z)$，就得到了

$$\Gamma(z)=\frac{1}{z}\prod_{k=1}^{\infty}\frac{(1+\frac{1}{k})^z}{1+\frac{k}{z}}$$

在欧拉无穷乘积的基础上，还可以再改写式子：

$$\begin{aligned}\Gamma(z)&=\lim_{n\to\infty}\frac{1}{z}\prod_{k=1}^{n}\left(1+\frac{1}{n}\right)^z\left(1+\frac{z}{n}\right)^{-1}=\lim_{n\to\infty}\frac{1}{z}\frac{(\frac{2}{1})^z(\frac{3}{2})^z\cdots(\frac{n}{n-1})^z(\frac{n+1}{n})^z}{(\frac{1+z}{1})(\frac{2+z}{2})\cdots(\frac{n+z}{n})}\\&=\lim_{n\to\infty}\frac{n!n^z}{z(z+1)\cdots(z+n)}\left(\frac{n+1}{n}\right)^z=\lim_{n\to\infty}\frac{n!n^z}{z(z+1)\cdots(z+n)}\\&=\lim_{n\to\infty}\frac{n!n^z}{(z)_{n+1}}\end{aligned}$$

(3) 魏尔斯特拉斯的无穷乘积定义

$$\Gamma(z)=\frac{e^{-\gamma z}}{z}\prod_{n=0}^{\infty}\left(1+\frac{z}{n}\right)^{-1}e^{\frac{z}{n}},\quad z\neq 0,-1,-2,\cdots$$

接上面，改写的欧拉无穷乘积可以再次改写：

$$\begin{aligned}\Gamma(z)&=\lim_{n\to\infty}\frac{n!n^z}{z(z+1)\cdots(z+n)}=\lim_{n\to\infty}\frac{1}{z}\prod_{k=1}^{n}\left(\frac{k}{z+k}\right)\exp\left[z\ln(n)\right]\\&=\lim_{n\to\infty}\frac{1}{z}\prod_{k=1}^{n}\left(1+\frac{z}{k}\right)^{-1}\exp\left[z\left(\ln(n)-\sum_{k=1}^{n}\frac{1}{k}\right)\right]\prod_{k=1}^{n}e^{z/k}\\&=\frac{e^{-\gamma z}}{z}\prod_{k=1}^{\infty}\left(1+\frac{z}{k}\right)^{-1}e^{z/k}\end{aligned}$$

其中

$$\gamma=\lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k}-\ln(n)\right)=0.577215664\cdots$$

称为欧拉-马歇罗尼常数 (Euler-Mascheroni constant)，接下来有一节会专门提到它.

(4) 余元公式(reflection formula)

$$\Gamma(1+z)\Gamma(1-z)=\frac{\pi z}{\sin(\pi z)},\quad z\notin\mathbb{Z}$$

证明：利用魏尔斯特拉斯无穷乘积

$$\begin{aligned}\Gamma(z)\Gamma(-z)&=\frac{e^{\gamma z}}{z}\prod_{n=1}^{\infty}\left(1+\frac{z}{n}\right)^{-1}e^{z/n}\frac{e^{-\gamma z}}{-z}\prod_{n=1}^{\infty}\left(1+\frac{-z}{n}\right)^{-1}e^{-z/n}\\&=-\frac{1}{z^2}\prod_{n=1}^{\infty}\left[\left(1+\frac{z}{n}\right)\left(1+\frac{z}{n}\right)\right]^{-1}=-\frac{1}{z^2}\prod_{n=1}^{\infty}\left(1-\frac{z^2}{n^2}\right)^{-1}\end{aligned}$$

之前讨论零点分解定理时，得到了这个式子：

$$\frac{\sin(\pi z)}{\pi z}=\prod_{n=1}^{\infty}\left(1-\frac{z^2}{n^2}\right)$$

所以有

$$\Gamma(z)\Gamma(-z)=-\frac{\pi}{z\sin(\pi z)}$$

考虑到 $\Gamma(1+z)=z\Gamma(z),\Gamma(1-z)=-z\Gamma(-z)$，所以有

$$\Gamma(1+z)\Gamma(1-z)=\frac{\pi z}{\sin(\pi z)}$$

(5) 勒让德倍加公式 (duplication formula)

$$\Gamma(2z)=\frac{\Gamma(z)\Gamma(z+\frac{1}{2})}{2^{1-2z}\sqrt{\pi}}$$

证明需要用到接下来介绍的 B 函数.

$$B(a,a):=\int_0^1 x^{a-1}(1-x)^{a-1}\mathrm{d}x=\int_0^1(x-x^2)^{a-1}\mathrm{d}x=\int_0^1\left[\frac{1}{4}-\left(x-\frac{1}{2}\right)^2\right]^{a-1}\mathrm{d}x$$

由于被积函数关于 $x=\dfrac{1}{2}$ 对称，所以

$$B(a,a)=2\int_0^{\frac{1}{2}}\left[\frac{1}{4}-\left(x-\frac{1}{2}\right)^2\right]^{a-1}\mathrm{d}x$$

令 $\left(x-\dfrac{1}{2}\right)=\dfrac{1}{4}t$，则 $\mathrm{d}x=-\dfrac{1}{4\sqrt{t}}\mathrm{d}t$，积分上下限也由 $0\sim1/2$ 改成了 $1\sim0$，即

$$B(a,a)=2\int_0^1\left(\frac{1}{4}-\frac{1}{4}t\right)^{a-1}\frac{1}{4}t^{-1/2}\mathrm{d}t=\frac{1}{2^{2a-1}}B\left(\frac{1}{2},a\right)$$

根据关系 $B(a,b)=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ 以及 $\Gamma\left(\dfrac{1}{2}\right)=\sqrt{\pi}$，在上式两端代入，得到

$$\frac{\Gamma(a)\Gamma(a)}{\Gamma(2a)}=\frac{\sqrt{\pi}}{2^{2a-1}}\frac{\Gamma(a)}{\Gamma\left(a+\frac{1}{2}\right)}$$

整理一下即可.

(5) 由倍加公式可以推知，当自变量为某些复数时，有如下关系：

$$|\Gamma(n+bi)|^2=\frac{b\pi}{\sinh(b\pi)}\prod_{k=1}^{n-1}(k^2+b^2),\quad n\in\mathbb{N}_+$$
$$|\Gamma(-n+bi)|^2=\frac{\pi}{b\sinh(b\pi)}\prod_{k=1}^{n}(k^2+b^2)^{-1},\quad n\in\mathbb{N}$$
$$|\Gamma(\frac{1}{2}\pm n+bi)|^2=\frac{\pi}{\cosh(b\pi)}\prod_{k=1}^{n}\left((k-\frac{1}{2})^2+b^2\right)^{\pm 1},\quad n\in\mathbb{N}$$

代入特殊值可以得到 $|\Gamma(bi)|^2=\dfrac{\pi}{b\sinh(b\pi)}$，$|\Gamma(1+bi)|^2=\dfrac{b\pi}{\sinh(b\pi)}$，$|\Gamma(\frac{1}{2}+bi)|^2=\dfrac{\pi}{\cosh(b\pi)}$

根据余元公式，可以由 $\Gamma\left(\frac{1}{2}\right)=\dfrac{\pi}{\sin(\frac{1}{2}\pi)}=\pi$ 导出 $\Gamma(\frac{1}{2})=\sqrt{\pi}$，再由递推公式得到

$$\Gamma\left(n+\frac{1}{2}\right)=\sqrt{\pi}\cdot\frac{1}{2}\cdot\frac{3}{2}\cdot\frac{5}{2}\cdots\frac{2n-1}{2}=\frac{(2n-1)!!}{2^n}$$

3.2 多伽马函数

定义 多伽马函数

多伽马函数(polygamma function) 是一系列函数 $\{\psi^{(n)}\}_{n=0}^{\infty}$ 的统称，定义为

$$\psi^{(0)}(z):=\frac{\mathrm{d}}{\mathrm{d}z}\ln\Gamma(z)$$
$$\psi^{(n+1)}(z):=\frac{\mathrm{d}}{\mathrm{d}z}\psi^{(n)}(z)$$

其中 $\psi^{(0)}(z)$ 简写为 $\psi(z)$，称为双伽马函数(digamma function)，是这一系列中最有研究价值的函数.

digamma 经过亿点点化简，可以得到几种等价的表达式：

(1) 高斯的结果：

$$\psi(z)=\int_0^{\infty}\left(\frac{e^{-t}}{t}-\frac{e^{-zt}}{1-e^{-t}}\right)\mathrm{d}t$$

(2) 高斯结果的优化：

$$\psi(z+1)=-\gamma+\int_0^1\left(\frac{1-t^z}{1-t}\right)\mathrm{d}t$$

(3) 狄利克雷的结果：

$$\psi(z)=\int_0^{\infty}\left(e^{-t}-\frac{1}{(1+t)^z}\right)\frac{dt}{t}$$

(4) 级数表达式：

$$\psi(z+1)=-\gamma+\sum_{n=1}^{\infty}\left(\frac{1}{n}-\frac{1}{n+z}\right),\quad z\notin\mathbb{Z}_-$$

digamma 函数具有如下性质：

(1) 递推公式

$$\psi(z+1)=\psi(z)+\frac{1}{z}$$

(2) 反射性质

$$\psi(1-z)=\psi(z)+z\cot(\pi z)$$

(3) 倍加公式

$$\psi(2x)=\frac{1}{2}\psi(x)+\frac{1}{2}\psi\left(x+\frac{1}{2}\right)+\ln 2$$

由于 $\psi(1)=-\gamma$，根据性质 1 可以得知

$$\psi(n)=-\gamma+1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n-1}=-\gamma+\sum_{i=1}^{n-1}\frac{1}{i}$$

3.3 欧拉 beta 函数

Beta 函数又称为欧拉第一类积分，它看起来有点像二项分布的概率密度函数。

$$B(x,y):=\int_0^1 t^{x-1}(1-t)^{y-1}\mathrm{d}t,\quad \mathrm{Re}(x)>0,\mathrm{Re}(y)>0$$

Beta 函数的表达式也可以经过简单的换元得到下面这些结果.

(1) 令 $t=\cos^2\theta$

$$B(x,y)=2\int_0^{\frac{\pi}{2}}(\sin\theta)^{2x-1}(\cos\theta)^{2y-1}\mathrm{d}\theta$$

(2) 令 $t=u^2$

$$B(x,y)=2\int_0^1 u^{2x-1}(1-u)^{y-1}\mathrm{d}u$$

(3) 令 $t=\dfrac{u}{a}$

$$B(x,y)=\frac{1}{a^{x+y-1}}\int_0^a u^{x-1}(a-u)^{y-1}\mathrm{d}u$$

(4) 令 $t=\dfrac{1+u}{2}$

$$B(x,y)=\frac{1}{2^{x+y-1}}\int_{-1}^1(1+u)^{x-1}(1-u)^{y-1}\mathrm{d}u$$

(5) 令 $t=\dfrac{u}{1+u}$

$$B(x,y)=\int_0^\infty\frac{u^{x-1}}{(1+u)^{x+y}}\mathrm{d}u$$

Beta 函数和 Gamma 函数有非常紧密的关系：

$$B(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

对 Γ 函数动手，令 $t=x^2$，分别得到

$$\Gamma(p)=\int_0^\infty t^{p-1}e^{-t}\mathrm{d}t=\int_0^\infty x^{2(p-1)}e^{-x^2}(2x)\mathrm{d}x=2\int_0^\infty x^{2p-1}e^{-x^2}\mathrm{d}x$$

因此

$$\Gamma(p)\Gamma(q)=4\int_0^\infty x^{2p-1}e^{-x^2}\mathrm{d}x\int_0^\infty y^{2q-1}e^{-y^2}\mathrm{d}y=4\iint_{\mathbb{R}_+^2}x^{2p-1}y^{2q-1}e^{-(x^2+y^2)}\mathrm{d}x\mathrm{d}y$$

用极坐标换元，再拆成两个积分

原式 $= 4\int_0^{\frac{\pi}{2}} \mathrm{d}\theta \int_0^{\infty} e^{-\rho^2}\rho^{2p+2q-2}\cos^{2p-1}\theta\sin^{2q-1}\theta\cdot\rho\mathrm{d}\rho = 4\int_0^{\frac{\pi}{2}}\cos^{2p-1}\theta\sin^{2q-1}\theta\mathrm{d}\theta\int_0^{\infty}e^{-\rho^2}\rho^{2p+2q-1}\mathrm{d}\rho$

左边的积分即为 $\frac{1}{2}\mathrm{B}(p,q)$；令 $t=x^2$ 后，右边的积分即为 $\frac{1}{2}\Gamma(p+q)$，即

$$\Gamma(p)\Gamma(q)=\mathrm{B}(p,q)\Gamma(p+q)$$

Beta 函数有如下性质：

(1) 对称性

$$\mathrm{B}(x,y)=\mathrm{B}(y,x)$$

(2) 帕斯卡恒等式 (Pascal's identity) 的某种变形

$$\mathrm{B}(x,y)=\mathrm{B}(x,y+1)+\mathrm{B}(x+1,y)$$

(3) 递推公式（复现律）(recurrence rule)

$$\mathrm{B}(x+1,y)=\mathrm{B}(x,y)\cdot\frac{x}{x+y}$$

(4) 对于其中一个自变量的余元公式

$$\mathrm{B}(x,1-x)=\frac{\pi}{\sin(\pi x)}$$

(5)

$$\mathrm{B}(x,y)\cdot\mathrm{B}(x+y,1-y)=\frac{\pi}{x\sin(\pi y)}$$

(6) 与组合数的关系

$$(x+1)\binom{x}{y}=\frac{1}{B(x-y+1,y+1)}$$

当 x,y 都很大时，根据斯特林公式，$\mathrm{B}(x,y)\sim\sqrt{2\pi}\frac{x^{x-\frac{1}{2}}y^{y-\frac{1}{2}}}{(x+y)^{x+y-\frac{1}{2}}}$，如果只是 x 很大时，那么可以用 Gamma 函数估计：$\mathrm{B}(x,y)\sim\Gamma(y)x^{-y}$.

3.4 点火公式

点火公式是民间称呼，它指的是这个公式

定理 华里士积分 (Wallis integeral)

$$W_n=\int_0^{\frac{\pi}{2}}\sin^n(x)\mathrm{d}x=\begin{cases}\dfrac{(n-1)!!}{n!!} & n\text{为奇数}\\ \dfrac{(n-1)!!}{n!!}\dfrac{\pi}{2} & n\text{为正偶数}\end{cases}$$

经过简单换元，可以得到：

(1) $\int_0^{\frac{\pi}{2}}\cos^n(x)\mathrm{d}x=W_n$

(2) $\int_0^{\pi}\sin^n(x)=2W_n$

(3) $\int_0^{\pi} \cos^n(x) = \begin{cases} 0 & n\text{为正奇数} \\ 2W_n & n\text{为正偶数} \end{cases}$

(4) $\int_0^{2\pi} \sin^n(x) = \int_0^{2\pi} \cos^n(x) = \begin{cases} 0 & n\text{为正奇数} \\ 4W_n & n\text{为正偶数} \end{cases}$

其中双阶乘定义为

$$n!! = \begin{cases} n \times (n-2) \times (n-4) \times \cdots \times 3 \times 1 & n\text{为正奇数} \\ n \times (n-2) \times (n-4) \times \cdots \times 4 \times 2 & n\text{为正偶数} \end{cases}$$

以前证明点火公式都是使用分部积分求出 $\{W_n\}$ 的递推公式，再求出 W_1 然后解出通项公式，具体如下

$$W_n = \int_0^{\frac{\pi}{2}} (\sin^{n-2} x)(1-\cos^2 x)\mathrm{d}x = \int_0^{\frac{\pi}{2}} \sin^{n-2} x\mathrm{d}x - \int_0^{\frac{\pi}{2}} \sin^{n-2} x \cos^2 x\mathrm{d}x$$

其中

$$\int_0^{\frac{\pi}{2}} \sin^{n-2} x \cos^2 x\mathrm{d}x = \frac{\sin^{n-1} x}{n-1} \cos x\Big|_0^{\frac{\pi}{2}} + \frac{1}{n-1}\int_0^{\frac{\pi}{2}} \sin^{n-1} x \sin x\mathrm{d}x = 0 + \frac{1}{n-1}W_n$$

这就求出了 $W_n = W_{n-2} - \dfrac{W_n}{n-1}$，接着又因为 $W_0 = 1, W_1 = \dfrac{\pi}{2}$，就求出了 $\{W_n\}$ 的通项公式。

现在可以发现，华里士积分实际上是 B 函数等价表达式的特殊情况：

$$W_n = \int_0^{\frac{\pi}{2}} \sin^n x \cos^0 x\mathrm{d}x = \int_0^{\frac{\pi}{2}} \sin^{2\frac{n+1}{2}-1} x \cos^{2(\frac{1}{2})-1} x = \frac{1}{2}B\left(\frac{n+1}{2}, \frac{1}{2}\right) = \frac{\Gamma(\frac{n+1}{2})\Gamma(\frac{1}{2})}{2\Gamma(\frac{n}{2}+1)}$$

根据这个关系，可以把华里士积分的 n 扩展到实数范围。接着再分类讨论 Γ 函数：

(1) 若 n 是奇数，则 $\dfrac{n+1}{2}$ 是整数，$\dfrac{n}{2}+1$ 则不是，因此

$$W_n = \frac{\Gamma(\frac{n+1}{2})\Gamma(\frac{1}{2})}{2\Gamma(\frac{n}{2}+1)} = \frac{\frac{n-1}{2}!\sqrt{\pi}}{2\sqrt{\pi}\frac{n!!}{2^{(n+1)/2}}} = \frac{\frac{n-1}{2}!!2^{\frac{n-1}{2}}}{n!!} = \frac{(n-1)!!}{n!!}$$

(2) 若 n 是偶数，则 $\dfrac{n+1}{2}$ 不是整数，$\dfrac{n}{2}+1$ 则是，因此

$$W_n = \frac{\Gamma(\frac{n+1}{2})\Gamma(\frac{1}{2})}{2\Gamma(\frac{n}{2}+1)} = \frac{\frac{(n-1)!!}{2^{n/2}}\sqrt{\pi}\cdot\sqrt{\pi}}{2\frac{n}{2}!!} = \frac{\pi}{2}\cdot\frac{(n-1)!!}{n!!}$$

这就导出了点火公式。

把点火公式拓展一下，推导方法同上，可以得到：

$$\int_0^{\frac{\pi}{2}} \sin^p x \cos^q x\mathrm{d}x = \frac{1}{2}B\left(\frac{p+1}{2}, \frac{q+1}{2}\right) = \begin{cases} \dfrac{\pi}{2}\cdot\dfrac{(p-1)!!(q-1)!!}{(p+q)!!} & p,q\text{都为偶数} \\ \dfrac{(p-1)!!(q-1)!!}{(p+q)!!} & \text{其他} \end{cases}$$

接下来讨论一下 W_n 的增长趋势. 我们知道，在区间 $[0, \dfrac{\pi}{2}]$ 上，$\sin(x) \le x$. 并且由于 $\sin(x)$ 在该区间上递增，所以有 $\sin(\sin(x)) \le \sin(x)$，照此下去，得到更一般的 $\sin^{n+1}(x) \le \sin^n(x)$，根据定积分的性质，$W_n$ 是递减的数列.

还没完，根据以上关系可以推知

$$
\begin{aligned}
&\sin^{2n+1}(x) \le \sin^{2n}(x) \le \sin^{2n-1}(x) \\
\implies\ & W_{2n+1} < W_{2n} < W_{2n-1} \\
\implies\ & 1 \le \frac{W_{2n}}{W_{2n+1}} \le \frac{W_{2n-1}}{W_{2n+1}} = \frac{2n-1}{2n}
\end{aligned}
$$

根据夹逼准则，取极限得到 $\lim\limits_{n\to\infty} \dfrac{W_{2n}}{W_{2n+1}} = 1$，然而

$$
\lim_{n\to\infty} \frac{W_{2n}}{W_{2n+1}} = \lim_{n\to\infty} \frac{\pi}{2} \frac{(2n-1)!!(2n+1)!!}{(2n)!!(2n)!!}
$$

其中

$$
\begin{aligned}
(2n)!! &= 2 \times 4 \times 6 \times \cdots \times (2n) = \prod_{k=1}^{n} 2k \\
(2n+1)!! &= 3 \times 5 \times 7 \times \cdots \times (2n+1) = \prod_{k=1}^{n} (2k+1) \\
(2n-1)!! &= 1 \times 3 \times 5 \times \cdots \times (2n-1) = \prod_{k=1}^{n} (2k-1)
\end{aligned}
$$

所以上式可以写为

$$
\lim_{n\to\infty} \frac{W_{2n}}{W_{2n+1}} = \frac{\pi}{2} \lim_{n\to\infty} \prod_{k=1}^{n} \left(\frac{2k+1}{2k} \cdot \frac{2k-1}{2k} \right) = 1
$$

整理一下，就得到了

定理　华里士乘积 (Wallis product)

$$
\frac{\pi}{2} = \prod_{n=1}^{\infty} \left(\frac{2n}{2n+1} \cdot \frac{2n}{2n-1} \right) = \left(\frac{2}{1}\cdot\frac{2}{3}\right)\left(\frac{4}{3}\cdot\frac{4}{5}\right)\left(\frac{6}{5}\cdot\frac{6}{7}\right)\cdots
$$

还有更简单的推导方法，直接代 $z = 1/2$ 到余元公式即可

$$
\Gamma^2\left(\frac{1}{2}\right) = \pi = -\frac{1}{2} \cdot \frac{1}{(1/2)^2} \prod_{n=1}^{\infty} \left(1 - \frac{(1/2)^2}{n^2}\right)^{-1} = 2 \prod_{n=1}^{\infty} \left(\frac{2n}{2n+1} \cdot \frac{2n}{2n-1} \right)
$$

3.5 高斯积分

上面的式子 $\Gamma^2\left(\dfrac{1}{2}\right) = \pi$ 告诉我们

$$
\sqrt{\pi} = \Gamma\left(\frac{1}{2}\right) := \int_0^{\infty} t^{-1/2} e^{-t} \mathrm{d}t = 2 \int_0^{\infty} e^{-t} \mathrm{d}\sqrt{t} \xlongequal{x=\sqrt{t}} 2 \int_0^{\infty} e^{-x^2} \mathrm{d}x
$$

考虑到被积函数是偶函数，所以这就得到了高斯积分

定理　高斯积分或欧拉-泊松积分

$$
\int_{-\infty}^{\infty} e^{-x^2} \mathrm{d}x = \sqrt{\pi}
$$

还记得以前是怎样求解这个积分的吗?

$$\begin{aligned}\int_{-\infty}^{+\infty} e^{-x^2}\mathrm{d}x &= \sqrt{\int_{-\infty}^{+\infty} e^{-x^2}\mathrm{d}x \cdot \int_{-\infty}^{+\infty} e^{-y^2}\mathrm{d}y} \\ &= \sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\mathrm{d}x\mathrm{d}y} = \sqrt{\int_0^{2\pi}\int_0^{+\infty} e^{-\rho^2}\rho\mathrm{d}\rho\mathrm{d}\theta} = \sqrt{\pi}\end{aligned}$$

更一般地，有

(1) $\displaystyle\int_{-\infty}^{\infty} e^{-a(x+b)^2} = \sqrt{\frac{\pi}{a}}$

(2) $\displaystyle\int_{-\infty}^{\infty} ae^{-\frac{(x-b)^2}{2c^2}} = \sqrt{2\pi}a|c|$

(3) $\displaystyle\int_{-\infty}^{\infty} e^{-ax^2+bx+c} = \sqrt{\frac{\pi}{a}e^{\frac{b^2}{4a}+c}}$

容易发现，上面这些积分和正态分布函数有点像. 确实，可以定义一系列函数

定义　正态分布函数及其衍生

(1) 误差函数 (error function)

$$\mathrm{erf}(z) := \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$$

(2) 互补误差函数 (complementary error function)

$$\mathrm{erfc}(z) := 1 - \mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_z^{+\infty} e^{-t^2}dt$$

(3) 虚误差函数 (imaginary error function)

$$\mathrm{erfi}(z) := -i\mathrm{erf}(iz) = \frac{2}{\sqrt{\pi}}\int_0^{iz} e^{-t^2}dt$$

(4) 标准正态分布函数 (standard normal CDF)

$$\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\mathrm{d}t$$

(5) 正态分布右尾函数 (standard normal CCDF)

$$Q(x) := 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty} e^{-\frac{t^2}{2}}\mathrm{d}t$$

当 $x \ll 1$ 和 $x \gg 1$ 时，误差函数可以用不同的麦克劳林展开式逼近:

$$\begin{aligned}\mathrm{erf}(z) &= \frac{1}{\pi}e^{-x^2}\sum_{k=0}^{\infty}\frac{(2x)^{2k+1}}{(2k+1)!!}, \quad x \ll 1 \\ \mathrm{erf}(z) &\sim 1 - \frac{e^{-x^2}}{\sqrt{\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k(2k-1)!!}{2^k}x^{-(2k+1)}, \quad x \gg 1\end{aligned}$$

3.6 斯特林公式

之前提到，当 n 很大时，有

$$n! = \Gamma(n+1) \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

此即为斯特林近似公式(Stirling's approximation)，接下来简单推导一下，然后再用欧拉-麦克劳林公式得出精确的斯特林渐进展开式.

证明：

$$n! = \Gamma(n+1) = \int_0^\infty x^n e^{-x}\mathrm{d}x \xlongequal[\mathrm{d}x=n\mathrm{d}t]{x=nt} \int_0^\infty (nt)^n e^{-nt} n\mathrm{d}t = n^{n+1}\int_0^\infty t^n e^{-nt}\mathrm{d}x = n^{n+1}\int_0^\infty e^{n(\ln t - t)}\mathrm{d}t$$

接下来使用的技巧称为拉普拉斯方法. 对于一个二阶可导的函数 $f(x)$，最值为 x_0，在最值处用泰勒公式展开

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 + o(x^2)$$

由于是最值，所以 $f'(x_0) = 0$，此时即为

$$f(x) = f(x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 + o(x^2)$$

所以

$$\int e^{nf(x)}\mathrm{d}x \approx e^{nf(x_0)}\int e^{\frac{1}{2}nf''(x_0)(x-x_0)^2}\mathrm{d}x$$

在这里，$f(x) = \ln x - x$，令 $f'(x_0) = \left(\frac{1}{x} - 1\right)_{x=x_0} = 0$，得到 $x_0 = 1$，于是

$$n! = n^{n+1}\int_0^\infty e^{n(\ln t - t)}\mathrm{d}t \approx n^{n+1}e^{-n}\int_0^\infty e^{-\frac{1}{2}n(x-1)^2}\mathrm{d}x \xlongequal[\mathrm{d}x=\sqrt{2/n}\mathrm{d}u]{u^2=\frac{1}{2}n(x-1)^2} n^{n+1}e^{-n}\sqrt{\frac{2}{n}}\int_0^\infty e^{-u^2}\mathrm{d}u$$

注意到当 n 很大时，$x<0$ 时，$e^{-\frac{1}{2}n(x-1)^2}$ 非常接近 0，所以可以把积分区间扩展到整个 $(-\infty,\infty)$：

$$n! \approx n^{n+1}e^{-n}\sqrt{\frac{2}{n}}\int_{-\infty}^\infty e^{-u^2}\mathrm{d}u = n^{n+1}e^{-n}\sqrt{\frac{2\pi}{n}} = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

这就得到了最基本的斯特林近似公式.

3.7 分数阶导数和分数阶微积分

我们学过 $\frac{\mathrm{d}}{\mathrm{d}x}$ 表示什么意思，也知道 $\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ 表示什么意思，但我们可没学过 $\frac{\mathrm{d}^{\sqrt{2}}}{\mathrm{d}x^{\sqrt{2}}}$，这节就把微分的积分的阶数扩展到任意实数.

令

$$\begin{cases} I^0[f](x) := f(x) \\ I^{n+1}[f](x) := \displaystyle\int_a^x I^n[f](t)\mathrm{d}t \end{cases}$$

将前几项写出来，注意区分自变量和被积分的哑变量.

$$\begin{aligned}
I^0[f](x_0) &= f(x_0) \\
I^1[f](x_1) &= \int_a^{x_1} f(x_0)\mathrm{d}x_0 \\
I^2[f](x_2) &= \int_a^{x_2}\int_a^{x_1} f(x_0)\mathrm{d}x_0\mathrm{d}x_1 \\
I^3[f](x_3) &= \int_a^{x_3}\int_a^{x_2}\int_a^{x_1} f(x_0)\mathrm{d}x_0\mathrm{d}x_1\mathrm{d}x_2 \\
I^n[f](x_n) &= \int_a^{x_n}\cdots\int_a^{x_2}\int_a^{x_1} f(x_0)\mathrm{d}x_0\mathrm{d}x_1\cdots\mathrm{d}x_{n-1}
\end{aligned}$$

可见 I^{n+1} 把 I^n 的自变量当做哑变量做了一次变上限积分，而该上限是 I^{n+1} 的自变量. 现在我们就来计算这一系列 I^n 是多少. 首先计算 $I^2[f](x)$，使用二重积分换序的方法.

$$I^2[f](x_2)=\int_a^{x_2}\int_a^{x_1}f(x_0)\mathrm{d}x_0\mathrm{d}x_1=\int_a^{x_2}\int_{x_0}^{x_2}f(x_0)\mathrm{d}x_1\mathrm{d}x_0=\int_a^{x_2}f(x_0)(x_2-x_0)\mathrm{d}x_0$$

变量 x_1 消失了. 接下来计算 $I^3[f](x)$，注意自变量的变化

$$\begin{aligned}I^3[f](x_3)&=\int_a^{x_3}\int_a^{x_2}f(x_0)(x_2-x_0)\mathrm{d}x_0\mathrm{d}x_2=\int_a^{x_3}\int_{x_0}^{x_3}f(x_0)(x_2-x_0)\mathrm{d}x_2\mathrm{d}x_0\\&=\int_a^{x_3}f(x_0)\frac{(x_3-x_0)^2}{2}\mathrm{d}x_0\end{aligned}$$

变量 x_2 消失了. 接下来计算 $I^4[f](x_4)$，注意自变量的变化

$$\begin{aligned}I^4[f](x_4)&=\int_a^{x_4}\int_a^{x_3}f(x_0)\frac{(x_3-x_0)^2}{2}\mathrm{d}x_0\mathrm{d}x_3=\int_a^{x_4}\int_{x_0}^{x_4}f(x_0)\frac{(x_3-x_0)^2}{2}\mathrm{d}x_3\mathrm{d}x_0\\&=\int_a^{x_4}f(x_0)\frac{(x_3-x_0)^3}{6}\mathrm{d}x_0\end{aligned}$$

根据以上结果，我们猜测：

$$I^n[f](x_n)=\int_a^{x_n}f(x_0)\frac{(x_n-x_0)^{n-1}}{(n-1)!}\mathrm{d}x_0$$

接下来用数学归纳法证明这个猜想，显然 $n=1$ 的情况成立，所以接下来只用证明递推的阶段.

$$\begin{aligned}I^{n+1}[f](x)&=\int_a^{x_{n+1}}I^n[f](x_n)\mathrm{d}x_{n+1}=\int_a^{x_{n+1}}\int_a^{x_n}f(x_0)\frac{(x_n-x_0)^{n-1}}{(n-1)!}\mathrm{d}x_0\mathrm{d}x_{n+1}\\&=\int_a^{x_{n+1}}\int_{x_0}^{x_{n+1}}f(x_0)\frac{(x_n-x_0)^{n-1}}{(n-1)!}\mathrm{d}x_{n+1}\mathrm{d}x_0=\int_a^{x_{n+1}}f(x_0)\frac{(x_{n+1}-x_0)^n}{n!}\mathrm{d}x_0\end{aligned}$$

证毕. 这就是

定理　柯西迭代积分公式 (Cauchy formula for repeated integration)

$$I^n[f](x)=\int_a^x\int_a^{x_{n-1}}\cdots\int_a^{x_2}\int_a^{x_1}f(x_0)\mathrm{d}x_0\mathrm{d}x_1\cdots\mathrm{d}x_{n-1}=\frac{1}{(n-1)!}\int_a^x f(t)(x-t)^{n-1}\mathrm{d}t$$

以上讨论了这么多，看起来就是随便定义了一个函数然后推导它的另一种表达式，和本节的主题似乎没有联系. 但假如现在令 $I^{-n}[f](x)=\dfrac{\partial^n}{\partial x^n}f(x)$，然后就可以发现 $\forall n,m\in\mathbb{Z}:I^n[I^m[f]](x)=I^{n+m}[f](x)$. 证明思路很简单，只要注意到变上限积分和导数是互逆运算即可. 如果将式中的阶乘推广为 Γ 函数，则可以把 n 的范围推广到任意实部为正的复数，而这就是

定义　黎曼-刘维尔积分 (Riemann-Liouville integral)

$$I^\alpha[f](x)=\frac{1}{\Gamma(\alpha)}\int_a^x f(t)(x-t)^{\alpha-1}\mathrm{d}t$$

例如求 $f(x)\equiv1$ 的任意阶积分，为了方便，令积分下限 $a=0$.

$$I^\alpha[f](x)=\int_0^x(x-t)^{\alpha-1}\mathrm{d}t=\frac{x^\alpha}{\alpha\Gamma(\alpha)}$$

当 $\alpha=1/2$ 时，得到 $I^\alpha[f](x)=2\sqrt{\dfrac{x}{\pi}}$. $f(x)\equiv1$ 积 0 次分得到的 x 的系数为 0，积 1 次分则为 1，积半次则为 1/2，非常符合直觉.

3.8 高维球的“体积”

回顾欧式空间中圆和球的定义，它们都是到定点的距离为定值的点的集合，唯一的不同只体现在维度上.

- 在二维平面，即 xy 直角坐标系下，定点在原点的圆的方程为 $x^2+y^2=r^2$，r 称为圆的半径.
- 在三维空间，即 xyz 直角坐标系下，定点在原点的球的方程为 $x^2+y^2+z^2=r^2$，r 称为球的半径.
- 由此可以推广到 n 维欧式空间中，在 $x_1x_2\cdots x_n$ 的直角坐标系下，定点在原点的超球的方程为 $x_1^2+x_2^2+\cdots+x_n^2=r^2$，$r$ 为该超球的半径.

4 勒奇超越函数系

之所以称为勒奇超越函数系，是因为除了本身之外，这一系列的函数都是勒奇超越函数的一种特殊情况. 然而勒奇超越函数仅仅是它们非常保守的推广，之后还可以再推广至超几何函数. 然而超几何函数的推广程度太大以至于值得单独开一节来讨论，所以本节讨论的函数局限在勒奇超越函数下.

4.1 黎曼 zeta 函数

黎曼 zeta 函数很有来头，它有好几种适用范围不同的定义:

(1) 欧拉原始的定义

$$\zeta(s):=\sum_{i=1}^{\infty}i^{-s}=1+\frac{1}{2^s}+\frac{1}{3^s}+\cdots,\quad \mathrm{Re}(s)>1$$

(2) 黎曼的积分定义

$$\zeta(s):=\frac{1}{\Gamma(s)}\int_0^{+\infty}\frac{x^{s-1}}{e^x-1}\mathrm{d}x,\quad \mathrm{Re}(s)>1$$

(3) 解析延拓

$$\zeta(s):=\frac{\Gamma(1-s)}{2\pi i}\oint_\gamma\frac{t^{s-1}}{e^{-t}-1}\mathrm{d}t,\quad s\neq 1$$

上面的前两种定义其实很容易相互推导，注意到第二种定义中前面除了一个 $\Gamma(s)$，所以尝试把它乘到第一种定义中:

$$\zeta(s)\Gamma(s)=\sum_{k=1}^{\infty}\frac{1}{k^s}\int_0^{+\infty}t^{s-1}e^{-t}\mathrm{d}t=\sum_{k=1}^{\infty}\int_0^{+\infty}\left(\frac{t}{k}\right)^{s-1}e^{-t}\frac{\mathrm{d}t}{k}$$

换元: $u=t/k$，则 $t=ku$，然后

$$\text{原式}=\sum_{k=1}^{\infty}\int_0^{+\infty}u^{s-1}e^{-ku}\mathrm{d}u=\int_0^{+\infty}u^{s-1}\sum_{k=1}^{\infty}e^{-ku}\mathrm{d}u$$

由等比数列求和公式，$\displaystyle\sum_{k=1}^{\infty}e^{-ku}=\lim_{k\to\infty}\frac{e^{-u}(1-e^{-u})^k}{1-e^{-u}}=\frac{e^{-u}}{1-e^{-u}}=\frac{1}{e^u-1}$，代入即得到

$$\text{原式}=\int_0^{+\infty}\frac{u^{s-1}}{e^u-1}\mathrm{d}u\Rightarrow\zeta(s)=\frac{1}{\Gamma(s)}\int_0^{+\infty}\frac{u^{s-1}}{e^u-1}\mathrm{d}u$$

黎曼进行了进一步的解析延拓，得到这样一个函数方程:

$$\zeta(s)=2^s\pi^{s-1}\sin\left(\frac{\pi s}{2}\right)\Gamma(1-s)\zeta(1-s)$$

ζ 函数的某些取值可以得到很奇妙的结果：

$$\zeta(2)=1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots=\sum_{i=1}^{\infty}\frac{1}{i^2}=\frac{\pi^2}{6}$$

这个等式称为巴塞尔问题，最先由欧拉解决。

欧拉首先注意到 $\sin(x)$ 的麦克劳林展开：

$$\sin(x)=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots\Rightarrow\frac{\sin(x)}{x}=x-\frac{x^2}{3!}+\frac{x^2}{5!}-\cdots$$

由于 $\frac{\sin(x)}{x}=0$ 的根是 $x=\pm k\pi,\quad k\in\mathbb{Z}^*$，在零点处必有 $1-\frac{x}{k\pi}=0$，所以可以把这个函数写成无穷乘积的形式[15]，再使用平方差公式：

$$\begin{aligned}\frac{\sin(x)}{x}&=\left(1-\frac{x}{\pi}\right)\left(1+\frac{x}{\pi}\right)\left(1-\frac{x}{2\pi}\right)\left(1+\frac{x}{2\pi}\right)\left(1-\frac{x}{3\pi}\right)\left(1+\frac{x}{3\pi}\right)\cdots\\&=\left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{(2\pi)^2}\right)\left(1-\frac{x^2}{(3\pi)^2}\right)\cdots\end{aligned}$$

把 x^2 项提出来，并和麦克劳林展开式的 x^2 项的系数比较：

$$-\frac{1}{\pi^2}-\frac{1}{(2\pi)^2}-\frac{1}{(3\pi)^2}-\cdots=\frac{1}{3!}$$

整理一下，就可以得到最终结果：

$$-\frac{1}{\pi^2}(1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots)=-\frac{1}{3!}\Rightarrow 1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots=\frac{\pi^2}{6}$$

$$\zeta(3)=1+\frac{1}{2^3}+\frac{1}{3^3}+\cdots=\sum_{i=1}^{\infty}\frac{1}{i^3}=1.20205\cdots$$

等式最右侧称为阿培里常数 (Apery constant)，是一个无理数，在量子电动力学中可以见到它。

$$\zeta(4)=1+\frac{1}{2^4}+\frac{1}{3^4}+\cdots=\sum_{i=1}^{\infty}\frac{1}{i^4}=\frac{\pi^4}{90}$$

这个数出现在黑体辐射中的斯特藩-玻尔兹曼定律 (Stefan-Boltzmann law) 中。

$$\zeta(1)=1+\frac{1}{2}+\frac{1}{3}+\cdots=\sum_{i=1}^{\infty}\frac{1}{i}$$

得到调和级数，这个级数不收敛。

根据黎曼的函数方程，可以发现

$$\zeta(-1)=2^{-1}\pi^{-2}\sin\left(\frac{-\pi}{2}\right)\Gamma(2)\zeta(2)=-\frac{1}{12}$$

同时注意到 $\zeta(-1)$ 表示的算式即为 $1+2+3+4+\cdots$，于是就得到了暴论：全体自然数的和为 $-\frac{1}{12}$. 错就错在欧拉对 ζ 函数的定义不允许以-1 作为自变量，所以 $\zeta(-1)$ 并不表示全体自然数的和。不过需要注意的是，全体自然数的拉马努金和确实是 $-\frac{1}{12}$.

根据黎曼的函数方程，还可以发现，当 $s=-2k,k\in\mathbb{N}_+$ 时，$\zeta(s)=0$，这些零点很好找到，称为平凡零点。zeta 函数还有另外一些零点，需要更多的研究才能找到，这样的零点称为非平凡零点。黎曼找到了很多，并发现它们的实部都是 $\frac{1}{2}$，于是他提出了一个著名的猜想：

[15]这种分解是否成立需要用到魏尔斯特拉斯分解定理来证明，但欧拉当时认为它是成立的

命题 黎曼猜想 (Riemann hypothesis)

$\zeta(s)$ 的非平凡零点都满足 $\mathrm{Re}(s)=\frac{1}{2}$.

黎曼猜想既是希尔伯特的 23 个难题之一，也是当今世界七大数学难题之一.

4.2 赫尔维茨 zeta 函数

赫尔维茨 (Hurwitz) 的 ζ 函数是黎曼 ζ 函数的一种推广，定义为

$$\zeta(s,q)=\sum_{n=0}^{\infty}\frac{1}{(n+q)^s}$$

4.3 狄利克雷 beta 函数

$$\beta(s)=\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^s}$$

4.4 狄利克雷 eta 函数

狄利克雷的 η 是黎曼 ζ 函数的交错级数，即

$$\eta(s)=\sum_{n=0}^{\infty}\frac{(-1)^{n-1}}{n^s}$$

4.5 勒让德 chi 函数

勒让德 χ 函数的特殊之处就在于它的泰勒级数恰好也是它的狄利克雷级数，定义为

$$\chi_{\nu}(z)=\sum_{n=0}^{\infty}\frac{z^{2n+1}}{(2n+1)^{\nu}}$$

4.6 多对数函数（容基耶尔函数）

多重对数是自然对数泰勒展开式的推广，定义为

$$\mathrm{Li}_s(z)=\sum_{n=0}^{\infty}\frac{z^n}{n^s}$$

4.7 勒奇超越函数

勒奇超越函数 (Lerch transcendent) 是以上几种特殊函数的推广，定义为

$$\Phi(z,s,\alpha)=\sum_{n=0}^{\infty}\frac{z^n}{(n+\alpha)^s}$$

它与以上几种函数有以下关系

$$
\begin{aligned}
\zeta(s) &= \Phi(1,s,1)\\
\zeta(s,q) &= \Phi(1,s,q)\\
\beta(s) &= 2^{-s}\Phi\left(-1,s,\frac{1}{2}\right)\\
\eta(s) &= \Phi(-1,s,1)\\
\chi_n(z) &= 2^{-n}z\Phi\left(z^2,2,\frac{1}{2}\right)\\
\mathrm{Li}_s(z) &= z\Phi(z,s,1)
\end{aligned}
$$

5 基本初等复合函数的积分

虽然基本初等函数的积分很简单，基本初等函数经过四则运算再积分也可以写出初等表达式，但是基本初等函数的复合函数的积分可能就不存在初等表达式。以下列出几个，重点在于这些函数的定积分的值。

(1) 菲涅尔积分 (Fresnel integral)

$$\mathrm{S}(x) = \int_0^x \sin(t^2)\mathrm{d}t \quad \text{和} \quad \mathrm{C}(x) = \int_0^x \cos(t^2)\mathrm{d}t$$

(2) 三角函数和双曲函数积分

$$
\begin{aligned}
\mathrm{sinc}(x) &= \frac{\sin(x)}{x} & \mathrm{Si}(x) &= \int_0^x \mathrm{sinc}(t)\mathrm{d}t\\
\mathrm{cosc}(x) &= \frac{\cos(x)}{x} & \mathrm{Ci}(x) &= \int_0^x \mathrm{cosc}(t)\mathrm{d}t\\
\mathrm{sinhc}(x) &= \frac{\sinh(x)}{x} & \mathrm{Shi}(x) &= \int_0^x \mathrm{sinhc}(t)\mathrm{d}t\\
\mathrm{coshc}(x) &= \frac{\cos(x)}{x} & \mathrm{Chi}(x) &= \int_0^x \mathrm{coshc}(t)\mathrm{d}t
\end{aligned}
$$

(3) 指数积分 (exponentiation integral)

$$\mathrm{Ei}(x) = \int_0^x \frac{e^{-t}}{t}\mathrm{d}t$$

(4) 对数积分 (logarithm integral)

$$\mathrm{li}(x) = \int_0^x \frac{\ln(t)}{t}\mathrm{d}t$$

6 柱函数（贝塞尔函数）系

6.1 为什么会有这个函数？

回顾线性代数的知识：

> 在线性空间 $\langle S;+,\cdot\rangle$ 中定义一个映射 f，对于任意常数 a,b，如果该变换满足
>
> $$f(ax+by) = af(x)+bf(y)$$

就称该映射具有线性。更进一步地，如果存在向量 x，使得存在常数 λ，满足 $f(x)=\lambda x$，则称该 λ 为该映射的特征值，x 称为该映射对应于 λ 的特征向量.

我们知道二维平面上的拉普拉斯算子 $\nabla^2=\dfrac{\partial^2}{\partial x^2}+\dfrac{\partial^2}{\partial y^2}$ 是 $\mathbb{R}^3$ 上的线性映射（不信你自己验证）. 自然地，二维向量的特征值就应该满足

$$\nabla^2 f(x,y)=\lambda f(x,y)$$

这个式子称为亥姆霍兹方程(Helmholtz equation). 亥姆霍兹方程在物理学中经常用到，它的解（也就是特征值）有无数多个. 换个坐标系，根据极坐标系和直角坐标系的关系：

$$\rho=\sqrt{x^2+y^2}\qquad \theta=\arg(x+iy)$$

将拉普拉斯算子改写为极坐标的形式：

$$\nabla^2=\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2}{\partial\theta^2}=\rho\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho^2}+\frac{1}{\rho^2}\frac{\partial^2}{\partial\theta^2}$$

根据分离变量(separation of variables)的思想，假设函数 $f(\rho,\theta)$ 能被写成两个单变量函数的乘积：$f(\rho,\theta)\equiv R(\rho)T(\theta)$，亥姆霍兹方程就可以改写成这个形式（将 $R(\rho)$ 简写为 R，$T(\theta)$ 简写为 T）：

$$R''T+\frac{1}{\rho}R'T+\frac{1}{\rho^2}RT''=\lambda RT$$

两边乘以 ρ^2，再除以 RT，得到

$$\rho^2\frac{R''}{R}+\rho\frac{R'}{R}+\frac{T''}{T}=\lambda\rho^2$$

左边是三个单项式相加的形式，其中参数 θ 只与第三个单项式 $\dfrac{T''(\theta)}{T(\theta)}$ 有关，然而右边却与 θ 无关，这说明 $\dfrac{T''(\theta)}{T(\theta)}$ 必须是一个常数，否则当 θ 发生变化后，左边的值就会发生变化，而这种变化是与 θ 无关的参数 ρ 所无法抵消的. 所以设 $\dfrac{T''(\theta)}{T(\theta)}=\mu$，这是一个很好解的常微分方程，再看另外一边：

$$\rho^2\frac{R''}{R}+\rho\frac{R'}{R}+\mu=\lambda\rho^2$$

两边乘 R，整理一下得到

$$\rho^2R''+\rho R'+(\mu-\lambda\rho^2)R=0$$

这个微分方程称为贝塞尔方程，它的解不一定能表示为初等形式，由其隐式定义的函数称为贝塞尔函数(Bessel function).

定义 贝塞尔函数

微分方程

$$x^2y\ +xy\ +(x^2-\alpha^2)y=0$$

的解称为 α 阶的贝塞尔函数，记作 $J_\alpha(x)$，其中 x 称为宗量，阶数和宗量的范围为复数集.

再换一种坐标系. 在三维空间，拉普拉斯算子为 $\nabla^2=\dfrac{\partial^2}{\partial x^2}+\dfrac{\partial^2}{\partial y^2}+\dfrac{\partial^2}{\partial z^2}$. 根据球坐标系与直角坐标系的关系：

$$r=\sqrt{x^2+y^2+z^2}\qquad \theta=\arg(x+iy)\qquad \varphi=\arccos(z/r)$$

将拉普拉斯算子改写为球坐标的形式，得到一个非常难看的式子：

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}$$

就像上面那样，根据分离变量的思想，假设函数 $f(r,\theta,\varphi)=R(r)T(\theta)P(\varphi)$，代入亥姆霍兹方程：

$$\frac{TP}{r^2}\left(r^2R'\right)'+\frac{RP}{r^2\sin^2\theta}\left(T'\sin\theta\right)'+\frac{RT}{r^2\sin^2\theta}P''=\lambda RTP$$

两边同除 RTP，再乘以 r^2，得到

$$\frac{\left(r^2R'\right)'}{R}+\frac{\left(T'\sin\theta\right)'}{T\sin^2\theta}+\frac{P''}{P\sin^2\theta}=\lambda r^2$$

左侧是三个单项式，注意其中的第二、第三个单项式与 r 无关，基于相同理由，我们又可以把它设为常数 μ，于是这就出来了一个单变量的微分方程，将括号展开：

$$r^2R''+2rR'+(\mu-\lambda r^2)R=0$$

这个微分方程称为球贝塞尔方程，它的解不一定能表示为初等形式，由其隐式定义的函数称为球贝塞尔函数(spherical Bessel function).

定义　球贝塞尔函数

微分方程

$$x^2y\ +xy\ +(x^2-\alpha^2)y=0$$

的解称为 α 阶的贝塞尔函数，记作 $J_\alpha(x)$，其中 x 称为宗量，阶数和宗量的范围为复数集.

6.2　解贝塞尔方程：Frobenius 法

我们知道二阶、线性、齐次、常系数的微分方程怎么解，但是把“常系数”去掉呢？就会解出一大堆特殊函数，可以说博物馆中的所有特殊函数都可以通过解这样的方程得到.

定义　贝塞尔函数

第一类贝塞尔函数

$$J_\alpha(x)=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!\Gamma(n+\alpha+1)}\left(\frac{x}{2}\right)^{2n+\alpha}$$

第二类贝塞尔函数

$$Y_\alpha(x)=\frac{J_\alpha(x)\cos(\alpha\pi)-J_{-\alpha}(x)}{\sin(\alpha\pi)}$$

汉克尔函数

$$H_\alpha^{(1)}(x)=J_\alpha(x)+iY_\alpha(x)$$
$$H_\alpha^{(2)}(x)=J_\alpha(x)-iY_\alpha(x)$$

修正贝塞尔函数

$$I_\alpha(x)=i^{-\alpha}J_\alpha(ix)$$
$$K_\alpha(x)=\frac{\pi}{2}\frac{I_{-\alpha}(x)-I_\alpha(x)}{\sin\alpha\pi}$$

球贝塞尔函数

$$j_n(x)=\sqrt{\frac{\pi}{2x}}J_{n+\frac{1}{2}}(x)$$
$$y_n(x)=\sqrt{\frac{\pi}{2x}}Y_{n+\frac{1}{2}}(x)$$

球汉克尔函数

$$h_n^{(1)}(x)=j_n(x)+iy_n(x)$$
$$h_n^{(2)}(x)=j_n(x)-iy_n(x)$$

命题　某个算不出来的积分

利用修正贝塞尔函数可以得到

$$\int e^{\cos(x)}\mathrm{d}x=\int e^{\sin(x)}\mathrm{d}x=\int e^{\cos(x)}\mathrm{d}t=I_0(1)x+2\sum_{n=1}^{\infty}I_n(1)\frac{\sin(nx)}{n}+C$$

令 $u=x-\frac{\pi}{2}$，则根据第二类换元法

$$\int e^{\sin(x)}\mathrm{d}x=\int e^{\sin(u+\pi/2)}\mathrm{d}\left(u+\frac{\pi}{2}\right)=\int e^{\cos(u)}\mathrm{d}u$$

接着对 $e^{\cos(x)}$ 作三角函数项的傅里叶展开

$$e^{\cos(x)}=\frac{1}{2}a_0+\sum_{n=1}^{\infty}(a_n\cos(nx)+b_n\sin(nx))$$

代入修正贝塞尔函数 $I_n(z)=\frac{1}{\pi}\int_0^\pi e^{z\cos(x)}\cos(nx)\mathrm{d}x$

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}e^{\cos(x)}\cos(nx)\mathrm{d}x=2I_n(1)$$
$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}e^{\cos(x)}\sin(nx)\mathrm{d}x=0$$

于是

$$e^{\cos(x)}=I_0(1)+2\sum_{n=1}^{\infty}I_n(1)\cos(nx)$$

逐项积分就是

$$\int e^{\cos(x)}\mathrm{d}t = I_0(1)x + 2\sum_{n=1}^{\infty}\frac{I_n(1)}{n}\sin(nx) + C$$

7 勒让德多项式（球多项式）

7.1 利用正交性导出

泰勒展开告诉我们，$\{x^n\}_{n=0}^{\infty}$ 是 $C(\mathbb{R})$ 的一个完备基底，但是这个基底却不是正交的，甚至在任意一个子集上都不正交，理由很简单，求内积时 $\langle x^n, x^n\rangle$ 和 $\langle x^{n-1}, x^{n+1}\rangle$ 的被积函数是相同的，而如果正交的话前者应为非零，后者应为零，矛盾. 于是我们自然就会思考将它们正交化，做出一组完备的正交基底.

> 回顾线性代数的知识，当向量 $\{\vec{\alpha}_k\}_{k=1}^{n}$ 线性无关时，它们可以张成 n 维线性空间，而该线性空间的一组完备正交基底 $\{\vec{\beta}_k\}_{k=1}^{n}$ 可以根据 $\{\vec{\alpha}_k\}_{k=1}^{n}$ 利用格拉姆-施密特正交化构造出来：
>
> $$\vec{\beta}_k = \begin{cases} \vec{\alpha}_k & k=1 \\ \vec{\alpha}_k - \displaystyle\sum_{m=0}^{k-1}\frac{\langle\vec{\alpha}_k, \vec{\beta}_m\rangle}{\langle\vec{\beta}_m, \vec{\beta}_m\rangle}\vec{\beta}_m & k>1 \end{cases}$$

接下来要做的事是选取一个好的函数空间，$C(\mathbb{R})$ 并不是一个好的选择，因为 $\lambda x.1$ 在 (R) 上的积分是无穷. 所以我们选择了 $C([-1,1])$. 遵循格拉姆-施密特正交化方法，在区间 $[-1,1]$ 上，有：

$$\begin{aligned}
p_0(x) &= 1 \\
p_1(x) &= x - \frac{\langle x, 1\rangle}{\langle 1, 1\rangle} \\
&= x - \frac{\int_{-1}^{1} x\mathrm{d}x}{\int_{-1}^{1} 1\mathrm{d}x} = x \\
p_2(x) &= x^2 - \frac{\langle x^2, 1\rangle}{\langle 1, 1\rangle} - \frac{\langle x^2, x\rangle}{\langle x, x\rangle}x \\
&= x^2 - \frac{\int_{-1}^{1} x^2\mathrm{d}x}{\int_{-1}^{1} 1\mathrm{d}x} - \frac{\int_{-1}^{1} x^3\mathrm{d}x}{\int_{-1}^{1} x^2\mathrm{d}x}x = x^2 - \frac{1}{3} \\
p_3(x) &= x^3 - \frac{\langle x^3, 1\rangle}{\langle 1, 1\rangle} - \frac{\langle x^3, x\rangle}{\langle x, x\rangle}x - \frac{\langle x^3, x^2-\frac{1}{3}\rangle}{\langle x^2-\frac{1}{3}, x^2-\frac{1}{3}\rangle}(x^2-\frac{1}{3}) \\
&= x^3 - \frac{\int_{-1}^{1} x^3\mathrm{d}x}{\int_{-1}^{1} 1\mathrm{d}x} - \frac{\int_{-1}^{1} x^4\mathrm{d}x}{\int_{-1}^{1} x^2\mathrm{d}x}x - \frac{\int_{-1}^{1} x^3(x^2-\frac{1}{3})\mathrm{d}x}{\int_{-1}^{1}(x^2-\frac{1}{3})^2\mathrm{d}x}(x^2-\frac{1}{3}) = x^3 - \frac{3}{5}x
\end{aligned}$$

我们希望得到的正交基底的性质尽可能模仿 $\{x^n\}_{n=0}^{\infty}$ 的性质，后者简单列举两条

- $(\lambda x.x^n)(1) = 1$
- 在区间 $[-1,1]$ 上，$\langle x^n, x^n\rangle = \dfrac{2}{2n+1}$

8 椭圆函数系

9 超几何函数系

超几何函数最初由高斯定义，在勒奇超越函数的基础上更推进一步. 最终得到的推广的超几何函数是一个包容性非常强的函数，以上提到的大部分特殊函数，包括勒奇超越函数系、贝塞尔函数系和椭圆函数系都是它的特殊情况，因此超几何函数可以非常深刻地揭示以上众多特殊函数的共性.

定义　超几何函数

上升阶乘 (rising factorial) 或伯赫哈默尔符号 (Pochhammer symbol)：

$$n^{(m)} = \begin{cases} 1 & m = 0 \\ \prod\limits_{k=m}^{n+m-1} k = n(n+1)\cdots(n+m-1) & m > 0 \end{cases}$$

合流超几何函数 (confluent hypergeometric function) 或库莫函数 (Kummer function)

$$M(a,b,z) = \sum_{n=0}^{\infty} \frac{a^{(n)}}{b^{(n)}} \frac{z^n}{n!}$$

超几何函数 (hypergeometric function)

$${}_2F_1(a,b;c;z) = \sum_{n=0}^{\infty} \frac{a^{(n)} b^{(n)}}{c^{(n)}} \frac{z^n}{n!}$$

推广的超几何函数 (generalized hypergeometric function)

$${}_pF_q(a_1,a_2,\cdots,a_p;b_1,b_2,\cdots,b_q;z) = \sum_{n=0}^{\infty} \frac{a_1^{(n)} a_2^{(n)} \cdots a_p^{(n)}}{b_1^{(n)} b_2^{(n)} \cdots b_q^{(n)}} \frac{z^n}{n!}$$