呼吁人大将宪法第一条“中华人民共和国是工人阶级领导的、以工农联盟为基础的、人民民主专政的社会主义国家。”修改为“中国是全体公民组成的，致力于保障公民基本权利，为公民的生存与发展提供良好公共环境的法治国家。中国全体公民的基本权利一律平等。中国不设国教。” —— **同一个世界，同一个反特权反垄断的梦想！**

初中后数学物理教呈(V110604)

丁建华（手机: 158-3818-9433）

本文各章曾先后在水木社区 (newsmth.net) 数学版征求意见，网友 dragonheart6、IQ60、Journeyer、Insomnia、yuxiboh、SCIbird、snoopyzhao、Emmons 、zc、hunterkiller、canper、Cracker、DrFaust、Adiascem、daishu59、shangzh87、primenumber、dnd、liyafan、vcpshine、freephoenix、Hecke、shallpion、maming、LosHeaven、temporary、Liapunov、nimi、chobitss、upp 等或指出了一些错误、或提出了很好的建议，或表达了热情的支持，或其中二者或三者皆有。本文作者在此向他们表示衷心的感谢！

本文作者学识有限，文中错漏之处很可能还有，欢迎大家继续批评指正。

Contents

Chapter 1

引子

小娟、小明和小亮初中毕业了，在菜市场找到正蹲在三轮车上卖菜的楚忠厚，要报考他的“研究生”，攻读数学物理专业的“初中后”学位。楚老师说：“好吧，让我们从积分 (integral)讲起。”

“这是我今天卖菜的钱。”楚老师打开钱包，一边数一边说：“两张十块的，三张二十的，一张五十的，一共是一百三。这就是一种积分。”

三个代表重要学生面面相觑，一头雾水。小娟说：“拜托，我们可是认真的。”

“这就是认真的牙。”楚老师顺手挑出几个西红柿，说：“你们看，一、二、三、四、五，一共五个。这也是一种积分。”

小亮拉着大家就走。小明说：“等一下。楚老师在伽利顿大学流学多年，不可能只是会数数而已。我们还是听听他怎么解释吧。”

Chapter 2

集合论

2.1 集合

2.1.1 元素、集合与子集

楚老师微笑着点点头，说："从集合论的角度讲，这五个西红柿可以看做是五个元素 (element)。不妨分别记为 a, b, c, d, e。把 a, b 放一块，就构成一个集合 (set)，不妨记为 $F = \{a, b\}$。称元素 a, b 属于集合 F，记为 $a, b \in F$。有时也称集合 F 含有元素 a 和 b。

（鉴于本文体例特殊，继续上一段引述时，段首引号一律省略。）因为"革命不分先后"，所以 $\{b, a\}$ 和 $\{a, b\}$ 表示同一个集合。

谎言重复一千遍，就能够变成真理，但是真理重复一千遍，却仍旧是同一个真理。所以 $\{a, a, a, a, b\}$ 和 $\{a, b\}$ 表示同一个集合。

如果一个集合 A 中的所有元素均属于集合 B，则称集合 A 包含于集合 B，也称集合 A 是集合 B 的子集 (subset)，记为 $A \subseteq B$。例如 $\{b, d\} \subseteq \{b, c, d\}$。有时也称集合 B 包含集合 A。注意"包含"与"含有"是完全不同的两个概念。

如果集合 A 包含于集合 B，而且集合 B 也包含于集合 A，即 $A \subseteq B$ 而且 $B \subseteq A$，则称集合 A 和集合 B 相等，记为 $A = B$。例如 $\{b, c\} = \{c, b\}$。

如果集合 A 包含于集合 B，但是集合 B 不包含于集合 A，即 $A \subseteq B$ 而且 $A \neq B$，则称集合 A 真包含于集合 B，也称集合 A 是集合 B 的真子集 (proper subset)，记为 $A \subset B$。例如 $\{b, d\} \subset \{b, c, d\}$。

2.1.2 集合的交、差、并、补、异或、皆非、幂

集合 A 和集合 B 的交集 (intersection)，是指既属于 A，又属于 B 的元素所构成的集合，记为 $A \cap B$。例如 $\{a, b, c\} \cap \{b, c, d\} = \{b, c\}$。

集合 A 和集合 B 的差集 (subtraction, difference)，是指属于 A，但不属于 B 的元素所构成的集合，记为 $A \setminus B$。例如 $\{a,b,c\} \setminus \{b,d\} = \{a,c\}$。

集合 A 和集合 B 的并集 (union)，是指属于 A，或属于 B，（含既属于 A 又属于 B，）的元素所构成的集合，记为 $A \cup B$。例如 $\{a,b,c\} \cup \{b,c,d\} = \{a,b,c,d\}$。

所讨论的问题中的所有元素构成的集合称为全集 (universal set)，常记为 U。在我们当前的例子中，$U = \{a,b,c,d,e\}$。

不含任何元素的集合称为空集 (empty set)，记为 $\varnothing$。例如 $\{a,b\} \cap \{c,d\} = \varnothing$。

全集与集合 A 的差集，称为集合 A 的补集 (complement)，记为 $A^{\complement} = U \setminus A$。也常有记为 $\overline{A}$ 的。例如 $\{a,b\}^{\complement} = \{c,d,e\}$。

集合的交、差、并、补是最常用的四种运算。更复杂的运算常常可以用这四种运算来表达。举例来说，集合 A 和集合 B 的异或并集 (exclusive union)，也称对称差 (symmetric difference)，是指那些要么属于 A 但不属于 B，要么属于 B 但不属于 A 的元素所构成的集合，记为 $A \uplus B$。例如 $\{a,b,c\} \uplus \{b,c,d\} = \{a,d\}$。显然，$A \uplus B = (A \cup B) \setminus (A \cap B)$。相比之下，前面所说的并集 $A \cup B$，包含了既属于 A，又属于 B 的元素，因此又称“或并集 (inclusive union)”。

最常用并不意味着最基本。最基本的运算只有一个，称为集合 A 与集合 B 的皆非集，是指既不属于 A，又不属于 B 的元素所构成的集合，记为 $A \divideontimes B$。例如在 $U = \{a,b,c,d,e\}$ 的情况下，$\{a,b\} \divideontimes \{b,c\} = \{d,e\}$。

显然，$A^{\complement} = A \divideontimes A$，$A \cup B = (A \divideontimes B)^{\complement}$，$A \setminus B = A^{\complement} \divideontimes B$，$A \cap B = A^{\complement} \divideontimes B^{\complement}$。

以集合 A 的所有子集为元素的集合，称为集合 A 的幂集 (power set)，记为 2^A。例如 $2^{\{b,c,d\}} = \{\varnothing, \{b\}, \{c\}, \{d\}, \{b,c\}, \{c,d\}, \{b,d\}, \{b,c,d\}\}$。

2.1.3 列表

列表 (list) 不同于集合。它也是由元素组成，但是顺序和重复是有意义的。例如 (a,b)、(b,a) 和 (a,a,b) 是三个不同的列表。注意，在这样的表达式中，集合用花括号，而列表一般用圆括号，偶尔也有用方括号或者尖括号的。

列表的长度是指列表中元素的个数。例如列表 (a,a,b) 的长度是 3。

长度为 2 的列表也称二元组，或者序偶 (ordered pair)。长度为 3 的列表也称三元组。依此类推，长度为 n 的列表也称 n 元组 (n-tuple)。

2.1.4 笛卡尔积与关系

集合 A 和集合 B 的笛卡尔积 (Cartesian product)，也称直积 (direct product)，是指第一个元素属于 A，第二个元素属于 B 的所有序偶所构成的集合，

记为 $A\times B$。例如 $\{a,b,c\}\times\{\alpha,\beta\}=\{(a,\alpha),(a,\beta),(b,\alpha),(b,\beta),(c,\alpha),(c,\beta)\}$。

两个集合的笛卡尔积的子集，称为这两个集合之间的关系 (relation)，也称二元关系。例如 $\{(a,\alpha),(a,\beta),(b,\alpha)\}$ 是 $\{a,b,c\}$ 与 $\{\alpha,\beta\}$ 之间的一个二元关系。

n 个集合 $A_1,A_2,\ldots,A_n$ 的笛卡尔积，是指第一个元素属于 A_1，第二个元素属于 A_2，依此类推，第 n 个元素属于 A_n 的所有 n 元组所构成的集合，记为 $A_1\times A_2\times\cdots\times A_n=\prod_{i=1}^{n}A_i$。它的子集，称为这 n 个集合之间的 n 元关系。

如果上述 n 个集合都相同，是同一个集合 A，那么上述笛卡尔积常简记为 A^n。

2.1.5 映射

设 $1<k<n$。如果上述 n 元关系中的所有 n 元组的前 k 个元素组成的 k 元组没有重复的，则称该关系为从 $X=A_1\times\cdots\times A_k$ 到 $Y=A_{k+1}\times\cdots\times A_n$ 的映射 (mapping)，或者函数 (function)，记为 $f\colon X\to Y$。X 称为定义空间，Y 称为值空间。前述的那些 k 元组构成的集合称为该映射的定义域 (domain)，记为 $\operatorname{Dom} f$。该 n 元关系中的所有 n 元组的后 $n-k$ 个元素组成的 $n-k$ 元组构成的集合称为该映射的值域 (range)，记为 $\operatorname{Ran} f$。显然，$\operatorname{Dom} f\subseteq X,\quad \operatorname{Ran} f\subseteq Y$。

例如 $\{(a,\alpha),(a,\beta),(b,\alpha)\}$ 不是从 $\{a,b,c\}$ 到 $\{\alpha,\beta\}$ 的映射，因为 a 有重复。而 $\{(a,\alpha),(b,\alpha)\}$ 则是，记为 $g\colon\{a,b,c\}\to\{\alpha,\beta\}$。显有 $\operatorname{Dom} g=\{a,b\}$，$\operatorname{Ran} g=\{\alpha\}$。我们也说 $g(a)=\alpha,\quad g(b)=\alpha$，而 $g(c)$ 没有定义。

2.2 关系型数据库

2.2.1 表、属性与记录

我建了一个网站，让农户把他们种的菜的地头价报给我。今天的报价是：

日期	品类	价格	数量	姓名	电话
20xx.4.1	西红柿	0.90/kg	100kg	张三	150xxxx3366
20xx.4.1	黄瓜	0.70/kg	200kg	张三	150xxxx3366
20xx.4.1	西红柿	0.85/kg	400kg	李四	138xxxx7788

用关系型数据库技术的术语来讲，这是一个表 (table)，反映了价格、数量与其他因素之间的关系。表中每一列称为一个属性 (attribute)，每一行称为一个实体 (entity)，也称记录 (record)，实际上就是一个 n 元组。关系是 n 元组的集合，表是记录的集合，两套术语的实质含义是一样的。

2.2.2 函数依赖

如果把日期、品类和姓名选为 A_1, A_2, A_3，把价格、数量和电话选为 A_4, A_5, A_6，那么这个关系是一个从 $A_1 \times A_2 \times A_3$ 到 $A_4 \times A_5 \times A_6$ 的映射。

如果网站不大，用户不多，我们可以只考虑简化的情形，即合理假设用户不重名，而且只拥有一部电话。那么，从姓名 A_3 到电话 A_6 也是一个映射。

这样的映射称为函数依赖 (functional dependency)。之所以这样叫，是因为映射也称函数。上例中，电话函数依赖于姓名，而报价表中所有属性，都函数依赖于日期、品类和姓名。

2.2.3 袋子

理论上，表是记录的集合，因此表中不应有重复的记录。然而实践上，几乎所有的数据库系统都允许用户在创建表的时候指定该表内是否可以有重复记录。如果可以有重复记录，那么这个表就不是记录的集合，而是记录的袋子 (bag)。

袋子既不同于集合，也不同于列表。在袋子中，重复有意义，而顺序没有意义。

如果一个表是袋子，那么数据库系统会自动为该表添加一个属性，称为记录标识码 RID。当用户向表中添加记录时，数据库系统自动为每条新添记录生成不重复的 RID。这样，该表就成了扩展后的记录的集合。而且，从 RID 到其他任何属性都是映射。

2.2.4 键与主键

如果一个表中的所有属性都函数依赖于一组属性，则称这组属性构成的集合是该表的一个键 (key)。如果一个键失去其中的任何一个属性都不再是键，则称这个键是主键 (primary key)。上例中，{ 日期、品类、姓名 } 是报价表的一个主键。

如果一个表是袋子，那么 { RID } 是该表的一个主键。

2.2.5 方案设计

注意到，上例中张三和他的电话出现了两次。这种不必要的重复称为冗余 (redundency)。冗余占用存储空间、增加数据出错的可能性、给数据的增删改造成困难，还会引起其他麻烦。所以，在设计数据库时，要仔细分析用户业务数据的特性，发掘并充分利用函数依赖，尽量消除冗余。

上例中，最好增加用户标识码 UID 属性，并把数据库改成两个表：

1. 用户表 TU。主键是 UID。

UID	姓名	电话
007	张三	150xxxx3366
009	李四	138xxxx7788

2. 报价表 TQ。主键是日期、品类、UID。

日期	品类	价格	数量	UID
20xx.4.1	西红柿	0.90/kg	100kg	007
20xx.4.1	黄瓜	0.70/kg	200kg	007
20xx.4.1	西红柿	0.85/kg	400kg	009

2.2.6 查询

可以用如下查询命令得到原来形式的报价信息：

select TQ.日期, TQ.品类, TQ.价格, TQ.数量, TU.姓名, TU.电话
from TQ, TU where TQ.UID = TU.UID ;

数据库系统首先根据 from 子句的内容，做一个报价表 TQ 和用户表 TU 的笛卡尔积 TQ×TU：

TQ.日期	TQ.品类	TQ.价格	TQ.数量	TQ.UID	TU.UID	TU.姓名	TU.电话
20xx.4.1	西红柿	0.90/kg	100kg	007	007	张三	150xxxx3366
20xx.4.1	黄瓜	0.70/kg	200kg	007	007	张三	150xxxx3366
20xx.4.1	西红柿	0.85/kg	400kg	009	007	张三	150xxxx3366
20xx.4.1	西红柿	0.90/kg	100kg	007	009	李四	138xxxx7788
20xx.4.1	黄瓜	0.70/kg	200kg	007	009	李四	138xxxx7788
20xx.4.1	西红柿	0.85/kg	400kg	009	009	李四	138xxxx7788

然后再根据 where 子句的内容，利用筛选条件 TQ.UID = TU.UID 选出适当的记录：

TQ.日期	TQ.品类	TQ.价格	TQ.数量	TQ.UID	TU.UID	TU.姓名	TU.电话
20xx.4.1	西红柿	0.90/kg	100kg	007	007	张三	150xxxx3366
20xx.4.1	黄瓜	0.70/kg	200kg	007	007	张三	150xxxx3366
20xx.4.1	西红柿	0.85/kg	400kg	009	009	李四	138xxxx7788

最后根据 select 子句的内容，选出适当的列，返回查询的结果集 (result set)。

2.2.7 实体/关系图

很明显，数据库中的表与用户业务中的概念有密切的联系。“一个用户有一个与他人不同的标识码，一个姓名和一个电话”，反映到数据库中，就是“用户表 TU 有 UID、姓名和电话三列，其中 {UID} 是主键”。这种不涉及具体数据的、对表的结构和整体特性所做的规划称为表的方案 (schema)。做这种规划时，常把业务概念称为实体 (entity)。例如用户和报价就是两个实体。一个用户可以提供多个蔬菜品类的报价，这被称为用户实体到报价实体之间的

一对多关系 (relationship)。实践上，常把用户业务系统中的实体和它们之间的关系画成实体/关系图，简称 E/R 图，以方便设计数据库。注意，我们前面也介绍过实体和关系的概念，但是此实体非彼实体，此关系非彼关系。一词有多义，具体指什么，要看上下文。

两个实体之间若是一对一关系，则可以把属性合并，用一个表来实现；若是一对多关系，则通常各用一个表来实现，并把“一”方的主键加到“多”方的表中；若是多对多关系，则除了各用一个表来实现实体之外，还要另做一个表，包含双方的主键，以反映这种多对多关系。例如，要设计一个关于生态环境的数据库，“动物”可以被选作实体，捕食与被捕食是动物与动物之间的多对多关系。这个多对多关系可以用一个表来实现：

捕食者	被捕食者
蝮蛇	青蛙
蝮蛇	田鼠
蛇鹫	蝮蛇
眼镜王蛇	蝮蛇

2.3 映射

2.3.1 全映射、部分映射、单射、满射、双射

继续讲数学。上述映射 $f\colon X \to Y$，如果 $\mathrm{Dom}\, f = X$，则称 f 为全映射 (entire mapping)，或全函数。否则，称 f 为部分映射 (partial mapping)，或部分函数。上例的映射 g，因为 $g(c)$ 没有定义，所以是个部分映射。

如果上述映射 $f\colon X \to Y$ 中，那些 $n-k$ 元组也没有重复的，则称该映射是“一到一 (one-to-one)”的，或称该映射为单射 (injective)。

如果上述映射 $f\colon X \to Y$ 是单射，那么将上述 k 元组与 $n-k$ 元组交换位置，就做成了一个从 Y 到 X 的映射，称为 f 的逆 (inverse)，或者反，记为 $f^{-1}\colon Y \to X$。

如果 $\mathrm{Ran}\, f = Y$，则称该映射是“到上 (onto)”的，或称该映射为满射 (surjective)。

既单且满的映射称为双射 (bijective)。过去也称一一 (one-one) 映射，或一对一映射，因为容易与一到一映射（即单射）混淆，所以逐渐废弃了。

因为双射是单的，所以存在逆。显然，双射的逆是既全且单的。

既全且双的映射称为全双射。显然，全双射的逆是既全且双的。

前例的映射 g，因为 α 有重复，所以不是单射；因为 $\mathrm{Ran}\, g \neq \{\alpha, \beta\}$，所以也不是满射。而 $\{(a,\beta),(b,\alpha)\}$ 是另一个从 $\{a,b,c\}$ 到 $\{\alpha,\beta\}$ 的映射。它既单且满，因而是一个双射。但它不是全映射。

2.3.2 等势

如果两个集合之间可以构做一个全双射，则称这两个集合等势 (equipotent)。显然，如果 $f\colon X \to Y$ 是单射，那么 $\mathrm{Dom}\, f$ 与 $\mathrm{Ran}\, f$ 等势；如果 f 是双射，那么 $\mathrm{Dom}\, f$ 与 Y 等势；如果 f 是全双射，那么 X 与 Y 等势。

普通的集合论教科书都说双射意味着等势。这其实是采取了一种态度：不把部分映射看做是映射。所以，凡谈到的映射都是全映射。但是他们又不明说这一点，导致学生在遇到部分映射的时候总是很困惑。我们从一开始就引入部分映射和全映射的概念，付出不多，却避免了很多麻烦，算是教学上的一个创新。

2.3.3 运算

形如 $f\colon A \to A$ 的映射也称集合 A 上的单目运算，或者一元运算。形如 $f\colon A \times A \to A$ 的映射也称集合 A 上的双目运算，或者二元运算。依此类推，形如 $f\colon A^n \to A$ 的映射也称集合 A 上的 n 目运算，或者 n 元运算。此时，映射的"名字" f 称为运算符 (operator)，A^n 中的 n 元组中的元素称为操作数 (operand)。操作数的位置是有意义的，例如 $a - b \neq b - a$。

一元运算的写法最多样。运算符可以写在操作数的前面，比如 $-n$；后面，比如 $n!$；上面，比如 $\overline{z}$；右上角，比如 $A^{\complement}$。实际上还有更多。

二元运算的写法通常是把运算符写在两个操作数中间。比如 $a + b$。但是也有例外，比如 $\log_a b$。甚至用两个操作数的字体大小和相对位置来表示特定运算，而无需操作符，比如 a^b。

三元运算的一个实例是 C 语言中的 $a ?\, b : c$，表示：如果 $a \neq 0$，那么值为 b，否则为 c。

集合的交、差、并是 2^U 上的双目运算，集合的补是 2^U 上的单目运算。

2.3.4 "无中生有"的代表集序列 (S_n)

现在，我们用递归的 (recursive) 方法构做一个集合的序列 (sequence)：

$$\begin{cases} S_0 = \varnothing, \\ S_k = \{S_0, S_1, \ldots, S_{k-1}\}, \quad (k > 0). \end{cases}$$

该序列的前几项为

$$\begin{aligned} S_0 &= \varnothing, \\ S_1 &= \{S_0\} = \{\varnothing\}, \\ S_2 &= \{S_0, S_1\} = \{\varnothing, \{\varnothing\}\}, \\ S_3 &= \{S_0, S_1, S_2\} = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}. \end{aligned}$$

设两个集合 S_m 和 S_n 取自该序列。如果 S_m 和 S_n 都不空，那么一定都含有 S_0。如果 $S_m \setminus \{S_0\} = S_m \setminus S_1$ 和 $S_n \setminus \{S_0\} = S_n \setminus S_1$ 都不空，那么一定都含有 S_1。如果 $S_m \setminus \{S_0, S_1\} = S_m \setminus S_2$ 和 $S_n \setminus \{S_0, S_1\} = S_n \setminus S_2$ 都不空，那么一定都含有 S_2。依此类推，最终只可能有三种结果：

1. $S_m \setminus S_k = \varnothing,\quad S_n \setminus S_k \neq \varnothing$。此时 $S_m \subset S_n$；
2. $S_m \setminus S_k = \varnothing,\quad S_n \setminus S_k = \varnothing$。此时 $S_m = S_n$；
3. $S_m \setminus S_k \neq \varnothing,\quad S_n \setminus S_k = \varnothing$。此时 $S_m \supset S_n$。

实际上，对序列中的任何一个集合 S_i，从中一次取走一个元素，必可在有限次操作后使其成为空集。也就是说，S_i 中元素的个数是有限的。

2.3.5 鸽笼原理

集合 A 如果与某个 S_i 等势，则称为有限集 (finite set)。

有限集与自己的真子集不等势。这个称为鸽笼原理。如果鸽子比笼子多，那么无论怎样安排，总有笼子装了多只鸽子。也就是说，鸽子的集合与笼子的集合之间，不可能构做一个全双射。

上述 S_m 和 S_n，第一，两者都是有限集；第二，两者要么相等，要么一个是另一个的真子集。所以，上述序列中，任何两个相异集合不等势。

2.4 基数

2.4.1 皮亚诺公理

自然数集 $\mathbf{N} = \{0, 1, 2, \ldots\}$ 的基本性质可以由五条皮亚诺公理 (Peano axioms) 来刻画：

1. 存在一个特殊的自然数 0；
2. 对任何自然数 n，都存在另一个自然数 $n' \neq n$，称为 n 的后继；
3. 0 不是任何自然数的后继；
4. 如果两个自然数的后继相等，那么这两个自然数也相等；
5. 一个自然数的集合 A，如果含有 0，而且对任何 $n \in A$，均有 $n' \in A$，那么 $A = \mathbf{N}$。

记 $\mathbf{S} = \{S_0, S_1, S_2, \ldots\}$，并称其中的元素为代表集；那么 $\phi\colon \mathbf{S} \to \mathbf{N}$，对所有的 $i \in \mathbf{N}$，有 $\phi(S_i) = i$，是一个全双射。把皮亚诺公理中的自然数换成代表集，0 换成 S_0，$\mathbf{N}$ 换成 $\mathbf{S}$，不难发现，$\mathbf{S}$ 与 $\mathbf{N}$ 具有完全相同的代数性质，不同的仅仅是名词解释而已。

2.4.2 基数

集合 A 的势 (potency)，也称基数 (cardinal number, cardinality)，记为 $|A|$。

显然，说 $|S_n| = n$ 是合理的。这就是前面构做的全双射 ϕ: $\mathbf{S} \to \mathbf{N}$，只是换了一种写法而已。实际上，$n$ 只是一个符号，是“该集合与 S_n 等势”的缩写。对那些学认数字的幼儿来说，2 就是“与两个手指等势”，3 就是“与三个手指等势”，依此类推。所以，$\mathbf{N}$ 的一切性质，本质上都是 $\mathbf{S}$ 的性质。

集合 A 如果是有限集，则必与某个 S_n 等势，从而 $|A| = |S_n| = n$。

集合 A 如果不是有限集，则称为无限集 (infinite set)。

因为 S_{n+1} 总可做，所以 $\mathbf{S}$ 不与任何 S_n 等势，因而不是有限集，是无限集。记 $|\mathbf{S}| = \aleph_0$。$\aleph$ 是希伯来语 (Hebrew) 字母表中的第一个字母，读作“阿里夫”。

因为存在上述全双射 ϕ: $\mathbf{S} \to \mathbf{N}$，所以 $|\mathbf{N}| = |\mathbf{S}| = \aleph_0$。

2.4.3 可数集与可数无穷大

集合 A 如果与 $\mathbf{N}$ 等势，则称为可数集 (countable set)。$\aleph_0$ 也称可数无穷大 (countable infinity)。

任何无限集都包含一个可数的子集。从这个意义上讲，可数无穷大是最小的无穷大。给定一个无限集 A，从 A 中可取一个元素 a_0，从 $A \setminus \{a_0\}$ 中可取一个元素 a_1，从 $A \setminus \{a_0, a_1\}$ 中可取一个元素 a_2，如此反复，因为 A 不是有限集，所以这种操作总可进行。所取出的元素构成的集合 $C = \{a_0, a_1, \ldots\}$ 就是 A 的可数的子集。 $\square$

2.4.4 无限集

无限集可与它的真子集等势。设集合 A 是无限集，集合 $C = \{a_0, a_1, \ldots\}$ 是如前所述做成的 A 的可数的子集。那么 $A' = A \setminus \{a_0\}$ 是 A 的真子集。我们这样构做映射 ϕ: $A \to A'$：对任何一个元素 $a \in A$，

- 如果 $a \notin C$，则定义 $\phi(a) = a$；
- 如果 $a \in C$，即 a 等于某个 a_i，则定义 $\phi(a) = \phi(a_i) = a_{i+1}$。

因为 C 是可数集，对任何 a_i，相应的 a_{i+1} 总存在，所以，易见该映射 ϕ 是全双射。即 A 与 A' 等势。 $\square$

例如，$\mathbf{N} = \{0, 1, 2, \ldots\}$ 与 $\mathbf{N} \setminus \{0, 1, \ldots, k-1\} = \{k, k+1, \ldots\}$ 等势，可做的全双射为 $\phi(i) = k + i$。$\mathbf{N} = \{0, 1, 2, \ldots\}$ 与偶数集 $\mathbf{E} = \{2, 4, 6, \ldots\}$ 等势，可做的全双射为 $\phi(i) = 2(i+1)$。

整数集，记为 $\mathbf{Z}$，是可数集。我们将所有整数这样列举出来：$0, -1, 1, -2, 2, -3, 3, \ldots$。让它们逐个对应 $0, 1, 2, 3, \ldots$。易见这是一个全双射。 $\square$

正有理数集，记为 $\mathbf{Q}^+$，是可数集。考虑一个无限大的表格，第 n 行第 m 列填有理数 m/n。我们可以用“卷角”的办法，把该表格的内容列举出来：$1/1, 2/1, 1/2, 3/1, 2/2, 1/3, \ldots$。让它们逐个对应 $0, 1, 2, 3, \ldots$。易见这是一个全双射。因此，该表格的格子是可数的。所有正有理数都在这个表格中，所以，$|\mathbf{Q}^+| \leq \aleph_0$。而 $\mathbf{N} \subset \mathbf{Q}^+$，所以，$|\mathbf{Q}^+| \geq \aleph_0$。因此，$|\mathbf{Q}^+| = \aleph_0$。 $\square$

有理数集，记为 $\mathbf{Q}$，是可数集。考虑两个无限大的表格，第一个表格的第 n 行第 m 列填有理数 $-m/n$，第二个表格的第 n 行第 m 列填有理数 m/n。我们这样列举 0 和这两个表格的内容：$0, -1/1, 1/1, -2/1, 2/1, -1/2, 1/2, \ldots$。让它们逐个对应 $0, 1, 2, 3, \ldots$。易见这是一个全双射。之后的推理与上段相同。 $\square$

因为任何有限集都不与自己的真子集等势，所以，如果一个集合与自己的某个真子集等势，就一定不是有限集，因此是无限集。

2.4.5 康托幂集定理

康托幂集定理 (Cantor's Powerset Theorem)：任何集合都与自己的幂集不等势。

假设不然，存在一个集合 T 与它的幂集 2^T 等势，即存在一个全双射 $h\colon T \to 2^T$，一方面，对集合 T 的任何一个元素 $t \in T$，存在且只存在一个 T 的子集 $h(t) \subseteq T$；另一方面，对 T 的任何一个子集 $S \subseteq T$，存在且只存在一个 T 的元素 $h^{-1}(S) \in T$。那么，所有满足 $t \notin h(t)$ 的元素构成一个子集 P，进而，应该存在一个元素 $p = h^{-1}(P)$。问题是：这个元素 p 是否属于子集 $h(p)$ 呢？

- 如果说 $p \in h(p)$，注意到 $p = h^{-1}(P)$，所以 $P = h(p)$，而按照 P 的构做方法，其中的任何元素都有 $t \notin h(t)$，既然 $p \in P$，那么 p 也该具有此性质，即 $p \notin h(p)$。矛盾。
- 如果说 $p \notin h(p)$，注意到 P 包含了所有满足 $t \notin h(t)$ 的元素，既然 $p \notin h(p)$，那就应该有 $p \in P$。而 $P = h(p)$，所以 $p \in h(p)$。也矛盾。

对于 p 是否属于 $h(p)$ 的问题，无论怎么回答都将导致矛盾。所以，我们只能放弃前面所做的假设，而不得不承认，与自己的幂集等势的集合不存在。 $\square$

因为在做子集的时候，对每个元素都有“取”和“不取”两种选择，所以 $|2^A| = 2^{|A|}$。从而，任何集合都与自己的幂集不等势就可以表示为 $2^{|A|} \neq |A|$。

对有限集 A，有 $|A| = n \in \mathbf{N}$，而对任何 $n \in \mathbf{N}$，$2^n \neq n$ 是熟知的事实。

对可数集 A，$|A| = \aleph_0$，$2^{|A|} \neq |A|$ 意味着 $2^{\aleph_0} \neq \aleph_0$。因此定义 $\aleph_1 = 2^{\aleph_0}$。依此类推，可定义 $\aleph_2 = 2^{\aleph_1}$，$\aleph_3 = 2^{\aleph_2}$ 等等一系列越来越大的无穷大基数。

2.4.6 基数的和、积、幂

给定任意两个集合 A 和 B，做 $\alpha\colon A \to (A \cup B) \times \{0,1\}$ 把 $a \in A$ 映射到 $(a,0)$，$\beta\colon B \to (A \cup B) \times \{0,1\}$ 把 $b \in B$ 映射到 $(b,1)$，那么 α 和 β 都是全单射，而且 $\operatorname{Ran}\alpha \cap \operatorname{Ran}\beta = \varnothing$，所以，定义集合 A 和 B 的基数的和为 $|A| + |B| = |\operatorname{Ran}\alpha \cup \operatorname{Ran}\beta|$。

定义集合 A 和 B 的基数的积为 $|A| \times |B| = |A \times B|$。

给定任意两个集合 A 和 B，所有全映射 $\theta\colon A \to B$ 构成的集合称为 B 的 A 次幂，记为 B^A。

$|\{0,1\}^A| = |2^A|$。构做一个映射 $\xi\colon \{0,1\}^A \to 2^A$：对任何一个全映射 $\theta \in \{0,1\}^A$，如果 $(a,0) \in \theta$，则不选 a；否则必定有 $(a,1) \in \theta$，故选 a。如此选出的元素构成 A 的一个子集 $T = \xi(\theta)$。显见该映射 ξ 是全双射。 □

给定任意两个集合 A 和 B，定义 $|B|$ 的 $|A|$ 次幂为 $|B|^{|A|} = |B^A|$。

2.4.7 连续统假说与连续无穷大

大于 0 且小于 1 的所有实数做成的集合记为 $\mathbf{I}$，有 $|\mathbf{I}| = \aleph_1$。因为这样的实数可以写成二进制小数的形式，位数是可数无穷大，每一位上有取 0 和取 1 两种选择。 □

$3^{\aleph_0} = \aleph_1$。因为任何 $t \in \mathbf{I}$ 均可写成三进制小数的形式，位数为可数无穷大，每一位上有取 0、1 和 2 三种选择。（思考题：证明 $\aleph_0^{\aleph_0} = \aleph_1$。） □

所有实数做成的集合称为实数集，记为 $\mathbf{R}$，有 $|\mathbf{R}| = \aleph_1$。因为任何实数 $x \in \mathbf{R}$ 都可写成二进制形式 $(x)_2 = \pm b_n b_{n-1} \cdots b_0 . b_{-1} b_{-2} \cdots$。我们将它按下述方法映射到 $t \in \mathbf{I}$：设 $(t)_2 = 0.c_{-1}c_{-2}\cdots$，

- 如果 $x > 0$，则 $c_{-1} = 0$，否则 $c_{-1} = 1$；
- $c_{-2} = b_0,\quad c_{-4} = b_1,\quad \ldots,\quad c_{-(2n+2)} = b_n$，以后偶数位上皆 0；
- $c_{-3} = b_{-1},\quad c_{-5} = b_{-2},\quad \ldots,\quad c_{-(2n+1)} = b_{-n},\quad \ldots$。

易见该映射是全双射。 □

无法合理地构做一个无限集，使它的势介于 $\aleph_0$ 和 $\aleph_1$ 之间。这被称为连续统假说 (Continuum Hypothesis)。它不能由其他的集合论公理推导出来，因此被认为是一个独立的公理。

连续统假说使我们确信，直线作为点的集合，它的势为 $\aleph_1$。因为我们可以在直线上建立坐标系，使其成为数轴，一个点对应一个实数，从而构做了一个全双射。因此，$\aleph_1$ 也称连续无穷大。

三维空间中的点与直线上的点一样多。三维空间中的点可以与实数的三元组 (x,y,z) 做成全双射。考虑 $0 < x < 1,\quad 0 < y < 1,\quad 0 < z < 1$ 的点构成

的集合，记为 $\mathbf{I}^3$。设 $t \in \mathbf{I}$ 的三进制形式为 $(t)_3 = 0.t_1t_2t_3t_4t_5t_6\cdots$，将它对应到 $(x)_3 = 0.t_1t_4\cdots$，　$(y)_3 = 0.t_2t_5\cdots$，　$(z)_3 = 0.t_3t_6\cdots$。易见这是一个全双射。所以，$\mathbf{I}^3$ 与 $\mathbf{I}$ 等势。而每个 $\mathbf{I}$ 都与 $\mathbf{R}$ 等势，所以三维空间作为点的集合，与直线作为点的集合，两者等势。这一事实可简写为 $\aleph_1 = 3 \times \aleph_1 = {\aleph_1}^3$。　□

2.5　罗素悖论

2.5.1　康托对角线法

关于 $\mathbf{I}$ 不可数，还有另一种证法。假设不然，$\mathbf{I}$ 是可数集，即我们可以将所有 $t \in \mathbf{I}$ 逐个列出：$t_0, t_1, t_2, \ldots$，从而构做一个全双射 $\phi(t_i) = i$。在实数的二进制形式中，某一位的非 (inverse) 是指，如果该位是 0，它的非就是 1；如果该位是 1，它的非就是 0。考虑这样一个实数 $s \in \mathbf{I}$，二进制形式为 $(s)_2 = 0.b_{-1}b_{-2}\ldots$，而 b_{-1} 是 t_0 对应的位的非，b_{-2} 是 t_1 对应的位的非，依此类推。那么，s 与任何 t_i 均不相等，但是它又的确是一个完全合法的属于 $\mathbf{I}$ 的实数。这与“所有 $t \in \mathbf{I}$ 均已列出”相矛盾。故 $\mathbf{I}$ 不可能是可数集。　□

这个证法称为康托对角线法 (diagonal argument)，其应用十分广泛。前述康托幂集定理的证明本质上也是这种方法。假设集合 A 与 2^A 等势，那么可以做一个“表格”，每一行对应一个子集 $S_i \in 2^A$，每一列对应一个元素 $a_j \in A$。如果 $S_i \ni a_j$，则第 i 行第 j 列填 $\surd$，否则就空着。考虑一个子集 $T \in 2^A$，它也对应着表格的一行，其中每格的内容这样确定：从表格的左上角向右下角看，如果表格的第 n 行第 n 列填有 $\surd$，则该行第 n 格空着，否则，就填 $\surd$。设这一行是表格中的第 m 行，那么按照 T 的构做方法，第 m 行第 m 列的内容将自己否定自己。因此，所做假设不能成立。注意，对该证明中所说的表格，行、列，应当做广义的理解。比如，如果 $|A| = \aleph_1$，那么该表格可以理解为一个正方形，第 i 行第 j 列的格子应当理解为正方形内部的坐标为 (x, y) 的点。　□

2.5.2　罗素悖论

所有集合的全体不是一个集合。假设不然，记该集合为 T，那么按定义 $T \in T$。考虑 T 中那些不含有自身的集合，它们的全体构成了 T 的一个真子集，记为 S。问题是：S 是否含有自身？（注意“含有 $(\ni)$”与“包含 $(\supseteq)$”的区别。任何集合 A 都包含自身，即 $A \supseteq A$，但未必含有自身。）

- 如果说 $S \in S$，而按照 S 的定义，S 中任何元素都不含有自身，所以 S 也应该有此性质，即 $S \notin S$。矛盾。
- 如果说 $S \notin S$，而按照 S 的定义，所有不含有自身的集合都在 S 中，所以 S 也应该在 S 中，即 $S \in S$。也矛盾。

所以，把 T 和 S 当作集合来处理是不可取的。 $\square$

"不含有自身"这种性质，称为罗素 (Russell) 性质。上述矛盾的情形，称为罗素悖论 (Russell's Paradox)。

2.5.3 类、内涵公理与外延公理

为解决罗素悖论，冯 • 诺依曼 (Von Neumann)、哥德尔 (Gödel) 和伯内斯 (Bernays) 引入了类的概念。类本身是无定义的，其特性完全由类与类之间的两种关系来刻画。

其一是成员关系。类 A 是类 B 的成员 (member)，记为 $A \in B$。出于习惯，也常说类 A 是类 B 的元素，说类 A 属于类 B。

给定一个类 A，存在一个关于类的单目谓词 (predicate) $P(x)$，定义为：如果类 $x \in A$，那么 $P(x)$ 为真；否则，为假。在实际问题中，这种谓词常常具有某种解释，称为对这个类的理解 (comprehension)，也称类的内涵，或性质 (property)。

给定一个关于类的单目谓词 $P(x)$，存在一个类 A，使得：类 A 的所有成员都使 $P(x)$ 为真，而且，所有使 $P(x)$ 为真并能够成为某个类的成员的类都是类 A 的成员。这被称为内涵公理 (Comprehension Axiom)。

其二是等同关系。类 A 和类 B 是等同的 (identical)，意味着 A 的任何成员都属于 B，而且 B 的任何成员也都属于 A。也称这两个类相等，记为 $A = B$。显然，对类的成员来说，重复和先后顺序都没有意义。

由既属于类 A 又属于类 B 的类所构成的类，称为类 A 和类 B 的交，记为 $A \cap B$。如果 $A \cap B = A$，则称类 A 包含于类 B，记为 $A \subseteq B$。也称 A 是 B 的子类。显然，如果 $A \subseteq B$ 而且 $B \subseteq A$，那么 $A = B$。

类所含有的成员的总体称为类的外延 (extensionality)。如果类 A 的内涵是 $P(x)$，那么类 A 的外延常被记做 $\{x : P(x)\}$。

外延相等的两个类等同。这被称为外延公理 (Axiom of Extensionality)。

类 A 的外延包含于类 B 的外延，当且仅当，类 A 的内涵 $P(x)$ 蕴涵类 B 的内涵 $Q(x)$（指凡使 $P(x)$ 为真的 x，都使 $Q(x)$ 为真）。这被称为外延与内涵的反比关系，可形式化地表示为 $\{x : P(x)\} \subseteq \{x : Q(x)\} \iff (P(x) \Rightarrow Q(x))$。

2.5.4 集合与本性类

现在，考虑罗素性质 $P(x)$，即 $x \notin x$。根据内涵公理，存在一个类 S，其所有成员都不含有自身，而且，所有不含有自身并能够成为某个类的成员的类都是类 S 的成员。这个类 S 称为罗素类。问：$P(S)$ 为真还是为假？设 $P(S)$ 为假，即 $S \in S$。可是，S 的所有成员都不含有自身，既然 S 也是 S 的成员之一，那么它也该具有这一性质，即 $S \notin S$。矛盾。改设 $P(S)$ 为真，即 $S \notin S$。

那么，根据内涵公理，所有不含有自身并能够成为某个类的成员的类都是 S 的成员。与以前不同的是，如果我们认定 S 不能成为任何类的成员。那么虽然 $S \notin S$ 成立，但是由于第二项条件不满足，因此不能推出 $S \in S$，从而不再导致矛盾。

如果一个类能够成为某个类的成员，则称为集合。否则，称为本性类，也称真类 (proper class)。代数学中有一个分支，叫范畴论 (category theory)，其中用到所有集合构成的类，所有群构成的类，等等，这些都是本性类。

2.5.5 全类与罗素类

内涵公理其实是说：对给定的关于类的单目谓词 $P(x)$，使它为真的所有集合总可构成一个类。

比如，恒真谓词 $T(x)$，无论 x 是什么，该谓词的值永远为真，常写成 $x = x$。与该谓词相关联的类，就是由所有集合构成的类 T，称为全类 (universal class)。

罗素性质 $x \notin x$ 也是一个谓词，与之相关联的类，就是由所有不含有自身的集合所构成的类 S，称为罗素类。由于罗素类本身不是集合，所以 $S \notin S$ 并不导致 $S \in S$。

集合的子类是集合。这是一条公理。因此，如果一个类的子类不是集合，那么这个类也不是集合。

罗素类 S 是全类 T 的子类，而且 S 不是集合，因此，T 也不是集合。

再次提醒列位看官，T 包含 S，即 $T \supseteq S$，但是 T 并不含有 S，即 $T \not\ni S$。实际上，S 是本性类，不可能被任何类所含有，包括不可能被 S 自己所含有。

引入类的概念，避免了罗素悖论，却产生了另一个问题：给定一个谓词 $P(x)$，如何判断与之关联的类是本性类，还是集合？把哥德尔不完全性定理 (Gödel's Incompleteness Theorem) 用到这个问题上，就是：不存在一劳永逸的办法。

Chapter 3

图灵机与递归函数论

不存在判断任意的非交互程序是否会最终运行结束的非交互程序。假设不然，记这个判断程序为 $P(x)$。利用 $P(x)$，可以编写一个程序 Q：如果 $P(Q)$ 判断 Q 会最终结束，就进入无限循环，永不结束；否则，就结束。问：Q 是否会最终结束？无论怎样回答都是矛盾。因此，程序 $P(x)$ 在逻辑上就不可能存在。 $\square$

3.1 图灵机

3.1.1 图灵机的结构与指令集

一个图灵机 (Turing Machine) 包括

- 一个可打印字符集 $\{S_0, S_1, \ldots\}$，也称字母表；
- 一个控制字符集 $\{R, L\}$；
- 一个状态集 $\{q_1, q_2, \ldots\}$；
- 一个指令集；
- 一个无限长列表，也称为带，和一个相配套的读写头。

其中 S_0 也常记做 B，表示空白；S_1 也常记做 1；R 表示读写头右移一位；L 表示读写头左移一位；读写头可以读出当前位置的可打印字符，以便确定下一步执行什么指令，也可以在当前位置写入指定的可打印字符；带上写有适当的可打印字符，而读写头处于适当的位置上；带的左右两端可以根据需要无限补充，补充的部分的初始内容总是 B。

图灵机的指令可分为四种类型：如果在状态 q_i 下读到字符 S_j，则

1. 指令 $q_iS_jS_kq_l$ 表示在当前位置上写入 S_k 并进入状态 q_l；
2. 指令 $q_iS_jRq_l$ 表示右移一位并进入状态 q_l；
3. 指令 $q_iS_jLq_l$ 表示左移一位并进入状态 q_l；
4. 指令 $q_iS_jq_kq_l$ 需要多些解释。如果一个图灵机含有这种指令，则需要指定一个自然数的集合 $A \subseteq \mathbf{N}$，称为该图灵机的相对集。这种指令也因此被称为是相对的。记当前带上字符 1 的个数为 n。该指令表示，如果 $n \in A$，则进入状态 q_k；否则，进入状态 q_l。如果一个图灵机的指令集不含有这种类型的指令，则称该图灵机是简单的。

图灵机的指令集里，没有两条指令的 q_i 和 S_j 都相同。这保证了在任何情况下，图灵机最多只有一种应执行的动作。当图灵机在状态 q_i 下读到了 S_j 时，如果指令集里没有针对 q_iS_j 的指令，则称该情形是最后的 (final)；如果指令集里即便有针对 q_iS_j 的指令，也都是相对的，则称该情形是终结的 (terminate)。对简单图灵机来说，最后的和终结的两者等价。

3.1.2 A 计算

我们用 $Pq_iS_jQ(Z/A)$ 来刻画指定了相对集 A 的图灵机 Z，在状态 q_i 下读到当前字符为 S_j，而带上读写头左边的字符串为 P，右边的字符串为 Q 时的情景，称为瞬时描述 (instant description)。当 Z 和 A 给定时，该瞬时描述常简记为 Pq_iS_jQ。

设 α 是一个瞬时描述，该图灵机执行一条指令后，新的瞬时描述为 β，我们称此过程为一步执行，或一步推导，记为 $\alpha \to \beta(Z/A)$。当 Z 和 A 给定时，该步推导常简记为 $\alpha \to \beta$。

如果一个瞬时描述 α_1，经有限步执行，所得到的瞬时描述 α_p是最后的，则称此过程为该图灵机的一次 A 计算，称 α_p 是这次计算的结果 (result)，记为 $\alpha_p = \mathrm{Res}_Z^A(\alpha_1)$。如果 α_1 不能在有限步内成为最后的，即该图灵机将永不停歇地运行下去，则称 $\mathrm{Res}_Z^A(\alpha_1)$ 无定义。

3.1.3 计算多元函数 $f(m_1, \ldots, m_k)$ 的值

为了用图灵机来计算自然数集 $\mathbf{N}$ 上的多元函数 $f(m_1, m_2, \ldots, m_k)$ 的值，我们安排初始的瞬时描述为 $\alpha_1 = q_1 1^{m_1+1} B 1^{m_2+1} B \cdots B 1^{m_k+1}(Z/A)$。用 $\langle\alpha\rangle$ 表示在瞬时描述 α 中，带上的字符 1 的个数。那么我们取 $\langle \mathrm{Res}_Z^A(\alpha_1)\rangle$ 为该多元函数在给定点上的值。如果 $\mathrm{Res}_Z^A(\alpha_1)$ 无定义，则称该多元函数在给定点上无定义。如果存在这样的点，则称该函数是部分函数，否则，称全函数。

3.1.4 一些常见函数的图灵机实现

后继函数

例如，后继函数 $S(x)=x+1$ 可以用只含有一条指令 q_1BBq_1 的图灵机来实现。实际上该图灵机根本不执行任何指令，因为初始的瞬时描述就是最后的！

清零函数

清零函数 $N(x)=0$。用图灵机实现这个函数，只需要两条指令

q_11Bq_2	; 如果已经为零则停机，否则清零当前位
q_2BRq_1	; 跳过已被清零的位置，重复上述过程

我们来看 $N(1)=0$ 的计算过程：

$$\begin{aligned}\alpha_1 &= q_111\\ &\rightarrow q_2B1\\ &\rightarrow Bq_11\\ &\rightarrow Bq_2B\\ &\rightarrow BBq_1B \text{（终结）}\end{aligned}$$

加法

加法 $f(x,y)=x+y$。似乎只要从初始位置连删两个 1 就可以了，问题是 x 可以取零，在初始位置上没有连续的两个 1 让你删。所以，我们只能删一个，然后找到下一个变量的位置，再删一个。该图灵机的指令集为：

q_11Bq_1	q_1BRq_2	; 删除一个 1 并跳过它
q_21Rq_2	q_2BRq_3	; 找到下一个变量并就位
q_31Bq_3		; 删除一个 1 并终结

我们跟踪一下 $1+2=3$ 的计算过程：

$$\begin{aligned}\alpha_1 &= q_111B111\\ &\rightarrow q_1B1B111\\ &\rightarrow Bq_21B111\\ &\rightarrow B1q_2B111\\ &\rightarrow B1Bq_3111\\ &\rightarrow B1Bq_3B11 \text{（终结）}\end{aligned}$$

减法

减法 $f(x,y) = x - y$。我们已经掌握了在带上寻找特定位置的技术，现在要做的，就是找到 x，删掉左端的 1，找到 y，删掉右端的 1，再找到 x，重复这一过程。下面的图灵机实现一个部分函数：当 $x < y$ 时，该函数无定义。

$q_1 1 B q_1$	$q_1 B R q_2$	; 删除 x 左端的一个 1 并跳过它
$q_2 1 R q_2$	$q_2 B R q_3$	; 找到 y 并就位
$q_3 1 R q_3$	$q_3 B L q_4$	; 删除 y 右端的一个 1
$q_4 1 B q_4$	$q_4 B L q_5$	;
$q_5 1 L q_6$		; 如果 y 已被删完，就停机，给出结果 $x - y$
$q_6 1 L q_6$	$q_6 B L q_7$	;　否则就寻找 x
$q_7 1 L q_8$	$q_7 B R q_9$	; 如果 x 还在，
$q_8 1 L q_8$	$q_8 B R q_1$	;　就找到其左端，并重复上述过程
$q_9 1 L q_9$	$q_9 B R q_9$	;　否则就在 y 的左端无休止地徘徊

真减

真减 $f(x,y) = x \ominus y$，当 $x \geq y$ 时，值为$x - y$；否则，为 0。只要把清零函数的程序“嫁接”到上述减法程序的适当位置即可。具体来说，裁掉上述减法程序最后一行，加上如下两行：

	$q_9 B R q_{10}$	; 把剩余的 y 清零
$q_{10} 1 B q_9$		;

设 $f(x)$ 是一个部分函数，而 $g(x)$ 定义为，当 $f(x)$ 有定义时，值为 $f(x)$，否则为 0，这样的 $g(x)$ 称为 $f(x)$ 的补全函数。例如，真减是减法的补全函数。

以上图灵机程序都来自《Computability and Unsolvability by Martin Davis (McGraw-Hill, 1958)》，有中译本《可计算性与不可解性（北京大学出版社，1984）》。书中还给出了实现恒等函数 $U_i^n(x_1, \ldots, x_n) = x_i$ 和乘法 $f(x,y) = xy$ 的图灵机程序。

自然数的集合 A 的特征函数 $C_A(x)$

自然数的集合 A 的特征函数 $C_A(x)$，当 $x \in A$ 时，值为 0；否则为 1。也就是说，集合 A 是方程 $C_A(x) = 0$ 的解集。

函数 $C_A(x)$ 可以用指定了相对集为 A 的下述图灵机 Z 来实现：

$q_1 1 B q_1$	$q_1 B q_2 q_3$	; 删除一个 1，判断 $x \in A$ 是否为真
	$q_2 B R q_4$	; 为真，则清零后停机
$q_4 1 B q_2$		;
	$q_3 B R q_5$	; 否则，也清零
$q_5 1 B q_3$	$q_5 B 1 q_3$	;　但是清零后再写个 1，才停机

A 的特征函数 $C_A(x)$，后继函数 $S(x)=x+1$，恒等函数 $U_i^n(x_1,\ldots,x_n)=x_i,\quad(1\le i\le n)$，加法 $x+y$，真减 $x\ominus y$，乘法 xy，称为初等递归函数。注意，它们都是全函数。

3.1.5 n 正则的图灵机与函数的复合

给定一个计算 n 元函数的图灵机 Z，记其所有状态的最大下标为 $\theta(Z)$。如果 Z 的指令集不含有针对 $q_{\theta(Z)}$ 的指令，Z 的任何的计算结果都已经被整理为 s 元函数的初始瞬时描述，除了最后的状态都是 $q_{\theta(Z)}$ 而非 q_1 之外，则称该图灵机 Z 是 n 正则的。称它所计算的函数是 n 元 s 维向量函数 $f^{(s)}(x_1,\ldots,x_n)$。

设 Z_1 是 n 正则的，计算 n 元 s 维向量函数 $f^{(s)}(x_1,\ldots,x_n)$，Z_2 计算 s 元函数 $g(x_1,\ldots,x_s)$，则把 Z_2 的所有状态的下标都加 $\theta(Z_1)$，把修改后的状态集和指令集与 Z_1 的合并，做出的新图灵机，计算复合函数 $h=g\circ f$。

实际上，Davis 在那本书中提供了一些“胶水”程序，最终证明：任何可用图灵机实现的函数，它们的复合也可用图灵机实现。

3.1.6 对全函数的取极小运算

给定一个可用图灵机实现的全函数 $f(y,x_1,\ldots,x_n)$，可以做出一个图灵机 Z，逐一计算 $f(0,x_1,\ldots,x_n)$，$f(1,x_1,\ldots,x_n)$，等等，如果算出 $f(i,x_1,\ldots,x_n)=0$，就给出值 i 并停机；否则，就一直算下去。这个图灵机实现了一个函数 $h(x_1,\ldots,x_n)$，称为对全函数 $f(y,x_1,\ldots,x_n)$ 的取极小运算，记为 $\min_y[f(y,x_1,\ldots,x_n)=0]$。

例如，$x/2=\min_y[x\ominus(y+y)=0]$，有 $0/2=0$，$2/2=1$，但是 $1/2$ 无定义。

虽然取极小运算是作用于全函数上的，但是取极小得到的函数常常是部分函数。如果取极小得到的函数是全函数，则称被取极小的全函数是正则的。注意，这个“正则”与上述图灵机的 n 正则毫不相干。

3.2 递归

3.2.1 A 部分递归的函数

一个函数，如果能够由 $C_A(x)$ 和其他初等递归函数经过有限次复合和取极小而得到，则称该函数是 A 部分递归的。显然，任何这样的函数都可以用指定了相对集 A 的图灵机来实现。

如果一个 A 部分递归函数表示成初等递归函数的复合和取极小时，能够使得每个取极小操作都是对正则函数进行的，则称它是 A 递归的。显然，任何 A 递归函数都是全函数。

3.2.2 A 递归的集合和谓词

如果一个 n 元组的集合 S 的特征函数 $C_S(x_1, \ldots, x_n)$ 是 A 递归的，则称该集合 S 是 A 递归的。

如果集合 R 和 S 都是 A 递归的，那么 $R \cap S$、$R \cup S$ 和 $R^{\complement}$ 也是 A 递归的。因为

$$\begin{aligned} C_{R\cap S} &= (C_R + C_S) \ominus (C_R C_S) \\ C_{R\cup S} &= C_R C_S \\ C_{R^{\complement}} &= 1 \ominus C_R \end{aligned}$$

如果谓词 $P(x_1, \ldots, x_n)$ 的外延 $E(P) = \{(x_1, \ldots, x_n) : P(x_1, \ldots, x_n)\}$ 是 A 递归的，则称该谓词 P 是 A 递归的。谓词 P 的外延的特征函数 $C_{E(P)}$，称为该谓词的特征函数，也记为 C_P。谓词的特征函数都是全函数，因此不存在 A 部分递归的谓词的说法。

如果谓词 P 和 Q 都是 A 递归的，那么它们的逻辑与 $P \wedge Q$、逻辑或 $P \vee Q$ 和 P 的逻辑非 $\neg P$ 也是 A 递归的。因为 $C_{P\wedge Q} = C_{E(P)\cap E(Q)}$，$C_{P\vee Q} = C_{E(P)\cup E(Q)}$，$C_{\neg P} = C_{E(P)^{\complement}}$。 □

3.2.3 有界存在量词、有界全称量词和有界极小算符

如果谓词 $P(y, x_1, \ldots, x_n)$ 是 A 递归的，那么有界存在量词 (bounded existential quantifier) $\exists_0^z y$ 作用于 P 和有界全称量词 (bounded universal quantifier) $\forall_0^z y$ 作用于 P 得到的谓词都是 A 递归的。因为

$$\begin{aligned} (\exists_0^z y)P(y, x_1, \ldots, x_n) &= \bigvee_{y=0}^{z} P(y, x_1, \ldots, x_n), \\ (\forall_0^z y)P(y, x_1, \ldots, x_n) &= \bigwedge_{y=0}^{z} P(y, x_1, \ldots, x_n). \quad \square \end{aligned}$$

有界极小算符 $\mathcal{M}_0^z y$ 作用于谓词 $P(y, x_1, \ldots, x_n)$，得到 $\min_y[y \le z \wedge P(y, x_1, \ldots, x_n)]$ 的补全函数。

如果谓词 $P(y, x_1, \ldots, x_n)$ 是 A 递归的，那么 $(\mathcal{M}_0^z y)P(y, x_1, \ldots, x_n)$ 也是 A 递归的。因为

$$\begin{aligned} &(\mathcal{M}_0^z y)P(y, x_1, \ldots, x_n) \\ = &\left(1 \ominus \prod_{y=0}^{z} C_P(y, x_1, \ldots, x_n)\right) \sum_{t=0}^{z} \prod_{y=0}^{t} C_P(y, x_1, \ldots, x_n). \quad \square \end{aligned}$$

3.2.4 一些常见的递归谓词

如果任何使谓词 $P(x_1,\ldots,x_n)$ 为真的 $(x_1,\ldots,x_n)$，均使谓词 $Q(x_1,\ldots,x_n)$ 为真，则称 P 蕴涵 Q，记为 $P\Rightarrow Q$。

如果 $P\Rightarrow Q$ 而且 $Q\Rightarrow P$，则称 P 等价于 Q，也说“P 当且仅当 Q”，记为 $P \iff Q$。

自然数 y 能够整除 x，记为 $y|x$。这个二元谓词是递归的，因为 $y|x \iff (\exists_0^x z)(x=yz)$。

自然数 x 是一个素数，记为 $\mathrm{Prime}(x)$。它是递归的，因为 $\mathrm{Prime}(x) \iff (x\neq 1)\wedge(\forall_0^x z)[(z=1)\vee(z=x)\vee\neg(z|x)]$。

$x<y$ 是递归的，因为它的特征函数是 $1\ominus(y\ominus x)$。

$x=y$ 是递归的，因为 $x=y \iff \neg(x<y)\wedge\neg(y<x)$。

不超过 $x/2$ 的最大整数，记为 $[x/2]$。这个函数是递归的，因为 $[x/2]=(\mathcal{M}_0^x y)(x<2y+2)$。

3.3 原始递归

3.3.1 原始递归运算

给定全函数 $f(x_1,\ldots,x_n)$ 和 $g(z,y,x_1,\ldots,x_n)$，存在唯一一个全函数 $h(z,x_1,\ldots,x_n)$ 满足

$$\begin{aligned} h(0,x_1,\ldots,x_n) &= f(x_1,\ldots,x_n),\\ h(z+1,x_1,\ldots,x_n) &= g(z,h(z,x_1,\ldots,x_n),x_1,\ldots,x_n). \end{aligned}$$

称 h 是由 f 和 g 经原始递归运算而得到的函数。可证：如果 f 和 g 都是 A 递归的，那么 h 也是 A 递归的。

3.3.2 A 原始递归的函数

$C_A(x)$，$S(x)=x+1$，$N(x)=0$，$U_i^n(x_1,\ldots,x_n)=x_i$， $(1\leq i\leq n)$ 称为初等原始递归函数。

如果一个函数能够由初等原始递归函数的有限次复合和原始递归运算得到，则称为 A 原始递归的。

任何 A 原始递归的函数都是 A 递归的，但是可证：存在 A 递归的函数不是 A 原始递归的。

A 原始递归函数都是全函数。不存在 A 部分原始递归函数的说法。注意到 A 递归函数也都是全函数。实际上，任何 A 递归的函数都可以由 A 原始递归函数只经过一次取极小运算而得到。可见 A 原始递归函数类在取极小运算下不封闭。

3.3.3 一些常见的原始递归函数

前承函数 $P(x)$ 是原始递归的，因为 $P(0)=N(x), \quad P(x+1)=U_1^1(x)$。

函数 x^y 是原始递归的，因为 $x^0=S(N(x)), \quad x^{y+1}=x^yU_1^2(x,y)$。

阶乘 $x!$ 是原始递归的，因为 $0!=S(N(x)), \quad (x+1)!=x!S(x)$。

按 $\mathrm{Prime}(x)$ 的定义，它是原始递归的。

如果谓词 $P(y,x_1,\ldots,x_n)$ 是 A 原始递归的，那么，按照有界极小算符的定义，$(\mathcal{M}_0^z y)P(y,x_1,\ldots,x_n)$ 也是 A 原始递归的。

第 n 个素数 $\mathrm{Pr}(n)$ 是原始递归的，其定义是 $\mathrm{Pr}(0)=0, \quad \mathrm{Pr}(n+1)=(\mathcal{M}_0^{\mathrm{Pr}(n)!+1}y)[\mathrm{Prime}(y)\wedge(y>\mathrm{Pr}(n))]$。

不超过 $x/2$ 的最大整数 $[x/2]$，按其定义，是原始递归的。

函数 $J(x,y)=[((x+y)^2+3x+y)/2]$ 是原始递归的。它从二维表的左上角开始，沿着从右上到左下的对角线方向，逐行地对所有单元进行编号。第 x 行第 y 列的单元编号为 $J(x,y)$。例如，$J(0,0)=0$，$J(0,1)=1$，$J(1,0)=2$，$J(0,2)=3$，$J(1,1)=4$，$J(2,0)=5$，依此类推。

函数 $K(z)=(\mathcal{M}_0^z x)(\exists_0^z y)[z=J(x,y)]$ 是原始递归的。它给出编号为 z 的单元的行号。

函数 $L(z)=(\mathcal{M}_0^z y)(\exists_0^z x)[z=J(x,y)]$ 是原始递归的。它给出编号为 z 的单元的列号。

3.4 图灵机的算术化

前面我们为初等递归函数、复合和取极小运算编制了图灵机程序，从而证明了任何 A 部分递归函数都可以用图灵机来实现。现在，我们把图灵机的符号、指令集、瞬时描述、执行过程等等概念都用递归的函数和谓词来表达，从而证明，任何可用图灵机实现的函数都是 A 部分递归的。

3.4.1 哥德尔配数法

定义从图灵机的控制符集、字母表和状态集的并集到自然数集 $\mathbf{N}$ 的函数 $v(x)$ 为：

$$v(R)=3, \quad v(L)=5, \quad v(S_i)=4i+7, \quad v(q_j)=4j+5, \quad \text{其中} \quad i\geq 0, \quad j\geq 1.$$

给定符号串 $M=c_1\cdots c_n$，定义其哥德尔数为 $\mathrm{gn}(M)=\prod_{k=1}^{n}\mathrm{Pr}(k)^{v(c_k)}$。如果 M 是空串 Λ，则定义 $\mathrm{gn}(\Lambda)=1$。

如果 $n=\mathrm{gn}(M)$，则称 M 是 n 表示的图灵串，记为 $M=\mathrm{Exp}(n)$。

给定符号串的序列 $M_1,\ldots,M_n$，定义其哥德尔数为

$$\mathrm{gn}((M_1,\ldots,M_n))=\prod_{k=1}^{n}\mathrm{Pr}(k)^{\mathrm{gn}(M_k)}.$$

如果 $n=\mathrm{gn}((M_1,\ldots,M_n))$，则称 $(M_1,\ldots,M_n)$ 是 n 表示的图灵串序列，记为 $(M_1,\ldots,M_n)=\mathrm{Els}(n)$。

$v(c_k)$ 都是奇数，而 $\mathrm{gn}(M_k)$ 都是偶数，所以，任何数 n，不可能既表示符号串，又表示符号串序列。

给定图灵机 Z，设 $M_i\quad(1\le i\le n)$ 是它的全部指令。这些指令的任何一种排列的哥德尔数 z，都称为该图灵机 Z 的哥德尔数，记为 $z=\mathrm{gn}(Z)$。显见，有 n 个指令的图灵机可以有 $n!$ 个不同的哥德尔数。

给定数 x，如果存在一个图灵机 Z，使得 $x=\mathrm{gn}(Z)$，则称 Z 是 x 表示的图灵机，记为 $Z=\mathrm{TM}(x)$。

给定图灵机 Z，指定相对集 A，从给定的初始瞬时描述开始，逐步执行，达到某个瞬时描述是最后的，这个过程称为该图灵机 Z 的一个 A 计算。各步的瞬时描述做成一个符号串序列，其哥德尔数称为这个 A 计算的哥德尔数。

3.4.2 图灵谓词 $T_n^A(z,x_1,\ldots,x_n,y)$ 和停机谓词 $z\top_A x$

给定相对集 A，图灵谓词 $T_n^A(z,x_1,\ldots,x_n,y)$ 定义为：如果 z 是一个图灵机 Z 的哥德尔数，y 是 Z 以 $q_1 1^{x_1+1}B1^{x_2+1}B\cdots B1^{x_n+1}$ 为初始瞬时描述的一个 A 计算的哥德尔数，那么该谓词的值为真，否则为假。注意，y 的具体值常常是天文数字。好在我们基本上不必考虑 y 的具体值，而只是关注它的意义和存在性。

二元谓词 $P(z,x)\iff(\exists y)T_1^A(z,x,y)$ 表示图灵机 $\mathrm{TM}(z)$ 在以 x 为输入时会最终停机，因此称为停机谓词，记为 $z\top_A x$。

3.4.3 图灵谓词是 A 原始递归的

Davis 在那本书中给出了构造图灵谓词的详细步骤，每一步所用的函数和谓词都是 A 原始递归的，最终证明了图灵谓词是 A 原始递归的。

例如，$n\,\mathrm{Gl}\,x$ 表示哥德尔数 x 的分解式中第 n 个素数上的幂次。即

$$n\,\mathrm{Gl}\,x=(\mathcal{M}_0^x y)[(\mathrm{Pr}(n)^y|x)\land\neg(\mathrm{Pr}(n)^{y+1}|x)].$$

$\mathrm{Len}(x)$ 表示哥德尔数 x 表示的图灵串的长度，或者图灵串序列的长度。即

$$\mathrm{Len}(x)=(\mathcal{M}_0^x y)[(y\,\mathrm{Gl}\,x>0)\land(\forall_0^x i)((y+i+1)\,\mathrm{Gl}\,x=0)].$$

$\mathrm{CU}(n,x)=C_{\neg(n\,\mathrm{Gl}\,x=11)}(n,x)$：如果 $\mathrm{Exp}(x)$ 的第 n 个符号是 1，则值为 1，否则为 0。

$$\mathrm{Corn}(x) = \sum_{n=1}^{\mathrm{Len}(x)} \mathrm{CU}(n,x) \text{ 表示 } \mathrm{Exp}(x) \text{ 中 1 的个数，即 } \langle \mathrm{Exp}(x) \rangle\text{。}$$

$U(y) = \mathrm{Corn}(\mathrm{Len}(y)\,\mathrm{Gl}\,y)$ 表示 $\mathrm{Els}(y)$ 中最后一个符号串中 1 的个数。

3.4.4 克莱因范式定理

克莱因 (Kleene) 范式定理：任何可用图灵机实现的函数 $f(x_1,\ldots,x_n)$，都可表示为 $U(\min_y T_n^A(z,x_1,\ldots,x_n,y))$ 的形式，其中 $z=\mathrm{gn}(Z)$ 是所用的图灵机 Z 的哥德尔数，$y_0=\min_y T_n^A(z,x_1,\ldots,x_n,y)$ 如果有定义的话，是该图灵机在给定输入下所做的 A 计算的哥德尔数。

如果一个 A 部分递归函数是全函数，那么它一定是 A 递归的。因为，根据范式定理，该 A 部分递归函数可以由一个 A 原始递归的图灵谓词只经过一次取极小运算，再做一次复合而得到。该函数是全函数，就意味着这唯一的一次取极小运算是作用在正则函数上的。所以，该函数是 A 递归的。 □

3.4.5 通用图灵机

函数 $\phi(z,x)=U(\min_y T_1^{\varnothing}(z,x,y))$ 是部分递归的，因此，必可用一个图灵机 Z_u 来实现。称 Z_u 为通用图灵机。如果图灵机 Z 实现 $f(x)$，即 $f(x)=U(\min_y T_1^A(\mathrm{gn}(Z),x,y))$，那么 $U(\min_y T_2^A(\mathrm{gn}(Z_u),\mathrm{gn}(Z),x,y))$ 实现同样的函数 $f(x)$。多元函数可以编码成一元函数，因此，在通用图灵机 Z_u 上，可以模拟任何图灵机的运行。

3.5 可计算性

3.5.1 递归性就是可计算性

说到图灵，更多的人想到的是图灵实验：让人通过电传与可能是机器、也可能是人的另一方对话，如果从对话内容中判断不出对方是机器还是人，那么就可以说，对方，如果是机器的话，一定具有和人一样的智能。

其实图灵还有另一个贡献：说一个函数或谓词是可计算的，就是指这个函数或谓词可以用图灵机来实现。换言之，不能用图灵机来实现的，就是不可计算的。历史上出现过的很多看似不同的计算、演算和推理系统，最终被证实，都可以用图灵机来实现。

照这个标准，递归性就是可计算性，A 部分递归的就是 A 部分可计算的，A 递归的就是 A 可计算的，等等。

如果谓词 $P(x_1,\ldots,x_n)$ 的外延的特征函数 $C_{E(P)}(x_1,\ldots,x_n)$ 是 A 递归的，则称该谓词是 A 可计算的。

3.5.2 A 半可计算的谓词

如果存在以谓词 $P(x_1, \ldots, x_n)$ 的外延为定义域的 A 部分递归函数，则称该谓词是 A 半可计算的。

3.5.3 克莱因枚举定理

克莱因枚举定理：对任何 A 半可计算谓词 $P(x_1, \ldots, x_n)$，均存在相应的图灵机 Z，使得

$$P(x_1, \ldots, x_n) \iff (\exists y) T_n^A(\mathrm{gn}(Z), x_1, \ldots, x_n, y).$$

因为，存在 A 部分递归函数 $f(x_1, \ldots, x_n)$，就意味着存在图灵机 Z 来实现该函数。根据克莱因范式定理，有

$$f(x_1, \ldots, x_n) = U(\min_y T_n^A(\mathrm{gn}(Z), x_1, \ldots, x_n, y)),$$

而 f 的定义域，就是在其上存在 A 计算的那些点。 □

3.5.4 A 半可计算性与A 递归性

如果 $P(x_1, \ldots, x_n)$ 和 $\neg P(x_1, \ldots, x_n)$ 都是 A 半可计算的，那么 P 和 $\neg P$ 都是 A 递归的。因为，设实现 P 和 $\neg P$ 外延上的 A 部分递归函数的图灵机分别为 Z_1 和 Z_2，那么函数

$$h(x_1, \ldots, x_n) = \min_y [T_n^A(\mathrm{gn}(Z_1), x_1, \ldots, x_n, y) \vee T_n^A(\mathrm{gn}(Z_2), x_1, \ldots, x_n, y)]$$

是全函数，因而是 A 递归的。同时，易见

$$\begin{aligned} P(x_1, \ldots, x_n) &\iff T_n^A(\mathrm{gn}(Z_1), x_1, \ldots, x_n, h(x_1, \ldots, x_n)), \\ \neg P(x_1, \ldots, x_n) &\iff T_n^A(\mathrm{gn}(Z_2), x_1, \ldots, x_n, h(x_1, \ldots, x_n)). \end{aligned}$$

所以，P 和 $\neg P$ 都是 A 递归的。 □

停机谓词 $x \top_A x \iff (\exists y) T_1^A(x, x, y)$ 是 A 半可计算的。因为它的外延 $\{x : x \top_A x\}$ 就是 A 部分递归函数 $\min_y [T_1^A(x, x, y)]$ 的定义域。 □

停机谓词的否定 $\neg(x \top_A x)$ 不是 A 半可计算的。假设不然，存在一个图灵机 Z，使得 $\neg(x \top_A x) \iff \mathrm{gn}(Z) \top_A x$。那么，在 $x = \mathrm{gn}(Z)$ 这个点上，将出现矛盾。 □

函数 $f(x) = S(N(\min_y [T_1^A(x, x, y)]))$ 是 A 部分递归的。它的补全函数是谓词 $\neg(x \top_A x)$ 的特征函数，因此，该补全函数不是 A 递归的，不可能用图灵机来实现。

3.5.5 谓词的判定问题

如果谓词 $P(x_1,\ldots,x_n)$ 是递归的，那么存在一个图灵机 Z，实现 P 的特征函数 C_P，对任意给定的 $a_1,\ldots,a_n$，都能最终停机并给出正确的判定结果。因此，称 P 的判定问题是递归可解的。

如果谓词 $P(x_1,\ldots,x_n)$ 不是递归的，比如停机谓词 $x\top_{\varnothing}x$，那么判定 P 或者判定 $\neg P$，至少有一个不能用图灵机实现，因此，称 P 的判定问题是递归不可解的。注意，存在谓词 P，判定 P 和判定 $\neg P$ 都不能用图灵机实现。比如判定任意给定的图灵机 Z 是否实现一个全函数。

3.5.6 停机问题

对简单图灵机 Z，判定它对任意给定的初始瞬时描述 α 是否最终会停机，这个问题称为停机问题 (Halting Problem)。对很多具体的图灵机，其停机问题是递归可解的。通过分析它们的具体特性，可以写出相应的图灵机来实现判定停机的谓词。但是，存在着停机问题递归不可解的图灵机。比如，实现函数 $f(x)=\min_y T_1^{\varnothing}(x,x,y)$ 的图灵机 Z_0。不可能存在判定 Z_0 不停机的图灵机程序。

3.5.7 A 半可计算谓词的一些性质

如果 $P(x_1,x_2,\ldots,x_n)$ 是 A 半可计算的，那么 $(\exists x_1)P(x_1,x_2,\ldots,x_n)$ 也是 A 半可计算的。因为两个存在量词可以编码成一个存在量词。令 $w=2^{x_1}3^y$，根据枚举定理，

$$\begin{aligned}(\exists x_1)P(x_1,x_2,\ldots,x_n) &\iff (\exists x_1)(\exists y)T_n^A(\mathrm{gn}(Z),x_1,x_2,\ldots,x_n,y)\\ &\iff (\exists w)T_n^A(\mathrm{gn}(Z),1\,\mathrm{Gl}\,w,x_2,\ldots,x_n,2\,\mathrm{Gl}\,w). \quad \square\end{aligned}$$

如果 $P(x_1,x_2,\ldots,x_n)$ 是 A 半可计算的，那么 $(\forall_0^z x_1)P(x_1,x_2,\ldots,x_n)$ 也是 A 半可计算的。因为一个有穷序列可以编码成一个数，根据枚举定理，

$$\begin{aligned}&(\forall_0^z x_1)P(x_1,x_2,\ldots,x_n)\\ \iff &(\forall_0^z x_1)(\exists y)T_n^A(\mathrm{gn}(Z),x_1,x_2,\ldots,x_n,y)\\ \iff &(\exists w)(\forall_0^z x_1)T_n^A(\mathrm{gn}(Z),x_1,x_2,\ldots,x_n,(x_1+1)\,\mathrm{Gl}\,w). \quad \square\end{aligned}$$

如果 $P(x_1,\ldots,x_n)$ 和 $Q(x_1,\ldots,x_n)$ 是 A 半可计算的，那么 $P\wedge Q$ 和 $P\vee Q$ 也是 A 半可计算的。因为，根据枚举定理，可设

$$\begin{aligned}P(x_1,\ldots,x_n) &\iff (\exists u)T_n^A(\mathrm{gn}(Z_1),x_1,\ldots,x_n,u)\\ Q(x_1,\ldots,x_n) &\iff (\exists v)T_n^A(\mathrm{gn}(Z_2),x_1,\ldots,x_n,v)\end{aligned}$$

那么

$$\begin{aligned}
& P(x_1,\dots,x_n)\wedge Q(x_1,\dots,x_n)\\
\iff\ & (\exists u)(\exists v)\left(T_n^A(\mathrm{gn}(Z_1),x_1,\dots,x_n,u)\wedge T_n^A(\mathrm{gn}(Z_2),x_1,\dots,x_n,v)\right)\\
& P(x_1,\dots,x_n)\vee Q(x_1,\dots,x_n)\\
\iff\ & (\exists w)\left(T_n^A(\mathrm{gn}(Z_1),x_1,\dots,x_n,w)\vee T_n^A(\mathrm{gn}(Z_2),x_1,\dots,x_n,w)\right). \quad \square
\end{aligned}$$

如果 $P(x_1,x_2,\dots,x_n)$ 是 A 半可计算的，$f(y_1,\dots,y_m)$ 是 A 递归函数，那么 $P(f(y_1,\dots,y_m),x_2,\dots,x_n)$ 也是 A 半可计算的。因为，根据枚举定理，

$$\begin{aligned}
& P(f(y_1,\dots,y_m),x_2,\dots,x_n)\\
\iff\ & (\exists u)T_n^A(\mathrm{gn}(Z),f(y_1,\dots,y_m),x_2,\dots,x_n,u). \quad \square
\end{aligned}$$

3.6 递归可枚举

3.6.1 A 递归可枚举的集合

如果集合 S 可以成为某个 A 半可计算谓词 P 的外延，即 $S=\{x:P(x)\}$，则称该集合 S 是 A 递归可枚举的。

如果集合 S 是某个 A 部分递归函数 $f(x)$ 的值域，那么，S 是 A 递归可枚举的。因为，根据范式定理，可设 $f(x)=U(\min_y T_1^A(\mathrm{gn}(Z),x,y))$，那么集合 S 就是 A 半可计算谓词 $P(t)\iff(\exists x)(\exists w)\left(t=U(w)\wedge T_1^A(\mathrm{gn}(Z),x,w)\right)$ 的外延，即 $S=\{t:P(t)\}$。 $\square$

对任意给定的 A 递归可枚举集 S，都存在图灵机 Z，使得 S 的判定谓词 $x\in S$ 等价于 Z 的停机谓词 $\mathrm{gn}(Z)\top_A x$，这是枚举定理的直接推论。

3.6.2 用 A 原始递归函数“枚举” A 递归可枚举的集合

对任何非空的 A 递归可枚举的集合 S，均存在其值域是 S 的 A 原始递归函数。因为，设 S 是 A 半可计算谓词 $P(x)$ 的外延，即 $S=\{x:P(x)\}$，由枚举定理，存在图灵机 Z，使得 $P(x)\iff\mathrm{gn}(Z)\top_A x\iff(\exists y)T_1^A(\mathrm{gn}(Z),x,y)$。把关于 x,y 的 A 原始递归谓词 $T_1^A(\mathrm{gn}(Z),x,y)$ 的特征函数记为 $\tau_Z^A(x,y)$，那么，可以定义 $f(x)$ 如下：

$$\begin{aligned}
f(0) &= \min S,\\
f(n+1) &= \tau_Z^A(K(n+1),L(n+1))f(n)\\
&\quad+\left(1\ominus\tau_Z^A(K(n+1),L(n+1))\right)K(n+1).
\end{aligned}$$

显见该函数 $f(x)$ 是 A 原始递归的。它能够逐步枚举出满足谓词 $T_1^A(\mathrm{gn}(Z),x,y)$ 的 x，从而把 A 部分递归函数 $\min_y T_1^A(\mathrm{gn}(Z),x,y)$ 的定义域变成该函数 $f(x)$ 的值域。 $\square$

3.6.3 集合 A' 和 K

集合 S 是 A 递归的，当且仅当 S 和 S^{C} 都是 A 递归可枚举的。

集合 $A' = \{x : x\top_A x\}$ 是 A 递归可枚举的，但不是 A 递归的。

集合 $K = \varnothing' = \{x : x\top_\varnothing x\}$ 是递归可枚举的，但不是递归的。

3.6.4 实现全函数的图灵机的哥德尔数的集合

那些实现全函数的图灵机的哥德尔数的集合 R 不是递归可枚举的。假设不然，则应该存在值域为 R 的递归函数 $f(n)$。因为 $\mathrm{TM}(f(n))$ 实现全函数，所以 $U(\min_y T_1^\varnothing(f(n),x,y))$ 是全函数，因而是递归的。那么 $U(\min_y T_1^\varnothing(f(x),x,y))$ $+1$ 也是递归的。因此，应该存在 m 使得 $f(m) \in R$，并且

$$U(\min_y T_1^\varnothing(f(x),x,y)) + 1 = U(\min_y T_1^\varnothing(f(m),x,y)).$$

在 $x = m$ 这个点上会出现矛盾。 $\square$

Chapter 4

组合系统与形式逻辑

4.1 组合系统

4.1.1 字与字序列的哥德尔数

当我们讨论字问题的时候，只使用一个简单的符号集 $\{a_0, a_1, \ldots\}$，称为字母表，并常把 a_0 记为 1。注意，图灵机的符号 S_1 也记为 1，多数情况下，两者井水不犯河水。

符号集变了，哥德尔配数法的基本函数，从符号集到自然数集 $\mathbf{N}$ 的函数 $v(x)$，也要修改为 $v(a_i) = 2i + 1$。给字和字序列配数的方法不变。配数之后，关于字的谓词 R 就自然地转化成关于数值的谓词 R^* 了。说字谓词 R 是递归的，就是说数值谓词 R^* 是递归的。

谓词 $\mathrm{GN}(x) \iff \neg(\exists_1^{\mathrm{Len}(x)} y)[(y \,\mathrm{Gl}\, x = 0) \wedge ((y+1) \,\mathrm{Gl}\, x \neq 0)]$，表示存在正整数序列 a_i, $(1 \leq i \leq n)$，使得 $x = \prod_{i=1}^{n} \mathrm{Pr}(i)^{a_i}$。

函数 $x * y = x \displaystyle\prod_{i=0}^{\mathrm{Len}(y)\ominus 1} \mathrm{Pr}(\mathrm{Len}(x) + i + 1)^{(i+1)\,\mathrm{Gl}\, y}$ 等于 x 和 y 所表示的字拼接成的新字的哥德尔数，或者它们所表示的字序列拼接成的新序列的哥德尔数。

4.1.2 产生式

字谓词 $R^{g,h,k}_{g',h',k'}(X, Y) \iff (\exists P)(\exists Q)[(X = gPhQk) \wedge (Y = g'Ph'Qk')]$，称为与 g, h, k, g', h', k' 对应的产生式。记为 $gPhQk \to g'Ph'Qk'$。它是递归的，因为相应的数值谓词 $R^*(x, y)$ 有

$$
\begin{aligned}
R^*(x,y) \iff\ & \mathrm{GN}(x) \wedge \mathrm{GN}(y) \\
& \wedge (\exists_0^x p)(\exists_0^x q)[(x = \mathrm{gn}(g) * p * \mathrm{gn}(h) * q * \mathrm{gn}(k))
\end{aligned}
$$

$$\wedge(y = \mathrm{gn}(g') * p * \mathrm{gn}(h') * q * \mathrm{gn}(k'))]. \quad \Box$$

如果字 X_0 和 Y_0 使得产生式 $R(X_0, Y_0)$ 为真，即 $X_0 \to Y_0$，则称字 Y_0 是 X_0 关于产生式 R 的一个后承。

4.1.3 组合系统

一个组合系统 Γ，由一个非空字，和有限多个产生式所组成。那个非空字，称为该组合系统 Γ 的公理。

组合系统 Γ 中的一个证明，是指字的一个有限序列 $X_1, \ldots, X_n$，其中 X_1 是 Γ 的公理，而对每个 $i(1 < i \leq n)$，X_i 是 X_{i-1} 关于 Γ 的某个产生式的后承。每个 X_i 称为证明的一步。

一个证明的最后一步 W，称为定理，记为 $\vdash_\Gamma W$。

组合系统 Γ 的所有定理的哥德尔数的集合记为 T_Γ。它是递归可枚举的。因为，首先可以把 Γ 的所有产生式合并成一个递归的谓词：

$$R^*(x, y) \iff R_1^*(x, y) \vee R_2^*(x, y) \vee \cdots \vee R_p^*(x, y).$$

记 $a = \mathrm{gn}(X_1)$，$x = \mathrm{gn}(W)$，y 是整个证明，即这个字序列，的哥德尔数，则

$$\begin{aligned} x \in T_\Gamma \iff & \mathrm{GN}(x) \wedge (\exists y)[(1 \,\mathrm{Gl}\, y = a) \\ & \wedge (\forall_1^{\mathrm{Len}(y) \ominus 1} n) R^*(n \,\mathrm{Gl}\, y, (n+1) \,\mathrm{Gl}\, y) \wedge (\mathrm{Len}(y) \,\mathrm{Gl}\, y = x)]. \quad \Box \end{aligned}$$

4.1.4 半图厄系统

如果产生式 $R_{g',h',k'}^{g,h,k}(X, Y)$ 的 g, k, g', k' 都为空字 Λ，而 h, h' 都非空，则称 R 为与 h, h' 对应的半图厄 (semi-Thue) 产生式。

如果一个组合系统 Γ 的所有产生式都是半图厄的，则称 Γ 是半图厄系统。

Davis 在那本书中介绍了一系列引理，最终证明：对任意给定的简单图灵机 Z，都可以构做相应的半图厄系统来模拟其运行。

集合 S 是 $\varnothing$ 递归可枚举的，即递归可枚举的，当且仅当，存在简单图灵机 Z，使得 $x \in S \iff \mathrm{gn}(Z) \top_\varnothing x$。这是枚举定理在 $A = \varnothing$ 时的特殊情况。

因此，任何递归可枚举的集合都可以由相应的半图厄系统来生成。

我们已经知道，集合 $K = \varnothing' = \{x : x \top_\varnothing x\}$ 是递归可枚举的，但不是递归的。因此，存在生成 K 的半图厄系统 Γ_0。

4.1.5 半图厄系统 Γ_0 的字的判定问题

原则上可以编写一个图灵机程序，对输入 $x = 0, 1, 2, \ldots$，分别输出在 Γ_0 中可证的字 W，而且任何可证的字，都总有一天会被输出。但是，对于尚未被输出的字，我们不知道是我们等的时间不够长，还是它根本就不可证。

为了解决这个问题，可能有人会试图编写一个图灵机程序，对输入 $x = 0, 1, 2, \ldots$，分别输出在 Γ_0 中不可证的字 W，而且任何不可证的字，都总有一天会被输出。可惜，这个输出在 Γ_0 中不可证的字的图灵机程序，从逻辑上讲，就不可能存在。

也就是说，存在一个图灵机 Z_1，对于任意给定的字 W，如果 $\vdash_{\Gamma_0} W$，那么 Z_1 对于输入 $\mathrm{gn}(W)$ 会最终停机并报告结果；否则会一直运行下去，没有机会报告结果。而且，不存在图灵机 Z_2，对于任意给定的字 W，如果 $\neg \vdash_{\Gamma_0} W$，那么 Z_2 对于输入 $\mathrm{gn}(W)$ 会最终停机。这样，我们实际上就没有可行的统一的办法断定任意给定的字 W 在 Γ_0 中是不可证的。因此，我们说，这个半图厄系统 Γ_0 的字的判定问题是递归不可解的。

当然，我们仍然可以具体问题具体分析，判定某些字是不可证的。

4.2 形式逻辑

4.2.1 逻辑

一个逻辑 $\mathcal{L}$，由一个以字为元素的递归集 $\mathcal{U}$，和一个以多元递归字谓词为元素的有限集构成。$\mathcal{U}$ 中的字称为逻辑 $\mathcal{L}$ 的公理 (axiom)。那些字谓词称为逻辑 $\mathcal{L}$ 的推理规则 (rules of inference)。

如果字谓词 $R(Y, X_1, \ldots, X_n)$ 是逻辑 $\mathcal{L}$ 的一个推理规则，则称 Y 是 $\mathcal{L}$ 中的 $X_1, \ldots, X_n$ 用 R 得来的推论。就是说，如果 $\vdash X_1$，而且 $\vdash X_2$，一直到 $\vdash X_n$，那么 $\vdash Y$。

字序列 $X_1, \ldots, X_n$，如果对于每一个 $i (1 \leq i \leq n)$，或者 $X_i \in \mathcal{U}$，或者存在 $j_1, \ldots, j_k < i$，使得 X_i 是 $\mathcal{L}$ 中的 $X_{j_1}, \ldots, X_{j_k}$ 用一个推理规则得来的推论，则称该字序列为逻辑 $\mathcal{L}$ 中的一个证明。其中的每个字 X_i，称为该证明的一步。

一个证明的最后一步 W，称为定理，记为 $\vdash_{\mathcal{L}} W$。也称 W 在 $\mathcal{L}$ 中是可证的。

把组合系统 Γ 看做逻辑时，尽管逻辑中的证明比组合系统的证明在形式上更宽松，但两者可证的字的集合是等同的。因此，常把组合系统 Γ 说成是逻辑 $\mathcal{L}(\Gamma)$。

4.2.2 逻辑 $\mathcal{L}$ 的判定问题

逻辑 $\mathcal{L}$ 的所有定理的哥德尔数的集合 $T_{\mathcal{L}}$ 是递归可枚举的。因为，考虑判定 x 是一个证明的哥德尔数的谓词 $P(x)$，我们有

$$P(x) \iff \mathrm{GN}(x) \wedge (\forall_1^{\mathrm{Len}(x)} n)[(n \,\mathrm{Gl}\, x \in \mathcal{U}^*)$$

$$\vee(\exists_1^{n\ominus 1} i_1,\ldots,i_{n_1})R_1^*(n \operatorname{Gl} x, i_1 \operatorname{Gl} x,\ldots,i_{n_1} \operatorname{Gl} x)$$
$$\vee\cdots\vee(\exists_1^{n\ominus 1} i_1,\ldots,i_{n_k})R_k^*(n \operatorname{Gl} x, i_1 \operatorname{Gl} x,\ldots,i_{n_k} \operatorname{Gl} x)].$$

显然，$P(x)$ 是递归的。又因为

$$x \in T_{\mathcal{L}} \iff (\exists y)(P(y) \wedge (x = \operatorname{Len}(y) \operatorname{Gl} y)),$$

所以，$T_{\mathcal{L}}$ 是递归可枚举的。 □

如果集合 $T_{\mathcal{L}}$ 是递归的，则称逻辑 $\mathcal{L}$ 的判定问题是递归可解的；否则，是递归不可解的。

组合系统 Γ 可以看成是逻辑 $\mathcal{L}(\Gamma)$，而我们已经知道，存在着判定问题递归不可解的组合系统，因此，存在着判定问题递归不可解的逻辑。此时，集合 $T_{\mathcal{L}}$ 不是递归的，也即 ${T_{\mathcal{L}}}^{\complement}$ 不是递归可枚举的。

4.2.3 “对给定集合而言是半完全的”

给定一个可数的字序列 $W_0, W_1, W_2, \ldots$，把各个字的哥德尔数看做下标的函数，即 $f(n) = \operatorname{gn}(W_n)$，如果 $f(n)$ 是递归的，则称该字序列是递归的。

给定逻辑 $\mathcal{L}$ 和集合 Q，如果存在一个递归字序列 $W_0, W_1, W_2, \ldots$，使得 Q 是其中所有可证的字的下标集，即 $Q = \{n : \vdash_{\mathcal{L}} W_n\}$，则称 $\mathcal{L}$ 对于该集合 Q 而言是半完全的。

如果存在另一个递归字序列 $W_0', W_1', W_2', \ldots$，使得 $Q^{\complement}$ 是其中所有可证的字的下标集，即 $Q^{\complement} = \{n : \vdash_{\mathcal{L}} W_n'\}$，则称 $\mathcal{L}$ 对于集合 $Q^{\complement}$ 而言是半完全的。

如果逻辑 $\mathcal{L}$ 对于集合 Q 和 $Q^{\complement}$ 而言都是半完全的，则称逻辑 $\mathcal{L}$ 对于集合 Q 而言是完全的。

4.2.4 半完全性与判定问题的关系

因为 $\vdash_{\mathcal{L}} W_n \iff \operatorname{gn}(W_n) \in T_{\mathcal{L}}$，所以，如果 $T_{\mathcal{L}}$ 是递归的，那么 Q 必是递归的。这样，如果 Q 不是递归的，那么 $T_{\mathcal{L}}$ 也不是递归的，进而逻辑 $\mathcal{L}$ 有递归不可解的判定问题。

如果 Q 可以取任何递归可枚举集，那么 Q 可以取 K，而 K 不是递归的，因此，逻辑 $\mathcal{L}$ 有递归不可解的判定问题。

如果逻辑 $\mathcal{L}$ 有递归不可解的判定问题，那么 $\mathcal{L}$ 对于 $T_{\mathcal{L}}$ 而言是半完全的。因为，既然 $\mathcal{L}$ 有递归不可解的判定问题，那么一定有字不可证，记之为 F。然后我们定义字的序列 $W_0, W_1, W_2, \ldots$：如果存在字 X，使得 $n = \operatorname{gn}(X)$，则取 $W_n = X$，即 $W_{\operatorname{gn}(X)} = X$；否则，取 $W_n = F$。也就是说，把所有的字，无论是否可证，都填在它们的“第哥德尔数”的位置上，剩下的位置都填 F，这样，该序列中可证的那些字的下标就构成集合 $T_{\mathcal{L}}$。注意到谓词

$$\operatorname{Wd}(x) \iff \operatorname{GN}(x) \wedge (\forall_1^{\operatorname{Len}(x)} n)(\exists_0^x z)(n \operatorname{Gl} x = 2z + 1)$$

是递归的。所以

$$f(n) = \mathrm{gn}(W_n) = (1 \ominus C_{\mathrm{Wd}}(n))n + C_{\mathrm{Wd}}(n)\,\mathrm{gn}(F)$$

也是递归的。 $\square$

如果存在一个逻辑 $\mathcal{L}$，对于集合 Q 而言是半完全的，那么 Q 是递归可枚举的。因为 $Q = \{n : \mathrm{gn}(W_n) \in T_{\mathcal{L}}\}$，$T_{\mathcal{L}}$ 是递归可枚举的，而 $\mathrm{gn}(W_n)$ 是递归的函数。 $\square$

4.2.5 哥德尔不完全性定理的抽象形式

上一命题的逆反命题是：如果集合 Q 不是递归可枚举的，那么不存在任何逻辑 $\mathcal{L}$，对于 Q 而言是半完全的。

如果存在一个逻辑 $\mathcal{L}$，对于集合 Q 而言是完全的，那么 Q 是递归。等价命题是：如果集合 Q 是递归可枚举的，但不是递归的，那么不存在任何逻辑 $\mathcal{L}$，对于 Q 而言是完全的。

这两个关于"不存在任何逻辑 $\mathcal{L}$，对于 Q 而言如何如何"的断言，是哥德尔不完全性定理的抽象形式。

4.3 部分命题演算

4.3.1 合式公式、推理规则与公理

部分命题演算 $\mathcal{P}$ 的符号集由非运算符 $\neg$、蕴涵运算符 $\to$、括号 [和]，以及可数无穷多个命题变量名 $p_1, q_1, r_1, p_2, q_2, \ldots$ 所组成。该逻辑的字 W，如果符合下述递归定义，则称为合式公式 (well-formed-formula, or w.f.f.)。每个合式公式 W 都有一个真值 $W^\dagger$，要么为真 (true, T)，要么为假 (false, F)。合式公式的递归定义是：

- 命题变量 p_i 是合式公式。它们被视为自由变量，其真值可以自由地取真或取假。
- 如果 B 是合式公式，那么 $\neg B$ 也是合式公式，其真值为真，当且仅当，B 的真值为假，即 $(\neg B)^\dagger = \neg B^\dagger$。
- 如果 B 和 C 都是合式公式，那么 $[B \to C]$ 是合式公式，其真值为假，当且仅当 B 的真值为真而 C 的真值为假，即 $[B \to C]^\dagger = \neg(B^\dagger \wedge (\neg C^\dagger))$。

该逻辑有两个推理规则：

- 肯定式：三元谓词 $\mathcal{M}(X, Y, Z)$ 为真，当且仅当 Z 为 $[Y \to X]$。该推理规则也称假言推理 (syllogism)。就是说，如果 $\vdash [Y \to X]$，而且 $\vdash Y$，那么 $\vdash X$。

- 代入：二元谓词 $\mathcal{J}(X,Y)$ 为真，当且仅当 X 是这样得到的：将 Y 中的某个命题变量 p_i 的一切出现用某个合式公式 W 来替换。也记为 $\mathcal{J}(X,Y:p_i \leftarrow W)$。也可以说，三元函数 $\mathcal{J}(Y,p_i,W)$ 返回替换后的字，如果 $\vdash Y$，那么 $\vdash \mathcal{J}(Y,p_i,W)$。

该逻辑只有有限个公理。公理都是合式公式，并且真值只取真。

4.3.2 重言式

如果合式公式 W 的真值只取真，则称为重言式 (tautology)。显然，公理都是重言式。

部分命题演算 $\mathcal{P}$ 的所有定理都是重言式。因为，公理都是重言式，而两个推理规则又保持“是重言式”这一性质，即：如果 $\mathcal{M}(X,Y,Z)$ 为真，而且 Y,Z 都是重言式，那么 X 也是重言式；如果 $\mathcal{J}(X,Y)$ 为真，而且 Y 是重言式，那么 X 也是重言式。 $\square$

所有重言式的哥德尔数的集合是递归的。（思考题：证之。）但这并不意味着所有重言式都是可证的。

如果部分命题演算 $\mathcal{P}$ 的所有重言式都是可证的，则称 $\mathcal{P}$ 是完全的。

4.3.3 命题演算 $\mathcal{P}_0$

有一个特殊的部分命题演算，称为充满的部分命题演算，简称命题演算，记为 $\mathcal{P}_0$。它的公理集只包含如下三个重言式：

1. $\vdash [p_1 \to [q_1 \to p_1]]$；
2. $\vdash [[p_1 \to [q_1 \to r_1]] \to [[p_1 \to q_1] \to [p_1 \to r_1]]]$；
3. $\vdash [[\neg q_1 \to \neg p_1] \to [p_1 \to q_1]]$。

$\mathcal{P}_0$ 是完全的，其判定问题是递归可解的。

举例来说，$[p_1 \to p_1]$ 是重言式，证明如下：

$W_1 :\vdash [[p_1 \to [q_1 \to r_1]] \to [[p_1 \to q_1] \to [p_1 \to r_1]]], \{公理 2\}$

$W_2 :\vdash [[p_1 \to [q_1 \to p_1]] \to [[p_1 \to q_1] \to [p_1 \to p_1]]], \{\mathcal{J}(W_2, W_1 : r_1 \leftarrow p_1)\}$

$W_3 :\vdash [p_1 \to [q_1 \to p_1]], \{公理 1\}$

$W_4 :\vdash [[p_1 \to q_1] \to [p_1 \to p_1]], \{\mathcal{M}(W_4, W_3, W_2)\}$

$W_5 :\vdash [[p_1 \to [q_1 \to p_1]] \to [p_1 \to p_1]], \{\mathcal{J}(W_5, W_4 : q_1 \leftarrow [q_1 \to p_1])\}$

$W_6 :\vdash [p_1 \to p_1], \{\mathcal{M}(W_6, W_3, W_5)\}$ $\square$

4.3.4 递归不可解的判定问题

但是，半图厄系统都可译为部分命题演算，而前面讲过，存在判定问题递归不可解的半图厄系统，因此，存在判定问题递归不可解的部分命题演算。

判定所给的部分命题演算是否完全，这是递归不可解的。

判定所给的部分命题演算的公理是否都独立，这是递归不可解的。

4.4 一阶逻辑

4.4.1 项、合式公式与推理规则

一阶逻辑的符号集由非运算符 $\neg$、蕴涵运算符 $\rightarrow$、逗号 ,、方括号 [和]、圆括号 (和)，以及可数无穷多个个体变量名 $x_1, y_1, z_1, x_2, y_2, \ldots$、个体常量名、谓词名和函数名所组成。个体是指个体常量或个体变量。每个谓词或函数都有特定的变元数。

一阶逻辑 $\mathcal{F}$ 的项，是满足如下递归定义的字：

- 个体是项；
- 如果 $X_1, \ldots, X_n$ 都是项，a 是一个 n 元函数，那么 $a(X_1, \ldots, X_n)$ 是项。

一阶逻辑 $\mathcal{F}$ 的合式公式，是满足如下递归定义的字：

- 如果 $X_1, \ldots, X_m$ 都是项，p 是一个 m 元谓词，那么 $p(X_1, \ldots, X_m)$ 是合式公式；
- 如果 W_1, W_2 都是合式公式，那么 $[W_1 \rightarrow W_2]$ 是合式公式；
- 如果 W 是合式公式，那么 $\neg W$ 是合式公式；
- 如果 W 是合式公式，v 是个体变量，那么 $(v)W$ 是合式公式。

语义上常把合式公式理解为谓词或者语句。

一阶逻辑 $\mathcal{F}$ 有两个推理规则：

- 肯定式：三元谓词 $\mathcal{M}(X, Y, Z)$ 为真，当且仅当 Z 为 $[Y \rightarrow X]$。
- 概括：二元谓词 $\mathcal{G}(X, Y)$ 为真，当且仅当 X 为 $(v)Y$，其中 v 是个体变量。即：如果 $\vdash_{\mathcal{F}} Y$，那么 $\vdash_{\mathcal{F}} (v)Y$。

4.4.2 闭的合式公式

如果 $(v)A$ 是合式公式，字 W 包含 $(v)A$ 为子串，则称该子串 $(v)A$ 是 W 中以 v 为主体的一个闭子式。

个体变量 v 在 W 中作为闭子式的主体而出现，称为约束出现。如果 v 在 W 中的出现不是约束出现，则称为自由出现。

如果合式公式 W 中没有任何个体变量有自由出现，则称 W 为闭的 (closed)，简记为 c.w.f.f.。语义上常把闭的合式公式理解为语句。

4.4.3 影响域与代入

合式公式 W 中的所有以 v 为主体的闭子式的全体，称为 v 在 W 中的约束域。合式公式 X 中出现的所有个体变量在 W 中的约束域的并，称为 X 在 W 中的影响域。

字函数 $\mathcal{J}(W,v,X)$ 定义为：如果个体变量 v 在 W 中的任何自由出现，都不处于 X 在 W 中的影响域内，那么 $\mathcal{J}(W,v,X)$ 就是用 X 替换 v 在 W 中的所有自由出现而得到的字；否则，$\mathcal{J}(W,v,X)$ 就是 W。就是说，把 X 代入 W，只能在它的影响域之外进行。

4.4.4 一般公理与特殊公理

一阶逻辑的公理集是个递归的无限集，包括有限个（可以是零个）特殊公理和无限个一般公理。

所有一般公理被划分成五种图式，也称模板：

1. $\vdash [A \to [B \to A]]$；
2. $\vdash [[A \to [B \to C]] \to [[A \to B] \to [A \to C]]]$；
3. $\vdash [[\neg B \to \neg A] \to [A \to B]]$；
4. $\vdash [(v)A \to \mathcal{J}(A,v,X)]$；
5. $\vdash [(v)[A \to B] \to [A \to (v)B]]$，其中 v 在 A 中没有自由出现。

在这些图式中，用具体的个体变量替换 v，具体的项替换 X，具体的合式公式替换 A,B,C，就可以得到具体的公理。注意：一般公理不要求是闭的。

特殊公理都是闭的合式公式，即 c.w.f.f.。

4.4.5 充满的一阶逻辑 $\mathcal{F}_0$

没有特殊公理的一阶逻辑，称为充满的一阶逻辑，记为 $\mathcal{F}_0$。

如果一阶逻辑 $\mathcal{F}$ 的特殊公理为 $A_1, \ldots, A_n$，则记为 $\mathcal{F} = \mathcal{F}_0(A_1, \ldots, A_n)$。

$\vdash_{\mathcal{F}_0(A_1,\ldots,A_n)} W \iff \vdash_{\mathcal{F}_0} [A_1 \to [A_2 \to \cdots \to [A_n \to W] \cdots]]$。该式左边的意思是：认定 $A_1, \ldots, A_n$ 成立，可以推证 W 成立；右边的意思是：不认定 $A_1, \ldots, A_n$ 成立，但是可以推证，在 $A_1, \ldots, A_n$ 都成立的情况下，W 成立。这两种表述显然等价。

因此，任何一阶逻辑的推理能力都不大于 $\mathcal{F}_0$。

4.4.6 正规系统与一阶逻辑

如果产生式 $R^{g,h,k}_{g',h',k'}(X, Y)$ 的 h, k, g', h' 都为空字 Λ，而 g, k' 都非空，则称 R 为与 g, k' 对应的正规产生式。

如果一个组合系统 Γ 的所有产生式都是正规的，则称 Γ 是正规系统。

存在判定问题递归不可解的正规系统。因为，对任意给定的半图厄系统 τ，设其字母表是 $a_1, \ldots, a_n$，则可以构做一个正规系统 $\nu(\tau)$，其字母表是 $a_1, \ldots, a_n, a'_1, \ldots, a'_n$，使得 τ 的定理恰是 $\nu(\tau)$ 的定理中不含字母 $a'_1, \ldots, a'_n$ 的那些定理。Davis 在那本书中给出了具体的构做方法。这样，因为任何递归可枚举集都可以由半图厄系统生成，所以，也可以由正规系统生成。进而，K 可以由某个正规系统生成。因为 K 是递归可枚举的，但不是递归的，所以能够生成 K 的这个正规系统的判定问题是递归不可解的。

Davis 在那本书中还证明了：任何正规系统都可以译为一阶逻辑。这就意味着，存在判定问题递归不可解的一阶逻辑。进而，$\mathcal{F}_0$ 的判定问题是递归不可解的。

4.5 算术逻辑

4.5.1 数词与变量

在算术逻辑中，对于每个整数 n，有且只有一个字与之相关联，称为与 n 相关联的数词，记为 n^*。实际上就是整数的字符串形式。显然，我们可以合理规定函数 $f(n) = \mathrm{gn}(n^*)$ 是递归的。

在算术逻辑中，有可数无穷多个不同的字 $x_1, x_2, x_3, \ldots$，称为变量。我们合理规定函数 $g(n) = \mathrm{gn}(x_n)$ 是递归的。另外，规定不存在三个都不空的字 A, B, C，使得 AB 和 BC 都是变量。这样可以简化一些函数的构造过程。

4.5.2 合式公式

对每个算术逻辑 $\mathcal{U}$，

- 存在一个非空的递归的字集合，记为 $\mathrm{Wff}_{\mathcal{U}}$，其中的字称为合式公式。符号 A, B, C 等，常用来指代合式公式，称为原子。在很多方面，原子都类似于部分命题演算中的命题变量。但是，要注意，在算术逻辑中，变量 $x_1, x_2, \ldots$ 都不是合式公式。

- 存在一个递归的字函数 $\mathcal{I}(A, B)$，称为蕴涵。它的值是合式公式，当且仅当 A 和 B 都是合式公式。记 $\mathcal{I}(A, B)$ 为 $[A \to B]$。

- 存在一个递归的字函数 $\mathcal{N}(A)$，称为否定。它的值是合式公式，当且仅当 A 是合式公式。记 $\mathcal{N}(A)$ 为 $\neg A$。

- 存在一个递归的字函数 $\mathcal{G}(M, A)$，称为概括。它的值是合式公式，当且仅当 M 是变量而且 A 是合式公式。记 $\mathcal{G}(M, A)$ 为 $(M)A$。语义上，此处的 (M) 表示整数上的全称量词。

- 存在一个递归的字谓词 $\mathcal{B}(M, A)$，称为约束。当 A 是合式公式，M 是变量时，递归地，对任何合式公式 C，$\mathcal{B}(M, (M)C)$ 为真；如果 $\mathcal{B}(M, B)$ 为真，而 B 是 A 的子串，那么 $\mathcal{B}(M, A)$ 为真。其他情况下，$\mathcal{B}(M, A)$ 为假。

  如果 x_i 在 A 中有出现，但是 $\mathcal{B}(x_i, A)$ 为假，则称 x_i 在 A 中是自由的。

  如果合式公式 A 的任何合式子式都不含自由的变量，则称 A 是闭的。闭的合式公式简记为 c.w.f.f. 。

- 存在一个递归的字函数 $\mathcal{J}(X, M, N)$，称为代入。当 X 是合式公式，M 是变量或者原子时，它的值就是在 X 中用 N 替换 M 的所有出现而得到的字。

  如果 A 是合式公式，x_i 在 A 中是自由的，N 是变量或者数词，那么 $\mathcal{J}(A, x_i, N)$ 是合式公式。

- 存在一个递归的字函数 $\mathcal{C}(W)$，称为闭包。如果 W 是 c.w.f.f. ，那么 $\mathcal{C}(W)$ 就是 W；如果 W 是合式公式但不是 c.w.f.f. ，那么 $\mathcal{C}(W)$ 就是在 W 之前加上合适的 (x_i) 而得到的 c.w.f.f. ；如果 W 不是合式公式，那么 $\mathcal{C}(W)$ 也不是合式公式。

  合式公式 W 是闭的，当且仅当 $\mathcal{C}(W) = W$。而 $\mathcal{C}(W)$ 是递归的，所以 c.w.f.f. 的类是递归的。但这并不意味着，所有 c.w.f.f. 在 $\mathcal{U}$ 中都是可证的。

  可证 $\vdash_{\mathcal{U}} A \iff \vdash_{\mathcal{U}} (x_i)A$，进而 $\vdash_{\mathcal{U}} A \iff \vdash_{\mathcal{U}} \mathcal{C}(A)$。

4.5.3 公理与推理规则

算术逻辑有六个公理：

- 如果 A, B, C 是合式公式，那么

  (1) $\vdash [A \to [B \to A]]$；

  (2) $\vdash [[A \to [B \to C]] \to [[A \to B] \to [A \to C]]]$；

  (3) $\vdash [[\neg B \to \neg A] \to [A \to B]]$；

- 如果 A 是合式公式，x_i, x_j 在 A 中不是约束的，那么

  (4) $\vdash [\mathcal{J}(A, x_i, x_j) \to \neg(x_i)\neg A]$；

  (5) $\vdash [\mathcal{J}(A, x_i, n^*) \to \neg(x_i)\neg A]$；

- 如果 A, B 是合式公式，x_i 在 A 中不是自由的，那么

  (6) $\vdash [(x_i)[A \to B] \to [A \to (x_i)B]]$。

算术逻辑和部分命题演算一样，有两个推理规则：肯定式（即假言推理）和对原子的代入。一阶逻辑中的概括，在算术逻辑中是定理。

4.5.4 n 元合式公式

如果合式公式 W 有且只有 n 个自由变量，则称 W 是 n 元的。

设 $(x_1, \ldots, x_n)$ 是整数的 n 元组，W 是 n 元合式公式，那么，我们用 $W(x_1, \ldots, x_n)$ 表示用 x_i^* 替换 W 中 x_i 的一切出现，$i = 1, 2, \ldots, n$，而得到的合式公式。

对给定的数值谓词 $P(x_1, \ldots, x_n)$，如果在算术逻辑 $\mathcal{U}$ 中存在 n 元合式公式 W，使得对于每个 n 元组 $(x_1, \ldots, x_n)$，

- 如果 $P(x_1, \ldots, x_n)$ 为真，则 $\vdash_{\mathcal{U}} W(x_1, \ldots, x_n)$；
- 如果 $P(x_1, \ldots, x_n)$ 为假，则 $\vdash_{\mathcal{U}} \neg W(x_1, \ldots, x_n)$；

那么，就称 P 在 $\mathcal{U}$ 中是完全可表示的。

如果数值谓词 P 在算术逻辑 $\mathcal{U}$ 中是完全可表示的，那么 P 是递归的。因为，前面讲过，任何逻辑 $\mathcal{U}$ 的所有定理的哥德尔数的集合 $T_{\mathcal{U}}$ 是递归可枚举的。如果 P 在 $\mathcal{U}$ 中完全可表示，那么 P 和 $\neg P$ 都是半可计算的。 $\square$

4.5.5 算术逻辑的一致性与 ω 一致性

在算术逻辑 $\mathcal{U}$ 中，如果存在一个字 A，使得 $\vdash_{\mathcal{U}} A$，并且 $\vdash_{\mathcal{U}} \neg A$，那么，对任何合式公式 B，均有 $\vdash_{\mathcal{U}} B$。证明如下：

$$
\begin{aligned}
&W_1 :\vdash [A \to [B \to A]], \{\text{公理 } 1\}\\
&W_2 :\vdash [\neg A \to [B \to \neg A]], \{\mathcal{J}(W_1, A, \neg A)\}\\
&W_3 :\vdash [\neg A \to [\neg B \to \neg A]], \{\mathcal{J}(W_2, B, \neg B)\}\\
&W_4 :\vdash \neg A, \{\text{所设}\}\\
&W_5 :\vdash [\neg B \to \neg A], \{\mathcal{M}(W_5, W_4, W_3)\}\\
&W_6 :\vdash [[\neg B \to \neg A] \to [A \to B]], \{\text{公理 } 3\}\\
&W_7 :\vdash [A \to B], \{\mathcal{M}(W_7, W_5, W_6)\}\\
&W_8 :\vdash A, \{\text{所设}\}\\
&W_9 :\vdash B. \{\mathcal{M}(W_9, W_8, W_7)\} \quad \square
\end{aligned}
$$

上述算术逻辑 $\mathcal{U}$ 称为是不一致的 (inconsistent)。如果一个算术逻辑不是不一致的，则称为是一致的 (consistent)。

在算术逻辑 $\mathcal{U}$ 中，如果存在一个一元合式公式 W，使得：$\vdash_{\mathcal{U}} \neg(x_1)W$，并且，对于所有整数 n，均有 $\vdash_{\mathcal{U}} W(n)$，则称 $\mathcal{U}$ 是 ω 不一致的。如果一个算术逻辑不是 ω 不一致的，则称为是 ω 一致的。

显见，如果 $\mathcal{U}$ 是不一致的，那么一定也是 ω 不一致的。因此，如果 $\mathcal{U}$ 是 ω 一致的，那么一定也是一致的。“好”的算术逻辑至少得是 ω 一致的。

4.5.6 基础谓词类与算术逻辑的充足性

给定二元谓词类 Δ，如果对于任何递归可枚举集 Q，均存在谓词 $P(x,y) \in \Delta$，使得 $Q = \{x : (\exists y)P(x,y)\}$，则称 Δ 是一个基础。前面讲半可计算谓词的时候学过，二元递归谓词类，二元原始递归谓词类，形如 $T(\mathrm{gn}(Z), x, y)$ 的谓词类，都是基础。

给定算术逻辑 $\mathcal{U}$，如果存在一个基础 Δ，其中的所有谓词在 $\mathcal{U}$ 中都是完全可表示的，则称 $\mathcal{U}$ 是充足的。

此时，这些谓词都是递归的。而且，对于任何 $P(x,y) \in \Delta$，均存在二元合式公式 W，使得如果 $P(x,y)$ 为真，则 $\vdash_{\mathcal{U}} W(x,y)$；如果 $P(x,y)$ 为假，则 $\vdash_{\mathcal{U}} \neg W(x,y)$。称 W 是与谓词 P 相应的合式公式。

算术逻辑 $\mathcal{U}$，对某个基础中的任何谓词都有相应的合式公式，表明 $\mathcal{U}$ 有充足的能力对该基础中的任何谓词进行证明。但是，这并不意味着 $\mathcal{U}$ 能够对 $\mathrm{Wff}_{\mathcal{U}}$ 中的任意的合式公式进行证明。

4.5.7 ω 一致的、充足的算术逻辑

如果算术逻辑 $\mathcal{U}$ 是 ω 一致的、充足的，那么 $\mathcal{U}$ 对于任何递归可枚举集 Q 而言都是半完全的。因为，按所设，存在谓词 $P(x,y)$，使得 $Q=\{x:(\exists y)P(x,y)\}$。设与 P 相应的合式公式为 W，那么，记 V 是一元合式公式 $\neg(x_2)\neg W$，则我们可证 $Q=\{x:\vdash_{\mathcal{U}} V(x)\}$。

如果整数 $n\in Q$，那么存在整数 y_0，使得 $P(n,y_0)$ 为真，即 $\vdash_{\mathcal{U}} W(n,y_0)$。考虑到 $W(n,y_0)=\mathcal{J}(\mathcal{J}(W,x_2,y_0^*),x_1,n^*)$，公理 5 经适当代入后就是

$$\vdash_{\mathcal{U}} [W(n,y_0)\to \mathcal{J}(\neg(x_2)\neg W,x_1,n^*)].$$

再由假言推理得 $\vdash_{\mathcal{U}} \mathcal{J}(\neg(x_2)\neg W,x_1,n^*)$，即 $\vdash_{\mathcal{U}} V(n)$。

如果整数 n 满足 $\vdash_{\mathcal{U}} V(n)$，即 $\vdash_{\mathcal{U}} \mathcal{J}(\neg(x_2)\neg W,x_1,n^*)$。因为 $\mathcal{U}$ 是 ω 一致的，所以，存在某个整数 m，使得 $\neg\vdash_{\mathcal{U}} \neg W(n,m)$，从而 $P(n,m)$ 非假，即 $P(n,m)$ 为真，故 $n\in Q$。函数 $\mathrm{gn}(V(x))$ 递归为显然。 □

因此，这样的算术逻辑对于 K 而言是半完全的，进而有递归不可解的判定问题。

4.5.8 算术逻辑的完全性

如果对于算术逻辑 $\mathcal{U}$ 的任何闭的合式公式 A，或者 $\vdash_{\mathcal{U}} A$，或者 $\vdash_{\mathcal{U}} \neg A$，即 $\neg\vdash_{\mathcal{U}} A \iff \vdash_{\mathcal{U}} \neg A$，则称 $\mathcal{U}$ 是完全的。否则称它是不完全的。此时存在 $\mathcal{U}$ 的闭的合式公式 A，使得 A 和 $\neg A$ 都不是 $\mathcal{U}$ 的定理，即：A 从 $\mathcal{U}$ 内部是无法判定的。例如，有许多逻辑的一致性不能由其自身加以证明。

如果算术逻辑 $\mathcal{U}$ 是完全的，则它的判定问题是递归可解的。因为，可设 $\mathcal{U}$ 中所有 c.w.f.f. 的哥德尔数的集合为 F，前面讲过，F 是递归的。记 $P^+\subseteq F$ 为所有是定理的 c.w.f.f. 的哥德尔数的集合，$P^-=F\setminus P^+$。首先，显见 $P^+=F\cap T_{\mathcal{U}}$，因此 P^+ 是递归可枚举的。其次，因为 $\mathcal{U}$ 是完全的，$\neg\vdash_{\mathcal{U}} A \iff \vdash_{\mathcal{U}} \neg A$，所以 $P^-=F\cap\{x:\mathcal{N}^*(x)\in T_{\mathcal{U}}\}$，而 $P^{+\complement}=P^-\cup F^{\complement}$，因此，$P^{+\complement}$ 也是递归可枚举的。这样，P^+ 是递归的。前面讲过，$\vdash_{\mathcal{U}} W \iff \vdash_{\mathcal{U}} \mathcal{C}(W)$，所以，$T_{\mathcal{U}}=\{x:\mathcal{C}^*(x)\in P^+\}$，因此，$T_{\mathcal{U}}$ 是递归的。 □

4.5.9 哥德尔不完全性定理

如果算术逻辑 $\mathcal{U}$ 是 ω 一致的、充足的，那么 $\mathcal{U}$ 必有递归不可解的判定问题，进而必定是不完全的。这就是哥德尔不完全性定理。

该定理还有另一个证法：设 $\{x:\vdash_{\mathcal{U}} V(x)\}=K$，那么 $\{x:\vdash_{\mathcal{U}} \neg V(x)\}\subseteq K^{\complement}$。前面讲过，没有任何一个逻辑对 $K^{\complement}$ 而言是半完全的，因此 $\{x:\vdash_{\mathcal{U}}$

$\neg V(x)\} \neq K^{\mathsf{C}}$。这意味着，存在数 $n \in K^{\mathsf{C}}$，使得 $\neg \vdash_{\mathcal{U}} \neg V(x)$，而 $n \in K^{\mathsf{C}}$ 又表明 $\neg \vdash_{\mathcal{U}} V(x)$，因而 $\mathcal{U}$ 是不完全的。 □

以上从图灵机到哥德尔不完全性定理的这部分内容主要来自 Davis 的那本书，有适当改动。本文档仅限于学术交流之目的，而且如实说明了引述的出处，所以，这些引述不属于学术不端。:-)

现代数学的各种分支，归根结底都是逻辑：从一些基本定义和公理出发，通过演绎推理证明各种定理。集合论、概率论、几何、代数、分析等，概莫能外。所以，哥德尔不完全性定理的影响确实是广泛而深远的。

给定一个关于集合的谓词 $P(x)$，判断它的外延 $\{x : P(x)\}$ 是本性类，还是集合，尽管我还没有看到证明，但是我觉得可以合理猜测，这个判定问题是递归不可解的。

4.6 图灵机和算术逻辑

4.6.1 “忙碌的河狸”函数

从某种意义上讲，图灵机和算术逻辑在“处理能力”上是相当的。因此，存在无法用图灵机实现的函数，与存在无法用算术逻辑证明或证伪的命题，是基本相当的问题。

举一个不可计算函数的实例。前面讲过，实现后继函数 $S(x) = x + 1$ 的图灵机只含有一条指令 q_1BBq_1。初始瞬时描述为 q_11^{x+1}。该图灵机启动后，发现没有合适的指令可以执行，所以即刻停机。注意一个问题，该机实际上扫描了指令 q_1BBq_1，停机的决策就是根据这个指令做出的。所以，我们把这种情况也说成是“执行”了这个指令。那么，该机就是“执行一条指令后停机”。

现在，我们考虑初始带为全空的情况。显然，该机将反复执行 q_1BBq_1，永不停机。换成 q_11Bq_1，这个新的图灵机就会“执行一条指令后停机”。

问：有 n 个状态的图灵机，若初始带全空，在那些会最终停机的图灵机中，最多能执行多少条指令？这个函数称为“忙碌的河狸”函数 (The Busy Beaver Function)。显然，$B(1) = 1$，有人算过 $B(2) = 6$，原则上，对于给定的具体 n，可以计算 $B(n)$。但是，对于不同的 n，计算方法总会有所不同，没有任何单一的程序，可以对任意给定的 n 计算 $B(n)$。何以见得？

如果 $B(n)$ 是可计算的，那么停机问题就是递归可解的。因为，我们可以利用 $B(n)$ 来做一个判定任意给定的程序在给定输入下能否停机的程序。我们可以做一个引导程序生成器，对给定的程序和给定的输入，生成一个引导程序，从全空的初始带打印出给定的输入，再回到最左端，形成初始瞬时描述。然后，把待判定的图灵机拼接到引导程序的后面，做成一个新的图灵机。计算这个新的图灵机的状态数 n，调用 $B(n)$ 得到执行指令数的上限，用通用图灵机模拟新图灵机在初始全空的带上运行 $B(n)$ 步，若仍需继续模拟，则判定上述

图灵机在上述输入下不能停机；否则，判定能停机。此程序除了 $B(n)$ 外都是可计算的。

但是，我们已经知道停机问题是递归不可解的，所以，函数 $B(n)$ 不可能是可计算的。 □

图灵机计算能力的限制，来源于它的循规蹈矩、埋头苦干、只拉车不看路的行为方式。从哲学思辨的角度讲，这种行为方式在精确、严密方面有所得，在创造、适应方面有所失。但是，从历史和实践上看，正是这种行为方式在精确和严密方面的巨大进展，为创造和适应奠定了坚实的基础，极大地促进了人类各方面创造力的发挥和适应力的提高。

Chapter 5

测度论

好啦，我们已经从五个西红柿入手，讲到了集合论、数据库设计、可数集与连续统、康托幂集定理、罗素悖论、图灵机、递归函数论、形式逻辑、哥德尔不完全性定理。现在，我们回头继续谈积分。

5.1 可测集与正测度

5.1.1 球极投影与球心投影

如图 5.1 所示。 图中的粗直线按其横坐标做成实数轴，实数集 $\mathbf{R}$ 中的实数

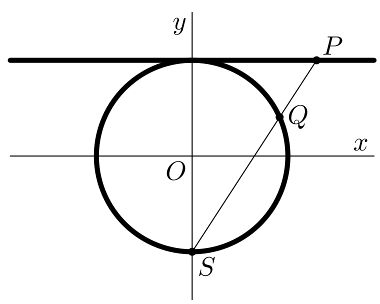

图 5.1: 球极投影

与该直线上的点一一对应。该直线上任意一点 P 与点 S 的连线与粗线圆有且只有一个交点 Q。这个从该直线上的点到该圆上的点的映射称为球极投影。注意到，该直线上没有点可以映射到点 S。但是，我们可以在该直线上添加一点，并规定它映射到点 S，从而使得球极投影成为全双射。添加的这一点称为无穷远点。显然，我们还得在实数集 $\mathbf{R}$ 中添加一个元素，使它与无穷远点对应。记

这个元素为 ∞。这样，从 $\mathbf{R}\cup\{\infty\}$ 到扩充后的直线，再到该圆，按上述映射做成一一对应。

如图 5.2 所示。 图中粗直线上任意一点 P 与点 O 的连线与粗线半圆有且

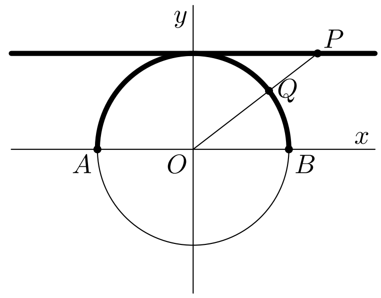

图 5.2: 球心投影

只有一个交点 Q。这个从该直线上的点到该闭的半圆上的点的映射称为球心投影。注意到，该直线上没有点可以映射到点 A 或点 B。但是，我们可以在该直线上添加两个无穷远点，并规定它们分别映射到点 A 和点 B，从而使得球心投影成为全双射。显然，我们还得在实数集 $\mathbf{R}$ 中添加两个元素，使它们分别与两个无穷远点对应。记与 A 对应的为 $-\infty$，与 B 对应的为 $+\infty$。这样，从 $\mathbf{R}\cup\{-\infty,+\infty\}$ 到扩充后的直线，再到该闭的半圆，按上述映射做成一一对应。

5.1.2 紧致集

设想一只袋鼠在一个圆形跑道上跳啊跳，它的足迹可以抽象为圆上的点的序列。这个序列本身不一定收敛。但是，至少存在一个子序列是收敛的。比如袋鼠在一个直径的两端来回跳的情况，我们只要隔一个取一个，就能得到一个常值点列，它收敛于该常值点。

给定度量空间 R 中的点集 A，如果 A 中的任何点列都有在 R 中收敛的子点列，则称 A 是 R 中的致密集，也称列紧集。

闭的致密集，称为紧致集。按定义，紧致集 A 中的任何点列都有在 A 中收敛的子点列。这是一个很好的分析性质。

按通常的度量，圆和闭线段都是紧致集。

实数集 $\mathbf{R}$ 不是紧致集。让袋鼠在实数轴上一直向前跳，每次跳的距离不小于给定常数，那么，这个序列没有任何子序列可以收敛。但是，按照上节的扩充方案，可以将 $\mathbf{R}$ 改造成为紧致集，称为 $\mathbf{R}$ 的紧致化。用球极投影，只添加一个无穷远点的扩充方案称为一点紧致化。用球心投影，添加两个无穷远点的扩充方案称为两点紧致化。

对于袋鼠在实数轴上一直向前跳的例子，如果把 $\mathbf{R}$ 做一点紧致化，那么这个序列将收敛于无穷远点 ∞；如果把 $\mathbf{R}$ 做两点紧致化，那么这个序列将收敛于无穷远点 $+\infty$。

5.1.3 扩展的实数系统

在测度论中，我们把集合 $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, +\infty\}$ 称为扩展的实数系统 (extended real number system)。在不致混淆的情况下，也可把 $+\infty$ 写成 ∞。元素 $-\infty$ 和 ∞ 与其他实数的运算定义为与通常的极限运算保持一致。比如对任何实数 $a \neq \pm\infty$，定义 $\dfrac{a}{\infty} = 0$，因为 $\lim\limits_{x\to\infty} \dfrac{a}{x} = 0$。

如果对应的极限运算是不定式，比如 $\infty - \infty$ 和 $\dfrac{\infty}{\infty}$，那么，这样的表达式就是不合法的。我们总可以通过做适当的限制来避免不合法的表达式。

对于普通实数 a，函数值 $f(a)$ 与极限值 $\lim\limits_{x\to a} f(x)$ 是不同的。比如 $f(x)$ 在 $x = a$ 处不连续的时候，两者的值就不相等。对于 $-\infty$ 和 ∞，情况也是一样。

理论上可以构造在 $-\infty$ 或 ∞ 处不连续的函数。比如，函数序列 $f_n(\theta) = \left(\dfrac{\theta}{\pi}\right)^n$，$n \in \mathbf{N}$，当 $n \to \infty$ 时，在 $[0, \pi]$ 上趋于函数 $f(\theta)$，该 $f(\theta)$ 在 $[0, \pi)$ 上恒等于 0，但是 $f(\pi) = 1$。把 θ 看做前述球心投影中 OP 与 x 轴的夹角，那么区间 $[0, \pi]$ 就对应闭的半圆。用球心投影将这个半圆投影到扩展的实数轴，那么，该函数在整个实数集 $\mathbf{R}$ 上恒等于零，唯独在 $-\infty$ 这一个点上，函数值为 1。（感谢 primenumber、dnd、shangzh87、DrFaust、IQ60。）

5.1.4 σ-代数

集合 X 上的一个 σ-代数 (σ-algebra)，也称集域代数，是 2^X 的一个子集 $\mathcal{A} \subseteq 2^X$，满足

1. $\varnothing \in \mathcal{A}$；
2. 对任何 $A \in \mathcal{A}$，均有 $A^{\complement} = X \setminus A \in \mathcal{A}$；
3. 对任何可数序列 $A_i \in \mathcal{A}$，其中$i \in \mathbf{N}$，均有 $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{A}$。

σ-代数 $\mathcal{A}$ 中的元素称为可测集 (measurable set)。集合 X 配上这样的一个 σ-代数 $\mathcal{A}$ 就构成一个可测空间 (measurable space)，记为 $(X, \mathcal{A})$。据说，之所以称为 σ-代数，是因为 σ 的读音非常接近于德语的并集 summe。

关于 σ-代数，有必要说明以下几点：

- 虽然定义中只说 $\varnothing \in \mathcal{A}$，但是根据第二条“如果 $A \in \mathcal{A}$，那么 $A^{\complement} \in \mathcal{A}$”，可知 $X = \varnothing^{\complement} \in \mathcal{A}$。实际上，$\{\varnothing, X\}$ 就已经构成了一个 σ-代数，称为平凡的 (trivial) σ-代数。

- 虽然定义中只说 $\mathcal{A}$ 在求并集运算下封闭，但是，根据第二条和德摩根律 (De Morgan's laws) $\left(\bigcup_{i=1}^{\infty} A_i\right)^{\complement} = \bigcap_{i=1}^{\infty} A_i^{\complement}$，$\mathcal{A}$ 在求交集运算下也封闭。
- 虽然定义中说 A_i 是可数序列，但是该定义包含了 A_i 是有限序列的情况。此时，后续的 A_i 都为空集 $\varnothing$。

这样，σ-代数实际上就是定义在 2^X 的子集上的，即 X 的某些子集做成的集合上的，关于并、交、补等运算封闭的代数系统。

2^X 本身是一个 σ-代数。它是 X 上所有 σ-代数中最大的。而平凡的 σ-代数，即 $\{\varnothing, X\}$，是最小的。

$\{\varnothing, A, A^{\complement}, X\}$ 是一个 σ-代数。称为由 $\{A\}$ 生成的 σ-代数。

$\{\varnothing, A, B, A^{\complement}, B^{\complement}, A\cap B, A\cup B, A^{\complement}\cap B, A^{\complement}\cup B, A\cap B^{\complement}, A\cup B^{\complement}, A^{\complement}\cap B^{\complement}, A^{\complement}\cup B^{\complement}, X\}$ 是一个 σ-代数。称为由 $\{A, B\}$ 生成的 σ-代数。

一般地，给定一个 X 的子集的集合 $\mathcal{P}$，我们把包含 $\mathcal{P}$ 的所有 σ-代数中最小的那个，称为由 $\mathcal{P}$ 生成的 σ-代数，记为 $\sigma(\mathcal{P})$。它也常表述为所有包含 $\mathcal{P}$ 的 σ-代数的交。

5.1.5 正测度

所谓正测度 (positive measure)，是指从 σ-代数到非负实数的函数 μ: $\mathcal{A} \to [0, \infty]$，满足：

1. $\mu(\varnothing) = 0$；
2. 可数加性：对任何互斥的 (mutually disjoint) 集合的序列 $(A_i \in \mathcal{A})$，

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

可测空间再配上测度，就称为测度空间 (measure space)，记为 $(X, \mathcal{A}, \mu)$。

5.1.6 计数测度与狄拉克测度

如果 X 是有限集，那么 $\mu(A_i) = |A_i|$ 就恰是一个测度，称为计数测度 (counting measure)。

给定元素 $x \in X$，如果 $x \in A$，则 $\delta_x(A) = 1$，否则 $\delta_x(A) = 0$，这也是一个测度，称为狄拉克测度 (Dirac measure)，也称点质量 (point mass)。乍一看，这似乎与可数加性不符：如果 $x \in A$ 而且 $x \in B$，那么 $\delta_x(A \cup B)$ 岂不是应该等于 2 了吗？仔细检查就会发现，可数加性要求 A 和 B 是互斥的，因此 $x \in A$ 与 $x \in B$，二者只可能取其一。

5.1.7 拓扑、开集与闭集

集合 X 上的一个拓扑 (topology)，是指 2^X 的一个子集 $\mathcal{T} \subseteq 2^X$，其中的元素称为开集，满足：

1. $\varnothing$ 和 X 是开集；
2. 有限个开集的交集是开集；
3. 任意多个开集的并集是开集。

集合 X 配上拓扑之后，就称为拓扑空间，记为 $(X, \mathcal{T})$。

拓扑空间 $(X, \mathcal{T})$ 中，集合 X 的一个子集 U，如果 $X \setminus U$ 是开集，则称 U 是闭集。显然：

1. $\varnothing$ 和 X 是闭集；
2. 有限个闭集的并集是闭集；
3. 任意多个闭集的交集是闭集。

开集与闭集是地位相当的概念。完全可以改用闭集来定义拓扑和开集。

拓扑空间 $(X, \mathcal{T})$ 中，集合 X 和 $\varnothing$ 既是开集，又是闭集。

对任意非空集 X，可以选择配置 2^X 作拓扑，称为离散拓扑。在该拓扑中，X 的任何子集，都既是开集，又是闭集。

以 n 维实欧氏空间 $\mathbf{R}^n$ 为 X，常定义这样的集合 U 为开集：对任何 $u \in U$，存在某个实数 $\varepsilon > 0$，使得 $\{x : |x - u| < \varepsilon\} \subseteq U$。注意，由单一的点构成的子集，称为孤点，在这个拓扑中不符合开集的定义，不是开集；而在离散拓扑中，孤点和其他子集一样，既是开集，又是闭集。

5.1.8 波雷尔 σ-代数

在数轴 $\mathbf{R}$ 上，常把任意多个开区间 (a, b) 的并集称为开集。说拓扑空间 $\mathbf{R}$，一般都是指这个拓扑。

给定拓扑空间 $(X, \mathcal{T})$，在所有 $\mathcal{T} \subseteq \mathcal{A}$ 的 σ-代数 $\mathcal{A}$ 中最小的那个，即由拓扑 $\mathcal{T}$ 生成的那个 σ-代数，称为波雷尔 (Borel) σ-代数。在拓扑空间上谈 σ-代数，通常都是指相应的波雷尔 σ-代数。特别是，我们把拓扑空间 $\mathbf{R}^n$ 上的波雷尔 σ-代数记为 $\mathcal{B}(\mathbf{R}^n)$。

5.1.9 勒贝格测度

在很多普通的应用中，我们可以说：在拓扑空间 $\mathbf{R}^n$ 上，指定闭的超矩形的 n 维体积为该闭集的测度，即

$$\lambda([a_1, b_1] \times \cdots \times [a_n, b_n]) = (b_1 - a_1) \cdots (b_n - a_n),$$

称为勒贝格测度 (Lebesgue measure)。

但是，拓扑空间 $\mathbf{R}^n$ 上可以有结构和性质都十分复杂的点集。有些点集甚至无法合理地定义其勒贝格测度。有保证的只是，任何点集 A 总有可数的开覆盖，记为 $S = \{(a_i, b_i) : i \in \mathbf{N}\}$。定义勒贝格外测度 (outer measure) 为：

$$\lambda^*(A) = \inf \left\{ \sum_{i=1}^{\infty} |b_i - a_i| : S = \{(a_i, b_i) : i \in \mathbf{N}\} \text{ 为 } A \text{ 的可数开覆盖} \right\}.$$

外测度不是测度，不满足可数加性，但是，它满足可数次加性 (countable subadditivity)：

> 如果 $(E_i : i \in \mathbf{N})$ 是数轴 $\mathbf{R}$ 上的点集序列，而 $E = \bigcup_{i=1}^{\infty} E_i$，那么 $\lambda^*(E) \le \sum_{i=1}^{\infty} \lambda^*(E_i)$。

证明从略。给定点集 A，如果对任何 $\varepsilon > 0$，均存在开集 B，使得 $A \subseteq B$，并且 $\lambda^*(B \setminus A) < \varepsilon$，则称 A 是勒贝格可测的。此时的外测度 λ^* 满足可数加性，自然成为勒贝格测度 λ。

5.2 勒贝格积分

5.2.1 指示函数及其积分

给定集合 X 的子集 $S \subseteq X$，可定义函数 I_S: $X \to \{0, 1\}$，如果 $x \in S$，则 $I_S(x) = 1$；否则 $I_S(x) = 0$。函数 $I_S(x)$ 称为子集 S 的指示函数 (indicator function)。

我们要学的第一个积分是：

$$\int_X I_S \, \mathrm{d}\mu = \int_S \mathrm{d}\mu = \mu(S).$$

即：可测集的指示函数的积分等于其测度。这是勒贝格积分 (Lebesgue integral) 的最简单的例子。当积分测度 (integral measure) μ 由测度空间 $(X, \mathcal{M}, \mu)$ 所确定，属于不言而喻时，$\mathrm{d}\mu$ 也可省略不写。

我在本文开头就说，数数是一种积分。现在，你们明白其中的深意了吧。数数是求计数测度，而求任何测度，都是对指示函数做积分。

中学讲的积分实际上是黎曼积分 (Riemann integral)。勒贝格积分不仅包含了黎曼积分，而且还具有更多“好”的性质。故，我们直接讲勒贝格积分。

上式表明，测度与勒贝格积分本质上是相通的，一致的。只要 S 是可测的，I_S 的积分就是可定义的。但是，一个函数 f 的积分可定义，并不意味着该函数可积 (integrable)。只有当 $\int_X |f| \,\mathrm{d}\mu < \infty$ 时，才说该函数是可积的。简单地说，测度无穷大仍说是可测的，而积分无穷大就不说是可积的。

5.2.2 可测函数

给定从可测空间 $(X, \mathcal{A})$ 到可测空间 $(Y, \mathcal{B})$ 的函数 $f\colon X \to Y$，如果值空间 (range space) 中的任何可测集的原象，都是定义空间 (domain space) 中的可测集，即 $\forall B \in \mathcal{B} : f^{-1}(B) \in \mathcal{A}$，则称该函数 f 是可测的。

给定从拓扑空间 $(X, \mathcal{S})$ 到拓扑空间 $(Y, \mathcal{T})$ 的函数 $f\colon X \to Y$，如果值空间中的任何开集的原象，都是定义空间中的开集，即 $\forall A \in \mathcal{T} : f^{-1}(A) \in \mathcal{S}$，则称该函数 f 是连续的。

连续函数不改变点集的邻接性质，可测函数不改变点集的测度性质。

正测度的两个基本性质可以很自然地成为积分的性质：

1. $\int_X I_\varnothing \,\mathrm{d}\mu = 0$；
2. 对互斥的集合序列 A_i，有 $\int_X \sum_{i=1}^{\infty} I_{A_i} \,\mathrm{d}\mu = \sum_{i=1}^{\infty} \int_X I_{A_i} \,\mathrm{d}\mu$。这被称为求和符号与积分符号的可交换性。

讨论勒贝格积分时，如果一个被积函数不是全的，我们总可以补充它，让它在原来的定义域之外取零，使它变成全函数。这并不影响它的可积性和积分的值。所以，除非特别说明，我们总假定所讨论的函数都是全的。

5.2.3 简单函数及其积分

如果一个可测函数的值域是有限集，则称它是简单的 (simple)。

在 $\overline{\mathbf{R}}$ 上取值的简单函数都可表为有限项的和，即 $\phi(x) = \sum_{i=1}^{n} a_i I_{S_i}(x)$，其中 a_i 是互异的函数值，$S_i = \phi^{-1}(\{a_i\}) = \{x : \phi(x) = a_i\}$ 是 $\{a_i\}$ 的原象。

因为我们已经假定所有简单函数都是全的，所以，上述所有 S_i 实际上构成对 X 的一个划分，即：如果 $i \neq j$，那么 $S_i \cap S_j = \varnothing$；而且，$\bigcup_{i=1}^{n} S_i = X$。

因为测度是线性的，即，如果 A_i 和 B_j 分别构成对 X 的划分，那么 $A_i \cap B_j$ 也构成对 X 的一个划分，并且

$$\sum_i a_i \mu(A_i) + \sum_j b_j \mu(B_j) = \sum_{i,j} (a_i + b_j) \mu(A_i \cap B_j).$$

所以，很自然地，勒贝格积分也是线性的。

简单函数的积分是明显的：

$$\int_X \phi \, \mathrm{d}\mu = \int_X \sum_{i=1}^{n} a_i I_{S_i} \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i \int_X I_{S_i} \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i \mu(S_i).$$

在做积分时，我们总是取 $0 \times \infty = 0$。这符合几乎所有应用上的需要。

现在，再来看看我钱包里的这些钱。一共六张，记为 a, b, c, d, e, f，构成集合 X，面值 $\psi(x)$ 是定义在 X 上的简单函数，有 $a_1 = 10$，$a_2 = 20$，$a_3 = 50$，而且，$S_1 = \{a, b\}$，$S_2 = \{c, d, e\}$，$S_3 = \{f\}$。总金额就是

$$\int_X \psi \, \mathrm{d}\mu = \sum_{i=1}^{3} a_i \mu(S_i) = 10 \times 2 + 20 \times 3 + 50 \times 1 = 130.$$

所以，我在本文开头就说，数钱是一种积分。

5.2.4 更多类型的可测函数

当然，学积分不只是为了数钱。我们希望知道更多类型的可测函数。

两个可测函数的复合是可测函数。

连续函数都是可测的。因为，波雷尔 σ-代数是包含相应拓扑的最小的 σ-代数。这意味着，一方面，任何开集都是可测集；另一方面，被任意给定的可测集所包含的最大的开集是存在且唯一的。因此，对值空间中的任意可测集，可以求出它所包含的最大的开集，如果函数是连续的，那么，该开集的原象是开集，因而是可测集。 $\square$

给定从可测空间 $(X, \mathcal{A})$ 到可测空间 $(\overline{\mathbf{R}}, \mathcal{B}(\mathbf{R})$ 的函数 $f \colon X \to \overline{\mathbf{R}}$，它是可测的，当且仅当，对任意实数 c，$(c, \infty]$ 的原象都是可测集。因为，我们有

$$\begin{aligned} (a, b) &= \left(\bigcup_{n=1}^{\infty} \overline{\mathbf{R}} \setminus (b - \frac{1}{n}, \infty] \right) \cap (a, \infty], \\ \{\infty\} &= \bigcap_{n=1}^{\infty} (n, \infty], \\ \{-\infty\} &= \overline{\mathbf{R}} \setminus \bigcup_{n=1}^{\infty} (-n, \infty]. \end{aligned}$$

其他可测集均可由这些类型的可测集通过运算而产生。 $\square$

上述证明中的 $(c, \infty]$，换成 $[c, \infty]$、$[-\infty, c]$、或 $[-\infty, c)$ 等，亦无不可。

5.2.5 上确界、下确界、上极限和下极限

给定实数集合 $\{a_n : n \in \mathbf{N}\}$，如果存在实数 U 使得 $\forall a_n : a_n \leq U$，则称 U 为该集合的上界。所有上界中的最小者，称为上确界 (supremum)，记为

$\sup_n a_n$；如果存在实数 L 使得 $\forall a_n : a_n \geq L$，则称 L 为该集合的下界。所有下界中的最大者，称为下确界 (infimum)，记为 $\inf_n a_n$。

给定有界的实数序列 $(a_n : n \in \mathbf{N})$，我们称序列 $\overline{a_n} = \sup_{k \geq n} a_k$ 为上序列 (upper sequence)，称序列 $\underline{a_n} = \inf_{k \geq n} a_k$ 为下序列 (lower sequence)。显见二者是单调的。又因为 (a_n) 有界，所以二者的极限存在。称 $\lim_{n\to\infty} \overline{a_n}$ 为上极限 (supremum limit)，记为 $\limsup_n a_n$；称 $\lim_{n\to\infty} \underline{a_n}$ 为下极限 (infimum limit)，记为 $\liminf_n a_n$。序列 (a_n) 的极限存在，当且仅当，$\limsup_n a_n = \liminf_n a_n$。

如果 $f_1, f_2, \ldots$ 是在 $\overline{\mathbf{R}}$ 上取值的可测函数序列，那么，函数

$$\sup_n f_n, \quad \inf_n f_n, \quad \limsup_n f_n, \quad \liminf_n f_n,$$

都是可测的。因为，令 $g(x) = \sup_n f_n(x)$，则 $g^{-1}((c, \infty]) = \bigcup_n f_n^{-1}((c, \infty])$。令 $g(x) = \inf_n f_n(x)$，则 $g^{-1}([-\infty, c)) = \bigcup_n f_n^{-1}([-\infty, c))$。而上极限和下极限的定义为：对于集合序列 $S_1, S_2, \ldots$，

$$\limsup_n S_n = \bigcap_{i=1}^{\infty} \bigcup_{k \geq i} S_k, \quad \liminf_n S_n = \bigcup_{i=1}^{\infty} \bigcap_{k \geq i} S_k.$$

它们显然都在 σ-代数的范围之内。 $\square$

5.2.6 函数序列的处处收敛与一致收敛

给定函数序列 $(f_n : n \in \mathbf{N})$、函数 f、和它们的共同的定义域的一个子集 A，如果对 A 中的每一个点 $x \in A$，均有 $\lim_{n\to\infty} f_n(x) = f(x)$，则称函数序列 (f_n) 在集合 A 上处处 (pointwise) 收敛于函数 f。处处收敛是条件最宽松，或者说“最弱 (weakest)”，的收敛。

给定函数序列 $(f_n : n \in \mathbf{N})$、函数 f 和它们的共同的定义域的一个子集 A，如果 $\lim_{n\to\infty} \sup_{x \in A} |f_n(x) - f(x)| = 0$，则称函数序列 (f_n) 在集合 A 上一致地 (uniformly) 收敛于函数 f。一致收敛也称均匀收敛。

易见函数序列 $(f_n(x) = x^n : n \in \mathbf{N})$ 在闭区间 $[0, 1]$ 上处处收敛于函数 $f(x)$：如果 $x = 1$，那么 $f(x) = 1$，否则 $f(x) = 0$。但是，对任意实数 $\delta > 0$，无论取多大的正整数 N，都总存在非常接近 1 的实数 $0 < x_{\delta,N} < 1$，使得即便 $n \geq N$，仍有 $|f_n(x_{\delta,N}) - f(x_{\delta,N})| > \delta$。这些“掉队者”，“拖后腿者”总是存在。实际上，$\lim_{n\to\infty} \sup_{x \in [0,1]} |f_n(x) - f(x)| = 1$，而不是 0。所以，函数序列 (f_n) 在区间 $[0, 1]$ 上并不能够一致地收敛于 f。（感谢 dragonheart6 和 zc。）

上例也说明，如果在 $n \geq N$ 之后，不再有“拖后腿者”，即 $\{x \in A : |f_n(x) - f(x)| > \delta\} = \varnothing$，那么，$\sup_{x \in A} |f_n(x) - f(x)| \leq \delta$，所以，函数序列

$f_n(x)$ 在集合 A 上一致地收敛于函数 $f(x)$。显然，一致收敛蕴含处处收敛。在这个意义上，也说一致收敛强于 (stronger than) 处处收敛。

5.2.7 可测函数的和、差、积、商运算

如果 $f,g\colon X\to\overline{\mathbf{R}}$ 是可测的，那么 $f+g$ 和 fg 也是可测的；当 $g\neq 0$ 时，f/g 也是可测的。因为，令 $h=f+g$，那么

$$h^{-1}([-\infty,c))=\bigcup_{r\in\mathbf{Q}} f^{-1}([-\infty,c-r))\cap g^{-1}([-\infty,r)).$$

注意到：有理数集是可数的。所以此式是个完全合法的 σ-代数表达式。对任意点 $g(x_0)=t$，总存在有理数 $r\in\mathbf{Q}$，使得 $g(x_0)=t<r$ 成立。当 r 取遍 $\mathbf{Q}$ 时，存在一个有理数序列 $r_1,r_2,\ldots$，无限逼近 t。所以，上式等号成立。

如果 f 是非负的，令 $h=-f$，那么

$$h^{-1}([-\infty,c))=f^{-1}((-c,\infty]).$$

所以，如果 f 是可测的，那么 $-f$ 也是可测的。

给定函数 $f(x)$，称函数 $f^+(x)=\max(f(x),0)$ 为 f 的正部 (positive part)，$f^-(x)=\max(-f(x),0)$ 为 f 的负部 (negative part)。显然，$f=f^+-f^-$，所以，函数 f 可测当且仅当它的正部和负部皆可测。

如果 f 和 g 都是非负的，令 $h=fg$，那么，与上段同理，我们有

$$h^{-1}([-\infty,c))=\bigcup_{r\in\mathbf{Q}} f^{-1}([-\infty,c/r))\cap g^{-1}([-\infty,r)).$$

对任意的可测函数 f 和 g，有 $fg=(f^+-f^-)(g^+-g^-)=f^+g^++f^-g^--f^+g^--f^-g^+$，因此，$fg$ 可测。

如果 $g>0$，那么，令 $h=1/g$，对于任意实数 $c>0$，我们有

$$h^{-1}([-\infty,c))=g^{-1}((1/c,\infty]),$$

对于 $c\leq 0$，$h^{-1}([-\infty,c))=\varnothing$，因此，$h$ 是可测的。

对于奇点 (singular point) 处的情况，注意到，如果函数 f 可测，那么 $f^{-1}(\{\infty\})$，$f^{-1}(\{-\infty\})$，$f^{-1}(\{0\})$，都是可测集。我们可以定义函数序列 $h_1,h_2,\ldots$，使得 h_i 在这些可测集上取 i、$-i$ 或 $1/i$，在其他点上取 f。这样，$i\to\infty$ 时，$h_i\to f$。对 f 参与的上述运算，用 h_i 去逼近，可证，在扩展的实数系 $\overline{\mathbf{R}}$ 上的合法表达式的规定下，（指排除 $\infty-\infty$ 等无定义的情况，）奇点的存在不影响上述 $f+g$，$f-g$，fg，f/g 的可测性。

综合以上内容，易见前述定理成立。 □

5.2.8 勒贝格积分

现在，我们可以给出具有一般性的勒贝格积分的定义了。

对于非负的可测函数 $f\colon X \to \overline{\mathbf{R}}$，其勒贝格积分定义为

$$\int_X f\,\mathrm{d}\mu = \sup\left\{\int_X \phi\,\mathrm{d}\mu : \phi\ \text{是简单函数，且}\ 0 \le \phi \le f\right\}.$$

如果此式收敛 (converge)，即 $\int_X f\,\mathrm{d}\mu < \infty$，则称该非负可测函数 f 是可积的。

对于 $X = \mathbf{R}$ 的情况，在求 $y = f(x)$ 的黎曼积分时，我们把定义域划分成微小区间段 Δx_i，段内典型值为 $f(x_i)$，然后我们求和式 $\sum f(x_i)\Delta x_i$ 在 $\Delta x_i \to 0$ 时的极限。在求 $y = f(x)$ 的勒贝格积分时，我们把值域划分成微小区间段 Δy_i，段内典型值为 y_i。y 值在此区间内的相应的 x 的取值范围，也即该区间段的原象，是一个可测集 S_i，可能是空集，也可能是一个或多个区间，甚至可能是整个数轴。既然是可测集，就可求测度 $\mu(S_i)$，进而可求和式 $\sum y_i\mu(S_i)$。该式在 $\Delta y_i \to 0$ 时的极限，就是勒贝格积分。如果 X 上的测度取为勒贝格测度 λ，那么，只要黎曼积分有定义，勒贝格积分就有定义且与之相等。所以，勒贝格积分包含黎曼积分。

如果可测函数 $f\colon X \to \overline{\mathbf{R}}$ 有正有负，但是 $\int_X f^+\,\mathrm{d}\mu$ 和 $\int_X f^-\,\mathrm{d}\mu$ 不同时为 ∞，则 $\int_X f\,\mathrm{d}\mu$ 定义为两者之差，即 $\int_X f\,\mathrm{d}\mu = \int_X f^+\,\mathrm{d}\mu - \int_X f^-\,\mathrm{d}\mu$。

对于一般可测函数 f，如果 $\int_X |f|\,\mathrm{d}\mu < \infty$，则称 f 是可积的。因为 $|f| = f^+ + f^- \ge f^\pm$，所以，$f$ 可积当且仅当 f^+ 和 f^- 都可积。此时，因为 $f = f^+ - f^-$，所以 $-\infty < \int_X f\,\mathrm{d}\mu < \infty$，即 f 的积分收敛。

5.3 收敛定理

5.3.1 零测度的集合

如果一个可测集 S 的测度为零，即 $\mu(S) = 0$，则称它是零测度的 (measure zero) 集合，简称零测集。

直线上可数个点的长度和，平面上可数个直线的面积和，空间中可数个平面的体积和，都是零测度的例子。（感谢 DrFaust。）尤其是，如果取勒贝格测度，那么实数轴上的有理数集是零测度的，即 $\lambda(\mathbf{Q}) = 0$。

5.3.2 几乎处处为真的谓词

如果使谓词 $P(x)$ 为假的 x 的集合是零测度的，则称该谓词 $P(x)$ 几乎处处 (almost everywhere，常简记为 a.e.) 为真。

如果非负可测函数 f 的积分为零，那么它一定几乎处处为零。因为，对任意正整数 n，

$$\mu\left(f^{-1}((1/n,\infty])\right) = n\int_X \frac{1}{n} I_{f^{-1}((1/n,\infty])}\,\mathrm{d}\mu \le n\int_X f\,\mathrm{d}\mu = 0,$$

而 $f^{-1}((0,\infty]) = \bigcup_n f^{-1}((1/n,\infty])$，所以 $\mu\left(f^{-1}((0,\infty])\right) = 0$。 □

如果可测函数 f 是可积的，那么 f 本身是几乎处处有界的，即 $|f| < \infty$ 几乎处处成立。如若不然，则根据定义，f 不可能是可积的。 □

5.3.3 几乎处处收敛与依测度收敛

如果函数序列 $(f_n : n \in \mathbf{N})$ 在集合 A 上，除了一个零测度的子集 B 之外，在其他点上处处都收敛，即 $B \subseteq A$ 而且 $\mu(B) = 0$，(f_n) 在 $A \setminus B$ 上处处收敛，则称 (f_n) 在集合 A 上几乎处处收敛。该收敛的条件又可表示为

$$\mu(\{x \in A : \lim_{n\to\infty} f_n(x) = f(x)\}) = \mu(A).$$

把一致收敛的条件放宽，允许集合 $\{x \in A : |f_n(x) - f(x)| > \delta\}$ 有正测度，但是该测度必须趋于零，即 $\lim\limits_{n\to\infty} \mu\left(\{x \in A : |f_n(x) - f(x)| > \delta\}\right) = 0$，则称函数序列 $f_n(x)$ 在集合 A 上依测度 (in measure) 收敛于函数 $f(x)$。

显然，一致收敛蕴含依测度收敛。

如果集合 A 的测度有限，即 $\mu(A) < \infty$，可测函数序列 $(f_n(x))$ 和可测函数 $f(x)$ 在集合 A 上几乎处处有界，那么几乎处处收敛蕴含依测度收敛。有人称此为 Lebesgue 定理。证明从略。

依测度收敛不蕴含几乎处处收敛。例如，令 $f_1(x)$ 在 $(0,1]$ 上为 1,其他点上为零；（以下不再重复说“其他点上为零”。）$f_2(x)$ 在 $(0,1/2]$ 上为 1；$f_3(x)$ 在 $(1/2,1]$ 上为 1；$f_3(x)$ 在 $(0,1/3]$ 上为 1；$f_4(x)$ 在 $(1/3,2/3]$ 上为 1；$f_5(x)$ 在 $(2/3,1]$ 上为 1；按此模式类推，显然，如果取勒贝格测度，那么，当 $n \to \infty$ 时，$f_n(x)$ 在 $(0,1)$ 上依测度收敛于恒零函数，但是在 $(0,1)$ 上的任何一点都不收敛。（感谢 dragonheart6 提供关于此反例的网址。）

如果函数序列 $(f_n(x))$ 在集合 A 上依测度收敛于函数 $f(x)$，则存在一个子序列 $(f_{n_i}(x) : i \in \mathbf{N})$ 在 A 上几乎处处收敛于 $f(x)$。这称为 F.Riesz 定理。证明从略。

5.3.4 几乎一致收敛与叶果洛夫定理

给定可测函数序列 $(f_n(x) : n \in \mathbf{N})$，可测函数 $f(x)$ 和它们的共同定义域的子集 A。如果对于任意实数 $\delta > 0$，均存在相应的子集 B_δ，使得 $\mu(B_\delta) < \delta$，并且 $(f_n(x))$ 在 $A \setminus B_\delta$ 上一致收敛，则称函数序列 $(f_n(x))$ 在集合A上几乎一致 (almost uniformly) 收敛于函数 $f(x)$。

几乎一致收敛蕴含依测度收敛。但是，从依测度收敛只能推出存在子序列几乎一致收敛。比如前面的依测度收敛的例子，区间 $(0,1)$ 除去任何子集都无法使 $(f_n(x))$ 在其上一致收敛，所以不是几乎一致收敛。

如果 $\mu(A)<\infty$，而且 $(f_n(x))$ 和 $f(x)$ 都是几乎处处有界的，那么 $(f_n(x))$ 在 A 上几乎处处收敛，当且仅当，$(f_n(x))$ 在 A 上几乎一致收敛。这称为叶果洛夫 (Egoroff) 定理。证明从略。

5.3.5 荒点集、迟滞集和另类集

在集合 A 中，逐点考察函数序列在该点是否收敛。不妨称在其上不收敛的点为荒点，意为荒凉、无人定居的地方。称荒点的集合为荒点集。那么，

- 如果荒点集是空集，则函数序列在集合 A 上处处收敛。
- 如果荒点集是零测度的，则函数序列在集合 A 上几乎处处收敛。

对所有正数 δ，不妨称集合 $B_{\delta,n}=\{x\in A:|f_n(x)-f(x)|>\delta\}$ 为迟滞集。那么，当 $n\to\infty$ 时，

- 如果在 $n\geq N$ 之后，迟滞集恒为空集，则函数序列在集合 A 上一致收敛。
- 如果迟滞集的测度 $\mu(B_{\delta,n})$ 趋于零（不必等于零），则函数序列在集合 A 上依测度收敛。

依测度收敛对迟滞集的“位置”并没有要求。比如前面的例子，$(0,1/n]$，$(1/n,2/n]$，……，$((n-1)/n,1]$，依次作迟滞集。这导致该函数序列在迟滞集之外，实际上在整个区间 $(0,1)$ 上，的任何一点上都不收敛，更不用说一致收敛了。

如果迟滞集最终局限于某些相对固定的“位置”，当 n 足够大时，离这些位置较远的地方就不会再出现迟滞集，从而函数序列可以在这些地方的一定范围上一致收敛，

给定大整数 N，称 $A_{\delta,N}=\bigcup\limits_{n\geq N}B_{\delta,n}$ 为另类集。当 $N\to\infty$ 时，如果另类集的测度 $\mu(A_{\delta,N})$ 趋于零，则函数序列在集合 A 上几乎一致收敛。

5.3.6 渐扩的和渐缩的可测集序列

给定测度空间 $(X,\mathcal{M},\mu)$，和可测集序列 $E_1\subseteq E_2\subseteq\cdots$，称序列 $(E_n:n\in\mathbf{N})$ 是渐扩的 (increasing)。记 $E=\bigcup\limits_{n=1}^{\infty}E_n$，称序列 (E_n) 扩至 (increase) E。记

为 $E_n \nearrow E$。易见 $\mu(E)=\lim_{n\to\infty}\mu(E_n)$。称 $\mu(E)$ 是从下连续的 (continuity from below)。

给定测度空间 $(X,\mathcal{M},\mu)$，和可测集序列 $E_1\supseteq E_2\supseteq\cdots$，称序列 $(E_n: n\in\mathbf{N})$ 是渐缩的 (decreasing)。如果其中有某个 E_k 满足 $\mu(E_k)<\infty$，那么，记 $E=\bigcap_{n=1}^{\infty}E_n$，称序列 E_n 缩至 (decrease) E。记为 $E_n\searrow E$。易见 $\mu(E)=\lim_{n\to\infty}\mu(E_n)$。称 $\mu(E)$ 是从上连续的 (continuity from above)。

5.3.7 单调收敛定理

单调收敛定理 (Monotone Convergence Theorem，简称 MCT)：如果非负可测函数序列 f_n 是处处 (pointwise) 单调增地趋近于函数 f 的，即对任何给定的 $x=x_0$，有 $f_1(x_0)\le f_2(x_0)\le\cdots\le f_n(x_0)\to f(x_0)$，那么，$f$ 是可测的，并且

$$\int_X f\,\mathrm{d}\mu=\int_X\lim_{n\to\infty}f_n\,\mathrm{d}\mu=\lim_{n\to\infty}\int_X f_n\,\mathrm{d}\mu.$$

因为，首先，由于 f_n 以 f 为上界，所以 $\lim_{n\to\infty}\int_X f_n\,\mathrm{d}\mu\le\int_X f\,\mathrm{d}\mu$。

其次，给定任意简单函数 $0\le\varphi\le f$，和实数 $0<t<1$，考虑集合 $A_n=\{x: f_n(x)\ge t\varphi\}$，因为 $f(x)\ge\varphi(x)>t\varphi(x)$，而 $f_n(x)\to f(x)$，所以，对任何 $x\in X$，存在相应的 n_x，使得 $f_{n_x}(x)\ge t\varphi(x)$，即 $x\in A_{n_x}$。故而 $\bigcup_n A_n=X$。那么 $\int_X t\varphi\,\mathrm{d}\mu\le\lim_{n\to\infty}\int_X f_n\,\mathrm{d}\mu$。令 $t\nearrow 1$，则得 $\int_X\varphi\,\mathrm{d}\mu\le\lim_{n\to\infty}\int_X f_n\,\mathrm{d}\mu$。对所有可能的简单函数 $0\le\varphi\le f$ 的积分求上确界 sup，即得 $\int_X f\,\mathrm{d}\mu\le\lim_{n\to\infty}\int_X f_n\,\mathrm{d}\mu$。

两个结果做对比，可见只能是等号成立。 □

该定理具有基础的重要性，以至于有些教材干脆用这个极限作为一般函数的勒贝格积分的定义。它也是取极限与求积分可交换的一个实例。

5.3.8 法图引理与控制收敛定理

法图引理 (Fatou's lemma)：在与 MCT 相同条件下，有

$$\int_X\liminf_n f_n\,\mathrm{d}\mu\le\liminf_n\int_X f_n\,\mathrm{d}\mu.$$

因为，令 $g_n=\inf_{k\ge n}f_k$，则 $g_n\le f_n$，令 $h=\liminf_n f_n$，则 $g_n\nearrow h$。代入单调收敛定理，有

$$\int_X\liminf_n f_n\,\mathrm{d}\mu=\int_X h\,\mathrm{d}\mu=\lim_n\int_X g_n\,\mathrm{d}\mu\le\liminf_n\int_X f_n\,\mathrm{d}\mu.\quad\square$$

控制收敛定理 (Dominated Convergence Theorem，简称 DCT。也称勒贝格定理)：给定可测函数序列 $f_n\colon X \to \overline{\mathbf{R}}$ 处处收敛于 f，如果存在一个可积函数 g，满足对所有 n 均有 $|f_n| \le g$，那么，f_n 和 f 都是可积的，而且

$$\lim_{n\to\infty}\int_X |f_n - f|\,\mathrm{d}\mu = 0.$$

因为，由法图引理，

$$\begin{aligned}\int_X \liminf_n (2g - |f_n - f|)\,\mathrm{d}\mu &\le \liminf_n \int_X (2g - |f_n - f|)\,\mathrm{d}\mu \\ &= \int_X 2g\,\mathrm{d}\mu + \liminf_n \left(-\int_X |f_n - f|\,\mathrm{d}\mu\right) \\ &= \int_X 2g\,\mathrm{d}\mu - \limsup_n \int_X |f_n - f|\,\mathrm{d}\mu.\end{aligned}$$

该方程左边等于 $\displaystyle\int_X 2g\,\mathrm{d}\mu$，所以

$$\limsup_n \int_X |f_n - f|\,\mathrm{d}\mu \le 0, \quad \text{即} \quad \lim_{n\to\infty}\int_X |f_n - f|\,\mathrm{d}\mu = 0. \quad \square$$

5.3.9 广义三角不等式、取极限与求积分的可交换性

广义三角不等式 (Generalized Triangle Inequality)：给定可积函数 $f\colon X \to \overline{\mathbf{R}}$，均有 $\left|\displaystyle\int_X f\,\mathrm{d}\mu\right| \le \displaystyle\int_X |f|\,\mathrm{d}\mu$。因为，从 $-|f| \le f \le |f|$ 显见 $-\displaystyle\int_X |f|\,\mathrm{d}\mu \le \displaystyle\int_X f\,\mathrm{d}\mu \le \displaystyle\int_X |f|\,\mathrm{d}\mu$。 $\square$

根据广义三角不等式，在 DCT 的条件下，$\left|\displaystyle\lim_{n\to\infty}\int_X (f_n - f)\,\mathrm{d}\mu\right| = 0$，即 $\displaystyle\lim_{n\to\infty}\int_X f_n\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu = \int_X \lim_{n\to\infty} f_n\,\mathrm{d}\mu$。这是取极限与求积分可交换的另一个实例。注意：虽然在许多应用问题中，控制函数 (dominating function) g 的存在性是自然满足的，但是，在概念上，我们不能忽略这个条件。

通常，某种研究对象的集合，比如点集、可积函数集、连续函数集等，如果关于极限运算封闭，则称它是完备的 (complete)。满足 MCT（或 DCT）条件的勒贝格可积函数集是完备的，但黎曼可积函数集不是。

如果控制函数 g 取常数，那么控制收敛定理就成了有界收敛定理 (Bounded Convergence Theorem，简称 BCT)。

5.3.10 求和与求积分的可交换性

Beppo-Levi 定理：如果可测函数序列 $f_n\colon X \to \overline{\mathbf{R}}$ 的绝对和函数是可积的，即 $\displaystyle\int_X \sum_{n=1}^{\infty} |f_n|\,\mathrm{d}\mu < \infty$，那么，它的和函数也是可积的，并且和函数的积

分等于各函数的积分的和，即

$$\sum_{n=1}^{\infty}\int_X f_n\,\mathrm{d}\mu=\int_X\sum_{n=1}^{\infty}f_n\,\mathrm{d}\mu.$$

因为，令 $g_N=\sum_{n=1}^N f_n$，$h_N=\sum_{n=1}^N|f_n|$，$h=\sum_{n=1}^{\infty}|f_n|$，显然，$g_N\le h_N\le h$，而 h 是可积的，所以，

$$\begin{aligned}\sum_{n=1}^{\infty}\int_X f_n\,\mathrm{d}\mu &= \lim_{N\to\infty}\sum_{n=1}^{N}\int_X f_n\,\mathrm{d}\mu\\ &= \lim_{N\to\infty}\int_X g_N\,\mathrm{d}\mu\\ &= \int_X\lim_{N\to\infty}g_N\,\mathrm{d}\mu\\ &= \int_X\sum_{n=1}^{\infty}f_n\,\mathrm{d}\mu.\quad\square\end{aligned}$$

这个也常称为求和与求积分的可交换性。当然，这里的求和，指可数个项目的求和，而不是普通的加法运算。

5.4 质量测度与拖回测度

5.4.1 质量密度函数与质量测度

测度空间 $(X,\mathcal{A},\mu)$ 上的任何非负的可测函数 $g\colon X\to[0,\infty]$ 都确定一个相应的测度

$$\nu(S)=\int_S g\,\mathrm{d}\mu,\quad S\in\mathcal{A}.$$

称函数 g 是测度 ν 在测度 μ 上的质量密度函数，称测度 ν 为在测度 μ 上按密度 g 生成的质量测度，记为 $\mathrm{d}\nu=g\,\mathrm{d}\mu$。因为，首先，$\nu(\varnothing)=0$ 为显然。其次，考虑任意集合 S 的任意可数划分 $\{S_n\}$，有 $I(S)=\sum_{n=1}^{\infty}I(S_n)$，而且

$$\nu(S)=\int_S g\,\mathrm{d}\mu=\int_X gI(S)\,\mathrm{d}\mu=\sum_{n=1}^{\infty}\int_X gI(S_n)\,\mathrm{d}\mu=\sum_{n=1}^{\infty}\nu(S_n).\quad\square$$

注意到，μ 在 μ 上的质量密度函数为 $g\equiv 1$，这就是前面讲的第一个勒贝格积分 $\mu(S)=\int_S\mathrm{d}\mu$。

如果 ν 是如上定义的 $(X,\mathcal{A})$ 上的测度，那么对任何可测函数 $f\colon X\to\overline{\mathbf{R}}$，有 $\int_X f\,\mathrm{d}\nu=\int_X fg\,\mathrm{d}\mu$。因为，如果 f 是非负简单函数，那么此关系就是积分的线性性质，前面已证。如果 f 是一般的非负可测函数，那么，我们可以用一个简单函数序列 ϕ_n 逐步扩至 f，即 $\phi_n\nearrow f$。此时，$\phi_n g\nearrow fg$。由 MCT，

$$\int_X f\,\mathrm{d}\nu=\lim_{n\to\infty}\int_X\phi_n g\,\mathrm{d}\mu=\int_X\lim_{n\to\infty}\phi_n g\,\mathrm{d}\mu=\int_X fg\,\mathrm{d}\mu.$$

如果 f 不是非负的，只要利用 $f=f^+-f^-$ 分别处理即可。 $\square$

5.4.2 拖回测度与换元法

给定测度空间 $(X, \mathcal{A}, \mu)$ 和可测空间 $(Y, \mathcal{B})$，如果函数 $g\colon X \to Y$ 和 $f\colon Y \to \overline{\mathbf{R}}$ 可测，那么我们可以为 Y 配置这样的测度 $\forall B \in \mathcal{B} : \nu(B) = \mu(g^{-1}(B))$，称为输运测度 (transport measure)，又称拖回测度 (pullback measure)。后者更形象些：把 B 从 Y 中沿着映射 g “拖回”到 X 中，求（其原象的）测度，作为 B 本身的测度。这一关系又常记为 $\mathrm{d}\nu = \hat{g}(\mathrm{d}\mu)$。

按照拖回测度的定义，显有 $\displaystyle\int_Y f\,\mathrm{d}\nu = \int_X (f \circ g)\ \hat{g}(\mathrm{d}\mu)$，称为换元法 (Change of Variables)。为了看得更清楚，可把变元明确地写出来，即

$$\int_Y f(y)\,\mathrm{d}\nu(y) = \int_X f(g(x))\ \hat{g}(\mathrm{d}\mu(x)).$$

注意，由于 $\mathrm{d}\mu$ 和 $\mathrm{d}\nu$ 都只是积分表达式中的形式符号，所以，$\hat{g}$ 也不是任何具体的映射，而只是从形式上提示我们，从 $\mathrm{d}\mu$ 到 $\mathrm{d}\nu$ 需要做某种变换。至于具体做怎样的变换，那要视可测函数 $g\colon X \to Y$ 的具体情况而定。

举例来说。设 $X = Y = \mathbf{R}$；当 $-1 \le x \le 1$ 时，$y = g(x) = x^2$，否则 $y = g(x) = 0$；当 $0 \le y \le 1$ 时，$f(y) = y$，否则 $f(y) = 0$；测度 μ 为勒贝格测度。试应用上述换元法进行分析。

先看 $g(x)$ 的正半支 $g_+(x)$：当 $0 \le x \le 1$ 时，$y = g_+(x) = x^2$，否则 $y = g_+(x) = 0$。这种情况下，Y 上任意区间 $B \subseteq [0,1]$ 沿着 g_+ 被“拖回”到 X 上求测度。${g_+}^{-1}(B) \subseteq [0,1]$。因为 μ 为勒贝格测度，所以，相应的勒贝格积分等于黎曼积分，

$$\int_X f(g_+(x))\ \hat{g}_+(\mathrm{d}\mu(x)) = \int_0^1 x^2 2x\,\mathrm{d}x = 2\int_0^1 x^3\,\mathrm{d}x = \frac{2}{4}x^4\Big|_0^1 = \frac{1}{2}.$$

也就是说，此处 $\mathrm{d}\nu(y) = \hat{g}_+(\mathrm{d}\mu(x)) = 2x\,\mathrm{d}\mu(x) = |2x|\,\mathrm{d}\mu(x)$。

再看 $g(x)$ 的负半支 $g_-(x)$：当 $-1 \le x \le 0$ 时，$y = g_-(x) = x^2$，否则 $y = g_-(x) = 0$。这种情况下，Y 上任意区间 $B \subseteq [0,1]$ 沿着 g_- 被“拖回”到 X 上求测度。${g_-}^{-1}(B) \subseteq [-1,0]$。

$$\int_X f(g_-(x))\ \hat{g}_-(\mathrm{d}\mu(x)) = \int_{-1}^0 x^2(-2x)\,\mathrm{d}x = -2\int_{-1}^0 x^3\,\mathrm{d}x = -\frac{2}{4}x^4\Big|_{-1}^0 = \frac{1}{2}.$$

也就是说，此处 $\mathrm{d}\nu(y) = \hat{g}_-(\mathrm{d}\mu(x)) = -2x\,\mathrm{d}\mu(x) = |2x|\,\mathrm{d}\mu(x)$。

现在来看 $g(x)$。Y 上任意区间 $B \subseteq [0,1]$ 沿着 g 被“拖回”到 X 上求测度。$g^{-1}(B) = {g_+}^{-1}(B) \cup {g_-}^{-1}(B)$。因此，$\nu(B) = \mu({g_+}^{-1}(B)) + \mu({g_-}^{-1}(B))$。进而，

$$\int_Y f(y)\,\mathrm{d}\nu(y) = \int_X f(g_+(x))\ \hat{g}_+(\mathrm{d}\mu(x)) + \int_X f(g_-(x))\ \hat{g}_-(\mathrm{d}\mu(x)) = 1.$$

注意，如果 ν 不是上述拖回测度，而是通常的勒贝格测度，那么

$$\int_Y f(y)\,\mathrm{d}\nu(y) = \int_0^1 y\,\mathrm{d}y = \frac{1}{2}y^2\Big|_0^1 = \frac{1}{2}.$$

可见，做勒贝格积分时，一定要明白是用什么测度来做这个积分。 □

Chapter 6

概率论

6.1　古典概率与计数问题

6.1.1　概率空间

在概率论中，可测空间 $(X, \mathcal{A})$ 常称为样本空间 (sample space)，记为 $(\Omega, \mathcal{E})$。Ω 中的元素称为样本点 (sample point)，常记为 ω。$\mathcal{E}$ 中的元素称为事件 (event)。设两事件 $E_1, E_2 \in \mathcal{E}$，则 $E_1 \cup E_2$ 表示事件 E_1 和 E_2 至少有一个发生，称为 E_1 和 E_2 的和；$E_1 \cap E_2$ 表示事件 E_1 和 E_2 都发生，称为 E_1 和 E_2 的积；${E_1}^{\complement}$ 表示事件 E_1 不发生，称为 E_1 的对立事件，或者逆事件。

事件 $E \in \mathcal{E}$ 的概率 (probability) $p(E)$ 是一种特殊的测度，它满足 $p(\Omega) = 1$。样本空间 $(\Omega, \mathcal{E})$ 配上概率 p，就构成概率空间 (probability space)，记为 $(\Omega, \mathcal{E}, p)$。

6.1.2　古典概率

古典概率问题通常涉及一组有限的 (finite)、完备的 (complete)、互斥的 (disjoint)、等可能的事件，称为基本事件 (basic event)。所有基本事件的集合就是 Ω，而一般事件 E 的概率就定义为 $p(E) = \dfrac{|E|}{|\Omega|}$，其中的 $|E|$ 是可测集 E 的计数测度。

例如，标准的六面骰子，掷出的点数超过四点的概率有多大？这里所有基本事件的集合为 $\Omega = \{1, 2, 3, 4, 5, 6\}$，“掷出的点数超过四点”这一事件为 $E = \{5, 6\}$，因而，概率 $p(E) = \dfrac{|E|}{|\Omega|} = \dfrac{1}{3}$。　□

6.1.3 排列与组合

从 n 个不同元素中选出 m 个，排成有次序的一列，称为一个排列 (arrangement)。可以如此构做的互不相同的排列的个数，称为 n 中选 m 的排列数，记为 A_n^m。

从 n 个不同元素中选出 m 个，并成无次序的一组，称为一个组合 (combination)。可以如此构做的互不相同的组合的个数，称为 n 中选 m 的组合数，记为 C_n^m。

显见，如上定义的排列是一种特殊的 m 元组，或曰长度为 m 的列表，其中的元素不会重复，因此常把排列记为 $(a_{i_1},\ldots,a_{i_m})$；如上定义的组合就是包含 m 个元素的子集，因此常把组合记为 $\{a_{i_1},\ldots,a_{i_m}\}$。

6.1.4 计数问题中的加法原理和乘法原理

如果一个集合 A，可以划分成两个集合 B 和 C，即 $B\cap C=\varnothing$ 并且 $A=B\cup C$，那么 $|A|=|B|+|C|$。这是计数问题中的加法原理。

如果一个序偶的集合 $A=\{(a_i,b_{ij}):1\le i\le n,1\le j\le m\}$，其中的序偶按第一个元素 a_i 分类，即，凡第一个元素相同的都分到同一类，可以分 n 类，而每一类都含有 m 个元素，那么 $|A|=n\times m$。这是计数问题中的乘法原理。

6.1.5 全排数、排列数和组合数公式

给定 n 个不同元素组成的集合 $A=\{a_1,a_2,\ldots,a_n\}$，考虑它的 n 中选 n 排列，也称全排列，简称全排。把所有排列按第一个元素分类，显然有 n 类。选定一类，设这一类中第一个元素为 a_{i_1}，那么，此类排列都切除第一个元素，恰好就是 $A\setminus\{a_{i_1}\}$ 的全排列。因此，$A_n^n=n\times A_{n-1}^{n-1}$。同理，$A_{n-1}^{n-1}=(n-1)\times A_{n-2}^{n-2}$。依此类推，易见 $A_n^n=n\times(n-1)\times\cdots\times 2\times 1=n!$。

类似地，考虑上述集合 A 中元素的所有全排列。把前 m 个元素依次都相等的分为一类，按定义，有 A_n^m 类。选定一类，设其中排列的前 m 个元素依次为 $a_{i_1},\ldots,a_{i_m}$，那么，此类排列都切除前 m 个元素，恰好就是 $A\setminus\{a_{i_1},\ldots,a_{i_m}\}$ 的全排列。因此，$A_n^n=A_n^mA_{n-m}^{n-m}$，即 $A_n^m=\dfrac{A_n^n}{A_{n-m}^{n-m}}=\dfrac{n!}{(n-m)!}$。

在 n 中选 m 的所有排列中，把如果不论元素次序，则可看做由相同元素构成的排列分为一类，按定义，每一类都对应且只对应一个组合：设某类中有一排列为 $(a_{i_1},\ldots,a_{i_m})$，那么，$a_{i_1},\ldots,a_{i_m}$ 的任意排列，即这 m 个元素的全排列，有 A_m^m 个，都在此类中，而且这个类对应着组合 $\{a_{i_1},\ldots,a_{i_m}\}$。这样，$n$ 中选 m 的所有排列被分成 C_n^m 类，每类包含 A_m^m 个排列，所以 $A_m^mC_n^m=A_n^m$，即：$C_n^m=\dfrac{A_n^m}{A_m^m}=\dfrac{A_n^n}{A_{n-m}^{n-m}A_m^m}=\dfrac{n!}{(n-m)!m!}$。

6.1.6 分划排列、分划组合与分划全排问题

记上段的 m 为 m_1。如果对给定的一个组合 $\{a_{i_1}, \ldots, a_{i_{m_1}}\}$，从剩下的 $n-m_1$ 个元素中取 m_2 个元素做成组合，则有 $C_{n-m_1}^{m_2}$ 种不同的取法。选定一种取法，剩下的 $n-m_1-m_2=m_3$ 个元素自然做成一个组合。这实际上是把集合 $A=\{a_1, a_2, \ldots, a_n\}$ 划分成互斥的三个子集 A_1, A_2, A_3，它们分别含有 m_1, m_2, m_3 个元素。由于“第一步”，“第二步”等都是有次序的，所以按照该过程，我们得到一个子集的排列 (A_1, A_2, A_3)。根据乘法原理，不同的结果有 $C_n^{m_1}C_{n-m_1}^{m_2}C_{n-m_1-m_2}^{m_3}=\dfrac{A_n^n}{A_{m_1}^{m_1}A_{m_2}^{m_2}A_{m_3}^{m_3}}=\dfrac{n!}{m_1!\,m_2!\,m_3!}$ 种。

同理，把集合 $A=\{a_1, a_2, \ldots, a_n\}$ 划分成互斥的 t 个子集 $A_1, A_2, \ldots, A_t$，它们分别含有 $m_1, m_2, \ldots, m_t$ 个元素，那么，可做的子集排列 $(A_1, A_2, \ldots, A_t)$ 一共有 $C_n^{m_1}C_{n-m_1}^{m_2}\cdots C_{m_t}^{m_t}=\dfrac{A_n^n}{A_{m_1}^{m_1}A_{m_2}^{m_2}\cdots A_{m_t}^{m_t}}=\dfrac{n!}{m_1!\,m_2!\,\cdots\,m_t!}$ 种。这个问题称为分划排列问题，简记为 $(m_1, m_2, \ldots, m_t)$。显然，组合数公式就是 $(m, n-m)$ 问题；全排列公式就是 $(1, 1, \ldots, 1)$ 问题，其中共有 n 个 1；排列数公式就是 $(1, \ldots, 1, n-m)$ 问题，其中共有 m 个 1。

分划排列问题的解的个数公式可以这样理解：由于 A_1 是集合，在集合 $A=\{a_1, a_2, \ldots, a_n\}$ 的元素的所有全排列中，A_1 的元素任意交换位置，简称置换 (permutation)，所产生的排列都对应同一个分划排列结果。这一性质称为组内置换对称性。把 A_n^n 个全排列，按照 $A_1, \ldots, A_t$ 的组内置换对称性分类，那么，每一类就对应一个分划排列结果。

条件与分划排列问题相同，问可做的组合 $\{A_1, A_2, \ldots, A_t\}$ 有多少种？这个问题称为分划组合问题，也称分堆问题，简记为 $\{m_1, m_2, \ldots, m_t\}$。

“堆”和“组”这样的概念，本身是无差别、无次序的，但是，你可以用“第一堆”、“第二堆”的指派，给它们贴上标签，赋予次序。举例来说，“有一堆是 $\{a, b\}$，还有一堆是 $\{c, d\}$”，与“有一堆是 $\{c, d\}$，还有一堆是 $\{a, b\}$”，描述的是同一个分划组合结果。但是，“第一堆是 $\{a, b\}$，第二堆是 $\{c, d\}$”，与“第一堆是 $\{c, d\}$，第二堆是 $\{a, b\}$”，描述的是两个不同的分划排列结果。

一个典型的分堆问题可能这样说：有三堆是每堆一个元素的，有两堆是每堆三个元素的。即，这是 $\{11133\}$ 问题。由于数字简单，我们省略了其中的逗号。显然，含一个元素的三个堆可以任意交换位置，含三个元素的两个堆可以交换位置。这称为组间置换对称性。如果 (11133) 问题的解有 P 个，那么，考虑到组间置换对称性，$\{11133\}$ 问题的解有 $N=\dfrac{P}{A_3^3A_2^2}$ 个。

实践上，分堆之后还要分给人、学校、旅游团等，典型的问题可能说：有三人是分到一本书的，有两人是分到三本书的。这里，所有的书都不相同。显然，分堆问题的每个解都是一个集合，集合中元素的每个全排列，都是这种问题的一个解。而且，这样产生的解不会有重复的。因此，如果分堆问题的解有

N 个，那么这种问题的解就有 $T = N \times A_t^t$ 个。正是由于这个特点，把这种问题称为分划全排问题，简记为 $\{m_1, m_2, \ldots, m_t\}!$。

也有人把分划排列问题称为定向分配问题，把分划全排问题称为不定向分配问题。

有个有趣的巧合：我们都知道，六本不同的书，甲、乙、丙三人各分两本，问分法几何，和每人分两本，问分法几何，答案应该是一样的。可是，概念上讲，前者是 (222) 问题，后者是 $\{222\}!$ 问题，答案怎么会一样呢？原来，这里的组间置换对称性是完全的。设 (222) 问题的解有 P 个，那么 $\{222\}$ 问题的解有 $N = \dfrac{P}{A_3^3}$ 个，进而 $\{222\}!$ 问题的解有 $T = N \times A_3^3 = P$ 个，组间置换对称性与全排列，两个因素所产生的效果恰好相互抵消。由此，有人把这类问题单列出来，称为均匀分配问题。

6.1.7 正整数的非负分拆问题

以上讨论中，$m_1, \ldots, m_t$ 是给定的。如果它们不是给定的呢？这就涉及到正整数 n 的非负分拆问题：$m_1 + \cdots + m_t = n$ 的非负整数解有多少个？简记为 t^+n 问题。

在 n 个 1 中掺进 $t-1$ 个 0，打乱顺序任意排，得到一个字符串，其中的 $t-1$ 个 0 把整个字符串分成 t 段，每一段中 1 的个数，就是相应的 m_i。例如 11001110，表示 $2+0+3+0=5$。这是 4^+5 问题的一个解。

从集合 $B = \{b_1, \ldots, b_{n+t-1}\}$ 中选取 n 个元素贴上标签 1，其他的都贴上标签 0，这样，排列 $(b_1, \ldots, b_{n+t-1})$ 的标签恰构成上述的字符串。于是，一个组合对应且只对应一个字符串，进而对应且只对应一个分拆方案。所以，$m_1 + \cdots + m_t = n$ 的非负整数解有 C_{n+t-1}^n 个。 □

例如，三本不同的书，三个人分，有多少种分法？

若是三本相同的书，那么，这就是 3^+3 问题，有 $C_{3+3-1}^3 = 10$ 个解。即：$(111), (012), (021), (102), (120), (201), (210), (300), (030), (003)$。

可这是三本不同的书牙！那就是十个分划排列问题的和。再看一下组间置换对称性，不难发现这还是三个分划全排问题，即 $\{111\}!$，$\{012\}!$，$\{003\}!$。因此，有 $\dfrac{A_3^3}{A_1^1 A_1^1 A_1^1} \dfrac{A_3^3}{A_3^3} + \dfrac{A_3^3}{A_0^0 A_1^1 A_2^2} \dfrac{A_3^3}{A_1^1 A_1^1 A_1^1} + \dfrac{A_3^3}{A_0^0 A_0^0 A_3^3} \dfrac{A_3^3}{A_2^2 A_1^1} = 27$ 种分法。 □

再例如，63 本不同的书，甲、乙、丙三个人分，甲至少分 10 本，乙至少分 20 本，丙至少分 30 本，问分法几何？

那些无论如何都必须分到的数目，称为底数。扣去所有底数，还剩三本可以自由分配，方案有 10 种，前面已经列出。然后，把底数加回去，就是所求的方案了，即：

$$(11,21,31), (10,21,32), (10,22,31), (11,20,32), (11,22,30), (12,20,31),$$
$$(12,21,30), (13,20,30), (10,23,30), (10,20,33).$$

很明显，由于指派不同的底数，破坏了组间置换对称性，这个题现在只能当作十个分划排列问题的和来求解了。 □

6.1.8 超几何分布

从 N 个白色的球和 M 个不是白色的球中，任取 n 个球，其中恰有 k 个白球，问取法几何。

先从 N 个白球中取 k 个，有 C_N^k 种取法。再从 M 个不是白色的球中取 $n-k$ 个，有 C_M^{n-k} 种取法。按乘法原理，共有 $C_N^k C_M^{n-k}$ 种取法。

从 $N+M$ 个球中任取 n 个，取法有 C_{N+M}^n 种。显见，如此取出的球中恰有 k 个白球的概率是 $p(k)=\dfrac{C_N^k C_M^{n-k}}{C_{N+M}^n}$。这称为超几何分布 (hypergeometric distribution)。该分布常见于产品抽检问题：白球相当于产品中的次品，从一个批次中抽检 n 个，发现其中有 k 个次品的概率恰符合该分布。 □

6.1.9 与统计物理有关的三个计数问题

麦克斯韦-波尔兹曼模型 (Maxwell-Boltzmann model)：r 个可识别的球，n 个足够大的箱子，问装法几何。

每个球都可以选择装到任何一个箱子里，因此有 n^r 种装法。 □

玻色-爱因斯坦模型 (Bose-Einstein model)：r 个不可识别的球，n 个足够大的箱子，问装法几何。

这是 n^+r 问题，有 C_{r+n-1}^r 种装法。 □

费米-狄拉克模型 (Fermi-Dirac model)：r 个不可识别的球，n 个最多只能装一个球的箱子，问装法几何。

从 n 个箱子中选 r 个来装球即可。有 C_n^r 种装法。 □

6.2 概率与条件概率

6.2.1 概率的加法定理与乘法定理

如果 E_1,E_2 是互斥的可测集，那么对任何测度 μ 均有 $\mu(E_1\cup E_2)=\mu(E_1)+\mu(E_2)$。计数测度当然也不例外。即，对于互斥的事件 E_1,E_2，我们有 $|E_1\cup E_2|=|E_1|+|E_2|$。两边同除以 $|\Omega|$，得 $p(E_1\cup E_2)=p(E_1)+p(E_2)$。这就是概率的加法定理。

定义 $p(B:A)=\dfrac{|B\cap A|}{|A|}$ 为条件概率 (conditional probability)。

显有 $p(A\cap B)=\dfrac{|A\cap B|}{|\Omega|}=\dfrac{|A|}{|\Omega|}\dfrac{|A\cap B|}{|A|}=p(A)p(B:A)$，这就是概率的乘法定理。

6.2.2 全概率公式与 Bayes 公式

如果事件 A 包含于一组互斥事件 $B_1, \ldots, B_n$ 的并，即 $A \subseteq B_1 \cup \cdots \cup B_n$，那么 $|A| = |A \cap B_1| + \cdots + |A \cap B_n|$，显见

$$\begin{aligned} p(A) &= p(A \cap B_1) + \cdots + p(A \cap B_n) \\ &= p(B_1)p(A : B_1) + \cdots + p(B_n)p(A : B_n). \end{aligned}$$

这称为全概率公式 (the law of total probability)。

如果条件同上段，那么 $p(A \cap B_i) = p(A)p(B_i : A) = p(B_i)p(A : B_i)$，所以

$$p(B_i : A) = \frac{p(B_i)p(A : B_i)}{p(A)} = \frac{p(B_i)p(A : B_i)}{p(B_1)p(A : B_1) + \cdots + p(B_n)p(A : B_n)}.$$

这称为 Bayes 公式。注意该方程的右手边，其分子是分母中的一项。这其实来源于

$$\frac{|A \cap B_i|}{|A|} = \frac{|A \cap B_i|}{|A \cap B_1| + \cdots + |A \cap B_n|}.$$

这个式子是很容易理解、很容易记忆的。

6.2.3 产品抽检问题

Bayes 公式的一个典型应用是产品抽检。

例如，某工厂每十个产品为一批次，假设任意批次中无次品，有一个次品和有两个次品的概率分别为 $0.6, 0.3, 0.1$，现在，每个批次抽检一个产品，若发现是次品，则整个批次退回，求：通过检验的批次按次品数目的概率分布。

记一个批次中无次品、有一个次品和有两个次品的事件分别为 B_0, B_1, B_2，一个批次通过检验的事件为 A。这样，$p(B_0) = 0.6, p(B_1) = 0.3, p(B_2) = 0.1$，无次品的批次通过检验的概率为 $p(A : B_0) = 10/10 = 1$，有一个次品的批次通过检验的概率为 $p(A : B_1) = 9/10 = 0.9$，有两个次品的批次通过检验的概率为 $p(A : B_2) = 8/10 = 0.8$，任意批次通过检验的概率为 $p(A) = 0.6 \times 1 + 0.3 \times 0.9 + 0.1 \times 0.8 = 0.95$。在通过检验的批次中，无次品的概率为 $p(B_0 : A) = 0.6 \times 1/0.95 = 0.632$，有一个次品的概率为 $p(B_1 : A) = 0.3 \times 0.9/0.95 = 0.284$，有两个次品的概率为 $p(B_2 : A) = 0.1 \times 0.8/0.95 = 0.084$。对比一下：

批次中的次品数	0	1	2
抽检前的概率	0.6	0.3	0.1
抽检后的概率	0.632	0.284	0.084

可见，产品的质量的确有所提升。

这是因为，有一个次品的批次被退回的概率为 $p(A^{\complement} : B_1) = 1/10 = 0.1$，有两个次品的批次被退回的概率为 $p(A^{\complement} : B_2) = 2/10 = 0.2$，而无次品的批次

不会被退回，所以，经过抽检，有次品的批次被退回了一部分，它们所占比重减小了，无次品的批次所占比重增加了。这些直观上的观念与概率是一致的。 □

在 Bayes 公式的各种应用中，$p(B_i)$ 通常都不是实测的，完全可靠的。比如上述抽检问题。如果我们实际测出了 $p(B_i)$，那就没有必要再抽检了，直接说这批产品合格或者不合格就行了。如果 $p(B_i)$ 不是实测的，不是可靠的，那么，它们该取什么样的值呢？回答是：在具体问题中，$p(B_i)$ 总是有大致稳定的分布规律，因此，我们可以凭经验对 $p(B_i)$ 做出初步估计，称为抽检前的假设概率。$p(B_i : A)$ 则称为抽检后的假设概率。通过比较抽检前后的假设概率，可以评价抽检方案在提高产品质量方面的效用。

于是有人会说，直观上，抽检应该使产品质量大幅提高才对，为什么这个例子中产品质量提升不明显呢？

前面讲过，抽检提高产品质量是通过退回有次品的批次来实现的。上例中，任意批次被退回的概率为 $p(A^{\mathsf{C}}) = 0.6 \times 0 + 0.3 \times 0.1 + 0.1 \times 0.2 = 0.05$，平均二十批才退一批，所以，对产品质量的影响有限。

如果假设任意批次中无次品，有一个次品和有两个次品的概率分别为 $0.1, 0.1, 0.8$，那么，$p(A) = 0.1 \times 1 + 0.1 \times 0.9 + 0.8 \times 0.8 = 0.83$。进而，在通过检验的批次中，无次品的概率为 $p(B_0 : A) = 0.1 \times 1/0.83 = 0.121$，有一个次品的概率为 $p(B_1 : A) = 0.1 \times 0.9/0.83 = 0.108$，有两个次品的概率为 $p(B_2 : A) = 0.8 \times 0.8/0.83 = 0.771$。对比抽检前后的概率分布：

批次中的次品数	0	1	2
抽检前的概率	0.1	0.1	0.8
抽检后的概率	0.121	0.108	0.771

任意批次被退回的概率为 $p(A^{\mathsf{C}}) = 0.1 \times 0 + 0.1 \times 0.1 + 0.8 \times 0.2 = 0.17$，平均二十批退 3.4 批，所以，对产品质量的影响有所增大，但还是很有限。 □

有人说，直观上，一个批次中有两个次品的概率为 0.8，好象八成产品不合格，完全是垃圾了，为什么抽检还没有明显效果呢？

按上述假设，取十个批次，可以合理预期有一个批次是无次品的，一个批次是有一个次品的，八个批次是有两个次品的。这就是说，一百个产品中有十七个次品，17% 的产品不合格。情况虽然算不上好，但是，毕竟与 80% 的产品不合格不是一回事儿。 □

抽检对产品质量的影响还取决于批次中次品数量的差异程度。如果没有差异，比如每批总有且只有一个次品，或者每批总有且只有九个次品，那么，抽检中，前者退回的少，后者退回的多，但是两者的产品质量在抽检前后都不发生任何改变。好在现实情况是，批次中次品数量不可能没有差异。所以，抽检还是有意义的。

要想增大抽检对产品质量的影响，可以提高抽检率。比如，每个批次抽检两个产品，只要发现有次品，整个批次退回。我们还是假设任意批次中无次品，有一个次品和有两个次品的概率分别为 $0.6, 0.3, 0.1$，求：按新的抽检方案，通过检验的批次按次品数目的概率分布。

$p(A : B_0) = 1$，$p(A : B_1) = 1 - \dfrac{C_9^1 C_1^1}{C_{10}^2} = 0.8$，$p(A : B_2) = 1 - \dfrac{C_2^1 C_8^1 + C_2^2 C_8^0}{C_{10}^2} = 0.622$，$p(A) = 0.6 \times 1 + 0.3 \times 0.8 + 0.1 \times 0.622 = 0.902$，$p(B_0 : A) = 0.6 \times 1/0.902 = 0.665$，$p(B_1 : A) = 0.3 \times 0.8/0.902 = 0.266$，$p(B_2 : A) = 0.1 \times 0.622/0.902 = 0.069$。对比如下：

批次中的次品数	0	1	2
抽检前的概率	0.6	0.3	0.1
抽检后的概率	0.665	0.266	0.069

抽检的效用明显提高了。 □

换刚才讲过的另一种情况，假设任意批次中无次品，有一个次品和有两个次品的概率分别为 $0.1, 0.1, 0.8$。那么$p(A) = 0.1 \times 1 + 0.1 \times 0.8 + 0.8 \times 0.622 = 0.6776$，$p(B_0 : A) = 0.1 \times 1/0.6776 = 0.148$，$p(B_1 : A) = 0.1 \times 0.8/0.6776 = 0.118$，$p(B_2 : A) = 0.8 \times 0.622/0.6776 = 0.734$。对比抽检前后的概率分布：

批次中的次品数	0	1	2
抽检前的概率	0.1	0.1	0.8
抽检后的概率	0.148	0.118	0.734

投入的成本高了，抽检的效果也提高了。 □

当然，实际生产中，抽检的主要作用还是预警，一旦发现退回批次的频率超出预期，要马上反馈给生产部门，设法从源头上减少次品的产生。单靠抽检人员堵截次品这一个手段来提高产品质量，是非常不经济的。

6.2.4 临床检测问题

Bayes 公式的另一个应用是临床检测。

设某疾病的实际发生率为万分之四。某项检测，在患者中有 95% 呈阳性，非患者中有 1% 呈阳性。某人的该项检测结果呈阳性，问：他患该病的概率有多大？

记此人的该项检测结果呈阳性为事件 A，患该病为事件 B。那么，$p(B) = 0.0004$，$p(A : B) = 0.95$，$p(A : B^{\complement}) = 0.01$，求 $p(B : A)$。首先，求此人的该项检测结果呈阳性的概率 $p(A) = 0.0004 \times 0.95 + 0.9996 \times 0.01 = 0.00038 + 0.009996 = 0.010376$。进而，$p(B : A) = 0.00038/0.010376 = 0.0366$。所以，患者呈阳性的概率 $P(A : B)$，与呈阳性者患病的概率 $P(B : A)$，是完全不同的两个概念。

不难发现，该项检测诊断性能不高，在于非患者中也有 1% 呈阳性。虽然这个比例看似不大，但是考虑到非患者有 99.96% 之众，这些呈阳性的非患者还是“冲淡”了检测结果。如果改进这项检测，使得非患者中只有 0.01% 呈阳性，那么 $p(A) = 0.00038 + 0.00009996 = 0.00047996$，从而 $p(B : A) = 0.792$。该项检测的诊断性能就大大提高了。 □

6.2.5 “直观”是把双刃剑

前面多次讲到“直观上，如何如何”，似乎每次都证明“直观上的直觉判断”是错误的。只有一次，说直观上的“比重”观念与概率是一致的。其实，直观是把双刃剑，用得好可以启发思路，引导你走上正途；用不好则会误导你走上歧途。在网上看过一个八卦贴，说不同的数学家对于什么是“显然的”有着不同的尺度。有的数学家说某事是“显然的”，那么此事通常可以由他所带的研究生在两周内证出来。有的数学家说某事是“显然的”，那么此事通常是他在十五分钟之前就介绍过其证明的。还有的数学家说某事是“显然的”，那么此事通常是错的。虽然直观、直觉可以培养，但是社会大众，包括本文作者在内，在概率问题上的直觉显然是培养得很不够。

例如，有个 Monty Hall 问题：一场娱乐节目。主持人 Monty Hall 要求嘉宾从三个门中选一个，其中只有一个门后面有奖品。嘉宾做出选择之后，主持人会打开剩下的两个门之一，而且这个门后面没有奖品。然后主持人会问嘉宾要不要更换选择。

有研究表明，如果增加门的数量，使得主持人打开四个空门，再问嘉宾换不换，那么，选择换的人就开始明显增多。这说明，一种“正确的直觉”开始起作用了。但是，如原题所说，主持人只打开了一个门，那么，一种广泛的、错误的、顽固的直觉使得相当多的人认为没必要换。

另一个大众直觉的例子是彩票。彩票的本质是一种娱乐消费。看电影是花钱买别人造的梦，买彩票是花钱买自己造的梦。尽管象福彩这样的彩票，其可以理性预期的回报不会好于亏损一半，但是，社会上还是有那么多人要买，绝大多数都是基于非理性的预期。真正明明白白地买彩票来娱乐的，恐怕是凤毛麟角。

即便是对大师来说，直觉也是双刃剑。

据说，有一次玻尔参加一个学术会议，台上有个博士生在讲解一个重大突破，全场为之振奋。只有玻尔一脸茫然。旁边的人看到了，还主动向玻尔解释这项重大突破是怎么怎么一回事。然后，经过数小时的热烈讨论，大家终于发现，该博士生在证明中隐含地使用了一个命题，而这个命题并不总是正确的。玻尔也不是一眼就能看透，只是直觉告诉他，一定有什么地方不对劲。

另一方面，在普朗克的黑体辐射公式已经成为量子论的一块基石之后，普朗克本人却依然直觉地认为，量子论“一定有什么地方不对劲”。

所以，没直觉不行，太信直觉也不行。大概“凭直觉大胆假设，凭逻辑和实验小心求证”是比较折中、比较可取的吧。

6.2.6 Bertrand 悖论

其实，在概率论自身的发展过程中，也遇到过直观上无法解释的悖论。

Bertrand 悖论 (paradox)：在单位圆上取任意弦，问弦长大于 $\sqrt{3}$ 的概率几何。

如果所谓“取任意弦”是指先任意取定圆上一点，再任意取圆上另一点，两点连线成弦，那么第一步不影响概率，第二步有 1/3 的概率取到弦长大于 $\sqrt{3}$。

如果所谓“取任意弦”是指先任意取定一个方向，再任意取沿此方向的平行线与圆相交成弦，那么第一步不影响概率，第二步有 1/2 的概率取到弦长大于 $\sqrt{3}$。

如果所谓“取任意弦”是指先任意取定圆内一点，再做过该点的半径的过该点的垂线，与圆相交成弦，那么有 1/4 的概率取到弦长大于 $\sqrt{3}$。

直观上，“取任意弦”就是取任意弦，应该只有一种含义，因而弦长大于 $\sqrt{3}$ 的概率应该有唯一的结果。可是实际做起来，却有三种做法，而且计算出来的概率各不相同。哪一个才是对的呢？

回答是：虽然问题中是以任意弦为样本点，但是，上述三种做法使得任意弦分别与圆周上各点、直径上各点、圆内各点一一对应，因此，“样本点等可能”这一概念被分别转化成了圆周上各点、直径上各点、圆内各点有均等机会被选取，由此产生了三个不同的样本空间 Ω，进而得出了三个不同的概率。因此，是原题有含糊之处，而不是三个概率有问题。 □

6.2.7 Buffon 实验

Buffon 实验 (experiment)：向间距为 a 的一组平行线上扔长为 ℓ 的针，问针压到线的概率几何。

记针的中点到平行线的距离为 d，针的方向与平行线的交角为 ϕ。由实验方案，可以合理假设 d 在 $[0, a/2]$ 之间均匀分布，ϕ 在 $[0, \pi/2]$ 之间均匀分布。因此，可取 $\Omega = \{(d, \phi) : 0 \le d \le a/2, 0 \le \phi \le \pi/2\}$。取测度 $\mu(E)$ 为 d-ϕ 平面上点集 E 的面积，那么，$\mu(\Omega) = a\pi/4$。如果针压到线，那么 $d \le (\ell/2)\sin\phi$，记此事件为 A，那么 $\mu(A) = \dfrac{1}{2}\displaystyle\int_0^{\frac{\pi}{2}} \ell \sin\phi \, \mathrm{d}\phi = \dfrac{\ell}{2}$。所以，$p(A) = \dfrac{\mu(A)}{\mu(\Omega)} = \dfrac{2\ell}{\pi a}$。据说，Buffon 曾经用该实验来计算 π 的值。 □

6.2.8 两事件的相互独立与一组事件的独立

如果事件 B 的发生不影响事件 A 的概率，即 $p(A:B)=p(A)$，则称 A 和 B 是相互独立的。此时有 $p(A\cap B)=p(A)p(B)$。

如果一个有限的事件组 $\{A_1,\ldots,A_n\}$ 中，任何一个事件 A_i 与其他事件、以及其他事件任意的积，都是相互独立的，则称这组事件是独立事件。此时有 $p(A_1\cap\cdots\cap A_n)=p(A_1)\cdots p(A_n)$。概率的这种性质是一般测度所没有的。另外需注意：事件组中的事件两两独立 (pairwise independent) 并不意味着这组事件是独立事件。

6.2.9 二项分布与泊松分布

如果某试验 (trial) 只有事件 A 和 A^{C} 两种结果，记 $p(A)=p,\quad p(A^{\mathsf{C}})=q$，那么，该试验独立地进行 n 次，事件 A 恰发生 m 次的概率为 $p(m)=C_n^m p^m q^{n-m}$，称为二项分布 (binomial distribution)。

当 n 取很大的值而 p 很小时，记 $\lambda=np$，有

$$
\begin{aligned}
p(m) &= C_n^m p^m q^{n-m}\\
&= \frac{n(n-1)\cdots(n-m+1)}{m!}\left(\frac{\lambda}{n}\right)^m\left(1-\frac{\lambda}{n}\right)^{n-m}\\
&= \frac{\lambda^m}{m!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{m-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n(1-\frac{m}{n})}\\
&\approx \frac{\lambda^m}{m!}e^{-\lambda}
\end{aligned}
$$

这种分布称为泊松 (Poisson) 分布。它是对满足上述条件的二项分布的近似，其准确性好于正态分布。

6.3 网上经常争论的几个概率问题

6.3.1 证人证词问题

某市某型号的汽车，蓝色的占 15%，绿色的占 85%。在法庭上，某证人作证说：他看到该案件所涉及的一辆该型号汽车是蓝色的。但是经技术鉴定，该证人在当时光线较弱的条件下，能正确判断该型号车的颜色的概率只有 80%。问：该车确为蓝色的概率几何。

如果没有该证人的证词，那么该车可能是该市该型号车中的任何一辆，因此，有 15% 的概率是蓝色的。

现在，根据证人的这份证词，我们可以排除两种情况：一种是，这辆车是蓝色的，但是证人看错了，以为是绿色的；另一种是，这辆车是绿色的，证人看对了，认为是绿色的。如果是这两种情况，那么证人不会做出上述证词。

因此，在获得了这份证词的情况下，我们只需考虑两种情况：一种是他看对了，这辆车确实是蓝色的；另一种是他看错了，这辆车其实是绿色的。

假设该市该型号车有十万辆，显见其中蓝车一万五，绿车八万五。证人看对了，把蓝车看成蓝车的情况有一万五中的 80%，即一万两千种；证人看错了，把绿车看成蓝车的情况有八万五中的 20%，即一万七千种。一共是两万九千种。其中该车确实是蓝车的情况有一万两千种。

所以，在获得证词后，所应考虑的情况中，该车为蓝色的概率是 $12/29 \approx 41.4\%$。与没有证词时相比，有大幅提高，但仍然没有超过 50%。

数学上讲，这是个贝叶斯公式的问题。记被此人指认为蓝色是事件 A，该车确为蓝色为事件 B。$p(B) = 0.15$，$p(A : B) = 0.8$, $p(A : \neg B) = 0.2$。所以

$$
\begin{aligned}
p(B : A) &= p(B)p(A : B)/(p(B)p(A : B) + p(\neg B)p(A : \neg B)) \\
&= 0.15 * 0.8/(0.15 * 0.8 + 0.85 * 0.2) = 0.6/1.45 = 12/29.
\end{aligned}
$$

6.3.2 三门抽奖问题

就是 Monty Hall 问题。如果主持人随意打开挑剩的两个门之一，那么在所有节目中有大约三分之一是主持人发现奖品。这种情况下，嘉宾无需更换选择。然而题目所述的情况是，主持人打开的门总是空的。

如果有六个门，嘉宾挑一个，中奖概率为六分之一。然后主持人打开四个空门，剩下一个，问你换不换。这时选择换的人就多了，因为这种情况下，比较容易看出来，最后这个门的中奖概率远高于嘉宾挑的那个门。既然嘉宾挑的那个门的中奖概率是六分之一，那么最后这个门的中奖概率一定是六分之五。

同理，题目所述情况下，最后那个门的中奖概率是三分之二。嘉宾挑的那个门的中奖概率是三分之一。可能是因为两者的差距不够悬殊，所以很多人凭“直觉”分辨不出，总以为是一半一半。

6.3.3 酒鬼问题

已知某酒鬼在晚上有 90% 的概率外出，在三家酒吧中随意挑一家喝酒。某天晚上，警察找了两家酒吧没有找到该酒鬼，问该酒鬼在第三家酒吧的概率几何。

假设连续一千天，天天发生这样的事，那么这一千天里，该酒鬼大约一百天在家，三百天在酒吧一，三百天在酒吧二，三百天在酒吧三。这样，该酒鬼不在警察找的那两家酒吧的日子，共有四百天，其中在家的日子有一百天，在第三家酒吧的日子有三百天，因此，在给定了上述条件的情况下，所求条件概率为四分之三。

有网友提出这样的疑问：假设该酒鬼在家、在酒吧一、在酒吧二，三种情况等概率，现在警察在酒吧一没有找到该酒鬼，求该酒鬼在酒吧二的概率。这个题与三门抽奖问题有何区别？

首先，警察未在酒吧一找到该酒鬼，剩下在家和在酒吧二两种情况等可能，因此，所求条件概率为二分之一。

其次，我们设想有两个警察，警察甲去查酒鬼家，警察乙去查两家酒吧。现在警察乙查了一家酒吧，没有找到该酒鬼。问警察乙在另一家酒吧找到该酒鬼的概率几何。这看着很像是三门抽奖问题，然而却不是。

假设三年内，连续九百天，这样的事天天发生，那么，平均来看，该酒鬼有三百天在家，三百天在酒吧一，三百天在酒吧二。对于前一题，该酒鬼不在酒吧一的日子共有六百天，其中有三百天他在酒吧二，所以，所求条件概率为二分之一。对于后一题，该酒鬼被警察甲找到的日子有三百天，被警察乙找一次就找到的日子有三百天，被警察乙找两次才找到的日子有三百天。也就是说，警察乙找一次没找到的日子一共有六百天。其中找第二次找到了的日子有三百天，找第二次也没找到的日子有三百天，所以，在“警察乙找一次没找到”的条件下，所求条件概率仍旧为二分之一。

现在，假设警察乙在两个酒吧都有线人，如果该酒鬼不在某酒吧，那么该酒吧的线人就会通知警察乙去查另一家酒吧。问警察乙在另一家酒吧找到该酒鬼的概率几何。

考虑上述九百天。除了该酒鬼在家的三百天，其余六百天警察乙都能在另一家酒吧找到该酒鬼，所以，所求概率为三分之二。这种情形才和三门抽奖问题一样。主持人打开一个空门，就相当于线人通知警察乙说该酒鬼不在这里。

如果两个线人把情况也通知了警察甲，那么警察甲找到该酒鬼的“概率”会不会变化呢？

可以说不会，也可以说会，取决于题目所说的“概率”的确切含义。

如果题中“概率”是指总的概率，那么，这两个线人所提供的信息对警察甲的行动没有任何实质的影响，警察甲总是要到而且只到酒鬼家，不能根据该情报改变计划，因而九百天里，总是只有三百天能找到该酒鬼，所以，从总体上讲，警察甲找到该酒鬼的概率是三分之一。

如果题中“概率”是指获得信息之后的条件概率，那么，虽然情报对警察甲的行动没有影响，但是，情报缩小了样本空间。已知该酒鬼不在某个酒吧的条件下，可考虑的天数只有六百天，其中有三百天在家，因此，在获得情报之后，警察甲找到该酒鬼的条件概率是二分之一。

6.3.4 小孩接电话问题

已知某家庭有两个孩子。星期二晚上给这家打电话，是孩子接的，并且得知这个孩子是女孩。问另一个孩子也是女孩的概率几何。

简单地说，孩子是男孩还是女孩互不影响，因此，另一个孩子是女孩的概率是二分之一。

详细点讲，该家庭两个孩子，甲男乙男，甲男乙女，甲女乙男，甲女乙女，四种情况等可能。得知其中一个是女孩，不失一般性，设得知孩子乙是女孩，那么还有甲男乙女，甲女乙女两种情况等可能。因此，孩子甲是女孩的概率为二分之一。

6.3.5 首先命中问题

甲、乙、丙三人，每次射击的命中率分别为 0.4，0.3，0.3。每天按照先甲，后乙，再丙，然后再循环的顺序练习射击。问甲在每天训练中首先命中的概率。

在所有的训练日中，甲第一枪即命中的概率为 0.4。甲第一枪不中，且随后乙、丙皆不中，甲第二枪命中的概率为 $0.6 \times 0.7 \times 0.7 \times 0.4$。这一枪也不中，且随后乙、丙皆不中，甲第三枪命中的概率为 $(0.6 \times 0.7 \times 0.7)^2 \times 0.4$。依此类推，构成等比数列。易见甲首先命中的概率为该等比数列之和。

6.4 随机变量

6.4.1 随机变量

如果 $(\Omega, \mathcal{E}, p)$ 为概率空间，则称可测函数 $X\colon \Omega \to \mathbf{R}$ 为该概率空间上的一个随机变量 (random variable)。

例如，我们把标准六面骰子的六个面上的图形记为 a, b, c, d, e, f。用 $\omega = a$ 表示掷骰子结果为 a 面向上，依此类推，那么 $\Omega = \{a, b, c, d, e, f\}$。设可测函数 $X\colon \Omega \to \mathbf{R}$ 为 $X(a) = 100, \quad X(b) = 200$，依此类推，直到 $X(f) = 600$。那么，X 就是一个随机变量。我们可以说：X 取值为 200 的概率等于 $p(\{b\}) = 1/6$，X 取值大于 400 的概率等于 $p(\{e, f\}) = 1/3$。

前面讲过，给定从可测空间 $(X, \mathcal{A})$ 到可测空间 $(\overline{\mathbf{R}}, \mathcal{B}(\mathbf{R}))$ 的函数 $f\colon X \to \overline{\mathbf{R}}$，它是可测的，当且仅当，对任意实数 c，$[-\infty, c)$ 的原象都是可测集。显见，当可测空间 $(X, \mathcal{A})$ 是样本空间 $(\Omega, \mathcal{E})$ 时，函数 $X\colon \Omega \to \mathbf{R}$ 是随机变量，当且仅当，对任意实数 c，$[-\infty, c)$ 的原象都是可测集。

具体到上例，$X^{-1}([-\infty, 100)) = \varnothing$，$X^{-1}([-\infty, 101)) = \{a\}$，$X^{-1}([-\infty, 300)) = \{a, b\}$，依此类推，直到$X^{-1}([-\infty, 601)) = \Omega$。很自然地，我们可以使用拖回测度来定义随机变量的取某可测集时的概率。例如，$p([-\infty, 100)) = p(\varnothing) = 0$，$p([-\infty, 101)) = p(\{a\}) = 1/6$，等等。

如果随机变量 X 作为函数，其值域 $X(\Omega)$ 是有限集或者可数集，则称 X 是离散的 (discrete)；如果其值域是一个区间，则称 X 是连续的 (continuous)。

理论上可以存在既非离散，又非连续的随机变量，但实践上这种情形意义不大。如果是在某阈值之下为离散的，阈值之上为连续的，那通常分解成两个子问题分别处理。比如，微观粒子能量低时，被势阱捕获处于束缚态，其能量就是离散型随机变量；能量高时，突破势阱处于游离态，其能量就是连续型随机变量。

6.4.2 累积分布函数、概率分布函数与分布密度函数

给定随机变量 $X\colon \Omega \to \mathbf{R}$，函数 $F_X(c) = p([-\infty, c)) = p(X^{-1}([-\infty, c)))$ 称为 X 的累积分布函数 (cumulative distribution function)，常简称 cdf。

给定随机变量 $X\colon \Omega \to \mathbf{R}$，和数轴上的点集 $A \in \mathcal{B}(\mathbf{R})$，函数 $P_X(A) = p(X^{-1}(A))$ 称为 X 的概率分布函数 (probability distribution function)，常简称 pdf。注意，pdf 的自变量是点集，不是实数，所以没有简单的图示法。另外，在不致混淆的情况下，我们也把 $P_X(A)$ 记做 $p(A)$。

给定连续型随机变量 $X\colon \Omega \to \mathbf{R}$，函数 $\phi_X(x) = \lim\limits_{\Delta x \to 0} \dfrac{P_X((x, x + \Delta x))}{\Delta x}$ 称为 X 的分布密度函数 (distribution density function)，常简称 ddf。它的图形 $y = \phi_X(x)$ 通常称为 X 的分布曲线。

6.4.3 随机变量的期望值

给定概率空间 $(\Omega, \mathcal{E}, p)$ 上的随机变量 X，如果 X 是可积的，则称 $E(X) = \displaystyle\int_\Omega X \,\mathrm{d}p$ 为 X 的期望值 (expected value)，也称数学期望 (mathematical expectation)，或平均值 (mean)。注意，求一个随机变量的期望值就是求该函数的特定类型的积分，反过来说，凡是求某函数的这种类型的积分，都可以说是求相应的随机变量的期望值。可见，期望值在概率论中具有基础的重要性。许多概念都是以随机变量的某个函数的期望值来表述的。

前面讲过，如果一个可测函数的值域是有限集，则称它是简单的。简单函数的积分为 $\displaystyle\int_X \phi \,\mathrm{d}\mu = \sum_{i=1}^{n} a_i \mu(S_i)$。相应地，离散型随机变量的期望值可以表示为 $E(X) = \sum x_i p(S_i)$。比如，上述六面骰子的例子，$E(X) = 100p(\{a\}) + 200p(\{b\}) + \cdots + 600p(\{f\}) = 350$。

6.4.4 随机变量序列的几乎肯定收敛和依概率收敛

前面讲的可测函数序列的收敛，在概率论中就成了随机变量序列的收敛。

给定概率空间 $(\Omega, \mathcal{E}, p)$ 和其上的随机变量序列 $(X_n : n \in \mathbf{N})$ 和随机变量 X，如果 $p\left(\{\omega : \lim\limits_{n \to \infty} X_n(\omega) = X(\omega)\}\right) = p(\Omega) = 1$，则称 (X_n) 以概率 1 收敛于 X，也称几乎肯定 (almost surely，常简记为 a.s.) 收敛于 X。显然，如果以概率为测度，那么几乎肯定收敛就是在 Ω 上几乎处处收敛。换个名字仅仅是为

了在措辞上更加自然、贴切；另外，总是“在 Ω 上”，不再选择集合 $A \subseteq \Omega$。注意，“几乎肯定收敛”并非“还不完全肯定是否收敛”。数学本身也是一门语言，与日常用语同词不同义的情况并不鲜见。

如果对于任意正实数 ε，均有 $\lim\limits_{n\to\infty} p(|X_n - X| < \varepsilon) = p(\Omega) = 1$，则称 (X_n) 依概率收敛于 X。这其实就是在 Ω 上依测度收敛。

6.4.5 平均值的稳定性

继续用骰子做例子，记相应的独立试验序列为 $(X_i : i \in \mathbf{N})$，那么随机变量序列 $\left(\dfrac{X_1 + \cdots + X_n}{n} : n \in \mathbf{N}\right)$ 在 $n \to \infty$ 时依概率收敛于 $p(X = 350) = 1$。也就是说，该序列的极限是：这个变量必定取 350，不再有任何的随机性。

连续型随机变量常采用分布密度和分布曲线来刻画，离散型随机变量则常采用分布列来刻画。本例中，X_1, X_2, X_3 等的分布列都是

100	200	300	400	500	600
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$\dfrac{X_1 + X_2}{2}$ 的分布列为

100	150	200	250	300	350	400	450	500	550	600
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$\dfrac{X_1 + X_2 + X_3}{3}$ 取 100 的概率为 $6^{-3} = 1/216$，取 133.33 的概率为 $3/216$。另一端的情况类似，取 600 的概率为 $1/216$，等等。我们已经可以看出规律了，随着 n 的增加，随机变量取两端的概率迅速衰减，最终只有取 350 这一种情况可以幸免。

6.4.6 随机变量的离差、绝对差、方差与标准差

由此可见，随机变量的取值在期望值附近集中的程度是个很重要的概念。$X - E(X)$ 称为离差 (deviation)。离差的绝对值的期望值 $E(|X - E(X)|)$ 称为绝对差 (absolute deviation)。离差的平方的期望值 $E((X - E(X))^2)$ 称为方差 (variance)，记为 $D(X)$，方差的正的平方根 $+\sqrt{D(X)}$ 称为标准差 (standard deviation), 或者均方差 (mean square deviation)，记为 $\sigma(X)$。显见，$\sigma^2(X) = D(X)$。注意，以上各种 deviation 都具有和随机变量 X 相同的量纲，只有方差 variance 的量纲是 X 的量纲的平方。

$D(X) = E(X^2) - (E(X))^2$。因为

$$D(X) = E((X - E(X))^2)$$

$$
\begin{aligned}
&= E(X^2 - 2E(X)X + (E(X))^2) \\
&= E(X^2) - 2E(X)E(X) + (E(X))^2 \\
&= E(X^2) - (E(X))^2. \quad \square
\end{aligned}
$$

6.4.7 两个随机变量的相关矩

给定同一概率空间上的两个随机变量 X 和 Y，它们离差的乘积的期望值 $E((X-E(X))(Y-E(Y)))$ 称为相关矩 (covariance moment)，记为 $\operatorname{cov}(X,Y)$。

如果随机变量 X 和 Y 是相互独立的，那么 $\operatorname{cov}(X,Y)=0$。因为，如果事件 A 和 B 相互独立，那么 $p(A\cap B)=p(A)p(B)$。所以

$$
\begin{aligned}
\operatorname{cov}(X,Y) &= \int_{\Omega^2} (X-E(X))(Y-E(Y))\,\mathrm{d}p_{\Omega^2} \\
&= \int_{\Omega} \left(\int_{\Omega} (X-E(X))(Y-E(Y))\,\mathrm{d}p \right) \mathrm{d}p \\
&= \int_{\Omega} (X-E(X))\,\mathrm{d}p \int_{\Omega} (Y-E(Y))\,\mathrm{d}p \\
&= E(X-E(X))E(Y-E(Y)) \\
&= 0. \quad \square
\end{aligned}
$$

如果随机变量 X 和 Y 是相互独立的，那么 $D(X+Y)=D(X)+D(Y)$。因为，既然 X 和 Y 相互独立，那么 $\operatorname{cov}(X,Y)=0$，所以

$$
\begin{aligned}
D(X+Y) &= E(((X+Y)-E(X+Y))^2) \\
&= E(((X-E(X))+(Y-E(Y)))^2) \\
&= E((X-E(X))^2 + 2(X-E(X))(Y-E(Y)) + (Y-E(Y))^2) \\
&= D(X) + 2\operatorname{cov}(X,Y) + D(Y) \\
&= D(X) + D(Y). \quad \square
\end{aligned}
$$

给定随机变量 X 和实数 k，有 $D(kX)=k^2D(X)$。因为 $D(kX)=E((kX-E(kX))^2)=E(k^2(X-E(X))^2)=k^2D(X)$。

具体到刚才骰子的例子。$D(X_1+\cdots+X_n)=nD(X_1)$，当 $n\to\infty$ 时，

$$
D(\frac{X_1+\cdots+X_n}{n}) = \frac{D(X_1+\cdots+X_n)}{n^2} = \frac{1}{n}D(X_1) \to 0,
$$

该随机变量将完全地等于其期望值，不再有机会取其他的值。

6.5 大数定律

6.5.1 切贝谢夫定理

实际上，我们不必强求 $D(X_i)$ 都相等，只要它们都有界，即都不大于某个

常数 K 即可。当 $n \to \infty$ 时，上述表达式的方差不大于 K/n，因此趋于零。另外，我们也不必强求 $E(X_i)$ 都相等，只要把上述表达式的期望值替换成各变量期望值的平均值即可。这样，我们就得到下述定理：

切贝谢夫定理 (Chebyshev's Theorem)：给定独立随机变量序列 $(X_i : i \in \mathbf{N})$，如果这些随机变量的方差是一致有界的，即存在同一个常数 K，使得所有 X_i 均有 $D(X_i) \le K$，则，记逐步平均序列 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 为 $\overline{X}_n$，有 $E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$，并且，对于任何正数 ε，恒有 $\lim_{n\to\infty} p(|\overline{X}_n - E(\overline{X}_n)| < \varepsilon) = 1$。

要证明这个定理，我们先看一下切贝谢夫不等式 (Chebyshev's Inequality)：给定随机变量 X，对任意正数 ε，有 $p\left(\frac{|X - E(X)|}{\varepsilon} \ge 1\right) \le \left(\frac{\sigma(X)}{\varepsilon}\right)^2$。因为，

$$\begin{aligned}
p\left(\frac{|X - E(X)|}{\varepsilon} \ge 1\right) &= \int_{\Omega} I_{|X-E(X)|\ge\varepsilon}\,\mathrm{d}p \\
&\le \int_{\Omega} \left(\frac{|X - E(X)|}{\varepsilon}\right)^2 I_{|X-E(X)|\ge\varepsilon}\,\mathrm{d}p \\
&\le \frac{1}{\varepsilon^2}\int_{\Omega} (X - E(X))^2\,\mathrm{d}p \\
&= \frac{D(X)}{\varepsilon^2} = \left(\frac{\sigma(X)}{\varepsilon}\right)^2. \quad \square
\end{aligned}$$

然后来证切贝谢夫定理。对逐步平均序列的通项 $\overline{X}_n$ 应用切贝谢夫不等式，因为 (X_i) 是独立随机变量序列，所以 $D(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i)$，因此

$$p(|\overline{X}_n - E(\overline{X}_n)| < \varepsilon) \ge 1 - \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n} D(X_i) \ge 1 - \frac{K}{\varepsilon^2 n}.$$

当 $n \to \infty$ 时，此式大于等于 1。但是概率不能大于 1，因此上式等于 1。 $\square$

6.5.2 伯努利定理

阐述同类大量随机事件的平均结果趋于稳定，几乎必然取特定值的定律统称大数定律 (Law of Large Numbers)。切贝谢夫定理是其中较重要的一个。另一个较为典型的大数定律是伯努利定理。

伯努利定理 (Bernoulli's Theorem)：在独立试验序列中，事件 A 的频率随着试验次数的不断增加，依概率收敛于 A 的概率 p。因为，可以定义随机变量 X 为：$X(A) = 1, \quad X(A^{\complement}) = 0$，那么，$A$ 的频率就是 $\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{m}{n}$，频率的期望值就是 $\frac{1}{n}\sum_{i=1}^{n} E(X_i) = p$。每个 X_i 的方差都是 $D(X_i) = (1-p)^2 p + (0-p)^2(1-p) = p(1-p) \le 1/4$，是有界的。由切贝谢夫定理立得。 $\square$

6.5.3 马尔科夫不等式

回头看一下切贝谢夫不等式的证明，不难发现改成绝对差也可以做成不等式。即

$$\begin{aligned} p\left(\frac{|X-E(X)|}{\varepsilon} \geq 1\right) &= \int_\Omega I_{|X-E(X)|\geq\varepsilon}\,\mathrm{d}p \\ &\leq \int_\Omega \frac{|X-E(X)|}{\varepsilon} I_{|X-E(X)|\geq\varepsilon}\,\mathrm{d}p \\ &\leq \frac{1}{\varepsilon}\int_\Omega |X-E(X)|\,\mathrm{d}p \\ &= \frac{E(|X-E(X)|)}{\varepsilon}. \quad \square \end{aligned}$$

这称为马尔科夫不等式。它有如下更为一般的表述：

马尔科夫不等式 (Markov's Inequality)：给定测度空间 $(X, \mathcal{M}, \mu)$ 和其上的非负可测函数 f，对任意正数 ε，有 $\mu([\varepsilon, \infty]) \leq \frac{1}{\varepsilon}\int_X f\,\mathrm{d}\mu$。

6.6 中心极限定理

6.6.1 半均分布的稳定性

回到刚才骰子的例子。当 $n \to \infty$ 时，

$$D(\frac{X_1+\cdots+X_n}{\sqrt{n}}) = \frac{D(X_1+\cdots+X_n)}{(\sqrt{n})^2} = D(X_1),$$

这个随机变量的方差不趋于零。但是

$$E(\frac{X_1+\cdots+X_n}{\sqrt{n}}) = \frac{E(X_1+\cdots+X_n)}{\sqrt{n}} = \sqrt{n}E(X_1) \to \infty.$$

似乎是没有意义啦。但是，且慢，如果 $E(X_1) = 0$ 呢？这很容易做到，因为 $E(X_1 - E(X_1)) = 0$。

给定一组随机变量 $X_1, \ldots, X_n$，称 $\frac{X_1+\cdots+X_n}{n}$ 为这组随机变量的平均结果，常记为 $\overline{X}$。可称 $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ 为这组随机变量的半均结果，可记为 $\ddot{X}$。显见 $\overline{X} = \frac{1}{\sqrt{n}}\ddot{X}$。

令 $Y_i = X_i - E(X_i)$，则 $D(Y_i) = D(X_i)$，并且 $E(Y_i) = 0$。进而

$$D(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}) = \frac{D(Y_1+\cdots+Y_n)}{(\sqrt{n})^2} = D(Y_1) = D(X_1),$$

而且

$$E(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}) = \frac{E(Y_1+\cdots+Y_n)}{\sqrt{n}} = \sqrt{n}E(Y_1) = 0.$$

这说明该随机变量 $\frac{Y_1+\cdots+Y_n}{\sqrt{n}}$ 有一个以零为“中心”的，固定分散程度的分布。

具体来说，Y_1, Y_2, Y_3 等的分布列都是

$$\begin{array}{cccccc} -250 & -150 & -50 & 50 & 150 & 250 \\ \hline \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array}$$

$\frac{Y_1+Y_2}{\sqrt{2}}$ 的分布列为

$$\begin{array}{ccccccccccc} -250\sqrt{2} & -200\sqrt{2} & -150\sqrt{2} & -100\sqrt{2} & -50\sqrt{2} & 0 & 50\sqrt{2} & 100\sqrt{2} & 150\sqrt{2} & 200\sqrt{2} & 250\sqrt{2} \\ \hline \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{array}$$

$\frac{Y_1+Y_2+Y_3}{\sqrt{3}}$ 取 $-250\sqrt{3}$ 的概率为 $6^{-3}=1/216$，等等。易见，随着 n 的增加，随机变量取两端的范围可以无限延伸，但是其概率迅速衰减，而方差却是固定的。它最终将趋于一个特定的连续分布。

6.6.2 正态分布

如果令 $Y_i=\frac{X_i-E(X_i)}{\sigma(X_i)}$，则 $D(Y_i)=\frac{D(X_i)}{(\sigma(X_i))^2}=1$，并且 $E(Y_i)=0$。称这样的随机变量 Y_i 是标准化的。显见，对任意的整数 $n\geq 1$，随机变量 $\frac{Y_1+\cdots+Y_n}{\sqrt{n}}$ 也是标准化的，并且可证，当 $n\to\infty$ 时，它趋于 $\phi(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$。该极限分布称为标准正态分布 (standard normal distribution)，记为 $N(0,1)$。

一般地，如果随机变量 X 的分布密度为 $\phi(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$，则称 X 服从正态分布 (normal distribution)。可证 $E(X)=a,\quad D(X)=\sigma^2$。因此，该正态分布常简记为 $N(a,\sigma^2)$。（感谢 dragonheart6 和 freephoenix。）相应的分布曲线常称为钟形曲线。该曲线关于期望值 $x=a$ 对称，而且有 $p(|x-a|<\sigma)=0.683,\quad p(|x-a|<2\sigma)=0.954,\quad p(|x-a|<3\sigma)=0.997$。

6.6.3 林德博尔格条件

如果一个随机变量是大量独立随机变量的和，但是任何个别变量对总和都只起微小作用，那么该总和变量的分布趋于正态分布。与这一现象有关的定理统称中心极限定理 (Central Limit Theorem)。历史上，人们就各种不同的条件，证明了一系列相应的中心极限定理。其中较为重要的一种条件是：

林德博尔格条件 (Lindeberg's condition)：给定独立随机变量序列 $(X_i : i\in$

$\mathbf{N}$)，记 $\sum_{i=1}^{n} D(X_i)$ 为 ${s_n}^2$。则把对任意给定的正数 ε，均有

$$\lim_{n\to\infty} \frac{1}{{s_n}^2} \sum_{i=1}^{n} \int_{|X_i - E(X_i)| > \varepsilon s_n} (X_i - E(X_i))^2 \,\mathrm{d}p = 0,$$

称为林德博尔格条件。

这个条件有更为通俗的解释：

$$\begin{aligned}
0 &= \lim_{n\to\infty} \sum_{i=1}^{n} \int_{\frac{|X_i - E(X_i)|}{s_n} > \varepsilon} \frac{(X_i - E(X_i))^2}{{s_n}^2} \,\mathrm{d}p \\
&\geq \lim_{n\to\infty} \sum_{i=1}^{n} \int_{\frac{|X_i - E(X_i)|}{s_n} > \varepsilon} \varepsilon^2 \,\mathrm{d}p \\
&= \varepsilon^2 \lim_{n\to\infty} \sum_{i=1}^{n} p\left(\frac{|X_i - E(X_i)|}{s_n} > \varepsilon\right) \\
&\geq \varepsilon^2 \lim_{n\to\infty} p\left(\sup_{1\leq i\leq n} \frac{|X_i - E(X_i)|}{s_n} > \varepsilon\right)
\end{aligned}$$

但是，概率不能小于零，而 ε 又是任意的，因此，林德博尔格条件等价于

$$\lim_{n\to\infty} p\left(\sup_{1\leq i\leq n} \frac{|X_i - E(X_i)|}{s_n} > \varepsilon\right) = 0.$$

如果带参数 $\theta \in \Theta$ 的随机变量序列 $(g_n(\theta) : n \in \mathbf{N})$，与依赖于参数的常数 $c(\theta)$，满足对于任意正数 ε 均有 $\lim_{n\to\infty} p\left(\sup_{\theta\in\Theta} |g_n(\theta) - c(\theta)| > \varepsilon\right) = 0$，则称 $g_n(\theta)$ 在参数空间 Θ 上一致地依概率收敛于 $c(\theta)$。（感谢 dragonheart6。）

取 $c(i) = 0$，$g_n(i) = \dfrac{|X_i - E(X_i)|}{s_n}$，那么，林德博尔格条件就是：当 $n \to \infty$ 时，$g_n(i)$ 在参数空间 $\mathbf{N}$ 上一致地依概率收敛于零。这就是“任何个别变量对总和都只起微小作用”的数学表达。

如果满足林德博尔格条件，那么

$$\lim_{n\to\infty} \phi\left(\frac{\sum_{i=1}^{n}(X_i - E(X_i))}{s_n}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

这是中心极限定理的较为一般的表述。

特别地，如果 $(X_i : i \in \mathbf{N})$ 是独立同分布 (independent identically distributed，常简称 iid) 随机变量序列，设 $E(X_i) = a$，$D(X_i) = \sigma^2$，则 ${s_n}^2 = \sum_{i=1}^{n} D(X_i) = n\sigma^2$。上述收敛条件就是当 $n \to \infty$ 时，一致地有 $\dfrac{1}{\sqrt{n}} \dfrac{X_i - a}{\sigma}$ 趋于零。显见林德博尔格条件满足。进而标准化的变量序列 $\left(Y_i = \dfrac{X_i - a}{\sigma} : i \in \mathbf{N}\right)$ 的逐步半均结果 $\dfrac{Y_1 + \cdots + Y_n}{\sqrt{n}}$ 趋于标准正态分布 $N(0,1)$。也就是说，$\ddot{X}_n$ 趋

于 $N(\sqrt{n}\,a,\sigma^2)$。注意到，$\overline{X}_n = \dfrac{1}{\sqrt{n}}\ddot{X}_n$，显见 $\overline{X}_n$ 趋于 $N(a,\dfrac{\sigma^2}{n})$。当 $n\to\infty$ 时，$\overline{X}_n$ 趋于 $N(a,0)$，正符合大数定律所述。

6.6.4 德莫威尔-拉普拉斯定理

对前面伯努利定理所述情形，即大量二项分布的和，也有一个相应的中心极限定理。此时，$\sum_{i=1}^{n}(X_i - E(X_i)) = m - np$，$\sigma = \sqrt{p(1-p)}$。所以，$\dfrac{m-np}{\sqrt{np(1-p)}}$ 服从标准正态分布 $N(0,1)$。这称为德莫威尔 (de Moivre) -拉普拉斯 (Laplace) 定理。

6.7 二维随机变量

6.7.1 两个随机变量的和

设两个独立随机变量 X 和 Y 服从相同类型的分布，如果它们的和 $X+Y$ 也服从该类型的分布，则称该类型的分布是可加的，也称可再生的。

正态分布是可加的。这乍一看好象不可能。因为假设 X 服从 $N(1,0.01)$，Y 服从 $N(100,0.01)$，那么这两个“钟形”都很瘦，而且相距甚远，加在一起应该是两个离得很远的钟形啊，怎么可能融合成一个“钟形”呢？事情完全不是这么回事儿。我们先澄清一下求和的含义。

设独立随机变量 X 和 Y 的分布列如下：

	100	200	300
$p(X)$	0.7	0.2	0.1
$p(Y)$	0.1	0.2	0.7
$p(X)+p(Y)$	0.8	0.4	0.8

前面所谓的“两个钟形”就是表上的 $p(X)+p(Y)$，这一行的和等于 2，根本就不符合 $p(\Omega)=1$ 的要求，不是任何随机变量的分布列。

那么 $X+Y$ 的分布列该怎么求呢？设 $Z = X+Y$，那么 $Z=200 \iff (X=100)\cap(Y=100)$，所以 $p(Z=200)=p(X=100)p(Y=100)=0.07$。$Z=300 \iff ((X=100)\cap(Y=200))\cup((X=200)\cap(Y=100))$，所以 $p(Z=300)=p(X=100)p(Y=200)+p(X=200)p(Y=100)=0.14+0.02=0.16$。依此类推，可得 Z 的分布列：

	200	300	400	500	600
$p(X+Y)$	0.07	0.16	0.54	0.16	0.07

原来，由于 X 与 Y 是相互独立的，所以它们实际上构成了两个维度，考虑映射 $f\colon \Omega^2 \to \mathbf{R}$：

	$X=100$	$X=200$	$X=300$
$Y=100$	0.07	0.02	0.01
$Y=200$	0.14	0.04	0.02
$Y=300$	0.49	0.14	0.07

称 X 和 Y 组成了二维随机变量 (X,Y)。该表格的含义是 $p((X,Y)=(a,b))=p((X=a)\cap(Y=b))$，称为 X 和 Y 的联合分布 (joint distribution)。因为我们已设 X 和 Y 相互独立，所以在该表中，$p((X=a)\cap(Y=b))=p(X=a)p(Y=b)$。而 $Z=X+Y$ 的五个值恰对应该表的从左下到右上这个方向的五条斜线，取某个值的概率，等于对应的斜线上的所有概率的和。

从上表还可以看出，X 在 100 处取大概率，Y 在 300 处取大概率，这意味着联合分布 (X,Y) 在 $(100,300)$ 这“一个”地方取较大概率。这与 $p(X)+p(Y)$ 有两处较大值完全不是一回事儿。

6.7.2 二维随机变量的联合分布

一般地，给定可测空间 $(\Omega_1,\mathcal{E}_1)$ 和 $(\Omega_2,\mathcal{E}_2)$，可做可测空间 $(\Omega_1\times\Omega_2,\mathcal{E}_1\times\mathcal{E}_2)$。如果此空间上的测度 p 满足 $p(\Omega_1\times\Omega_2)=1$，则称 p 是二元 (bivariate) 概率测度。称可测函数 $(X,Y)\colon\Omega_1\times\Omega_2\to\mathbf{R}$ 为二维随机变量，设 $A\in\mathcal{E}_1,\quad B\in\mathcal{E}_2$，那么 (A,B) 表示 A 并且 B，因此又常记做 $A\cap B$。$P_{(X,Y)}((A,B))=p((X,Y)^{-1}(A\cap B))$ 称为 X 和 Y 的联合概率分布函数。

给定连续型二维随机变量 $(X,Y)\colon\Omega_1\times\Omega_2\to\mathbf{R}$，函数

$$\phi_{(X,Y)}(x,y)=\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\frac{P_{(X,Y)}(((x,x+\Delta x),(y,y+\Delta y)))}{\Delta x\Delta y}$$

称为联合分布密度函数。它的图形 $z=\phi(x,y)$ 称为该联合分布的分布曲面。

6.7.3 独立正态分布是可加的

如果独立随机变量 X 服从 $N(a_x,{\sigma_x}^2)$，Y 服从 $N(a_y,{\sigma_y}^2)$，那么 (X,Y) 的联合分布密度就是 $\phi(x,y)=\dfrac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\dfrac{1}{2}(\dfrac{(x-a_x)^2}{{\sigma_x}^2}+\dfrac{(y-a_y)^2}{{\sigma_y}^2})\right)$，称为二维正态分布。

对于 $z=x+y$ 的任意给定值 $z=z_0$，有 $x=t,\quad y=z_0-t,\quad t\in\mathbf{R}$ 均能满足 $x+y=z_0$，所以随机变量 $Z=X+Y$ 的分布密度为 $\phi(z)=\displaystyle\int_{\mathbf{R}}\phi(t,z-t)\,\mathrm{d}t$。如果 X 和 Y 都如上段所述，则可证 Z 服从正态分布 $N(a_x+a_y,{\sigma_x}^2+{\sigma_y}^2)$。

我们在前面已经讲过，如果 X 和 Y 相互独立，那么 $E(X+Y)=E(X)+E(Y)$，$D(X+Y)=D(X)+D(Y)$，因此，上段结论的要点在于 Z 服从正态分布，而不是它的期望值和方差如何如何。

6.7.4 二维随机变量的边缘分布与条件分布

给定二维随机变量 (X,Y)，并设 X 是概率空间 $(\Omega_x,\mathcal{E}_x,p_x)$ 上的随机变量，Y 是概率空间 $(\Omega_y,\mathcal{E}_y,p_y)$ 上的随机变量，(X,Y) 分布密度为 $\phi(x,y)$，那么称 $\phi_X(x)=\int_{\Omega_y}\phi(x,y)\,\mathrm{d}p_y$ 为 X 的边缘分布密度 (marginal distribution density)，常形象地称这种计算为“积掉 (integrating out)” y；称 $\phi_Y(y)=\int_{\Omega_x}\phi(x,y)\,\mathrm{d}p_x$ 为 Y 的边缘分布密度。

条件同上段，设 $B\in\mathcal{E}_y$，那么称 $P((X=x):B)$ 为 X 在条件 B 下的条件分布 (conditional distribution)，记为 $p(x:B)$。尤其是，当 B 为 $Y=y$ 时，常把 $p(x:(Y=y))$ 记为 $p(x:y)$。显然，$p(x:B)=\dfrac{p((X=x)\cap B)}{p(B)}$。尤其是，$p(x:y)=\dfrac{p((X=x)\cap(Y=y))}{p_Y(y)}=\dfrac{p(x,y)}{p_Y(y)}$。类似地，$\phi(x:y)=\dfrac{\phi(x,y)}{\phi_Y(y)}$。如果 X 与 Y 是相互独立的，那么 $\phi(x:B)=\phi_X(x)$。同理可定义 Y 在条件 A 下的条件分布，兹不赘述。

6.7.5 两个随机变量的相关系数

现在看一个关于 XY 的命题：$E(XY)=E(X)E(Y)+\mathrm{cov}(X,Y)$。因为

$$\begin{aligned}\mathrm{cov}(X,Y) &= E[(X-E(X))(Y-E(Y))]\\ &= E(XY-E(X)Y-E(Y)X+E(X)E(Y))\\ &= E(XY)-E(X)E(Y).\quad\square\end{aligned}$$

将随机变量 X 和 Y 分别标准化成 X^* 和 Y^*，然后求相关矩 $\mathrm{cov}(X^*,Y^*)$，称之为相关系数 (correlation coefficient)，常记为 r。显见 $\mathrm{cov}(X,Y)=r\,\sigma_x\sigma_y$。易证 $|r|\le 1$，而且当 $Y=kx+b$ 时，等号成立。因此，r 又称线性相关系数。

如果随机变量 X 和 Y 相互独立，则 $\mathrm{cov}(X,Y)=0$，因此 $r=0$。但是，$r=0$ 并不能保证 X 和 Y 相互独立。

如果随机变量 X 服从 $N(a_x,{\sigma_x}^2)$，Y 服从 $N(a_y,{\sigma_y}^2)$，两者不是相互独立的，相关系数为 r，则它们的联合分布密度为

$$\phi(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\exp\left[-\frac{1}{2(1-r^2)}\left(\frac{(x-a_x)^2}{{\sigma_x}^2}-\frac{2r(x-a_x)(y-a_y)}{\sigma_x\sigma_y}+\frac{(y-a_y)^2}{{\sigma_y}^2}\right)\right].$$

6.8 随机变量的数字特征

6.8.1 随机变量的各阶原点矩与中心矩、偏态与峰态

给定随机变量 X 和正整数 k，称 $E(X^k)$ 为 X 的 k 阶原点矩 (k-th origin moment，也称 raw moment of order k)，记为 ν_k；称 $E((X-E(X))^k)$ 为 X 的

k 阶中心矩 (k-th central moment)，记为 μ_k。显然，一阶原点矩是期望值，一阶中心矩恒为零，二阶中心矩是方差。

如果随机变量 X 的分布密度或者分布列关于期望值是对称的，那么它的一切奇数阶中心矩都为零。特别是，对于正态分布，$\mu_3 = 0$。

注意到，μ_1 不可能不为零。那么，如果 μ_3 不为零，就意味着该变量的分布关于期望值不对称。因此，称 $\dfrac{\mu_3}{\sigma^3}$ 为偏态系数 (coefficient of skewness)，记为 γ_s。大致上，如果概率的峰值靠左，会带动期望值靠左，期望值右边的概率比重增加，称该分布偏向 (skew) 均值的右侧，求 μ_3 时，正值的权重增加，因此偏态为正。反之，如果概率峰值靠右，则偏态为负。

对于正态分布，有 $\mu_4 = 3\sigma^4$。称 $\dfrac{\mu_4}{\sigma^4} - 3$ 为峰态系数 (coefficient of kurtosis)，记为 γ_e，或 γ_k。大致上，如果概率峰峰顶的平台比正态分布的窄，那么将会有更多的概率分配给两边，求 μ_4 时，大离差的权重增加，因此，峰态为正。反之，如果峰顶的平台比正态分布的宽，则峰态为负。

6.8.2 中心矩与各阶原点矩的关系

注意到 $\nu_1 = E(X)$，由二项式定理，有

$$
\begin{aligned}
(X-\nu_1)^2 &= X^2 - 2X\nu_1 + {\nu_1}^2 \\
(X-\nu_1)^3 &= X^3 - 3X^2\nu_1 + 3X{\nu_1}^2 - {\nu_1}^3 \\
(X-\nu_1)^4 &= X^4 - 4X^3\nu_1 + 6X^2{\nu_1}^2 - 4X{\nu_1}^3 + {\nu_1}^4
\end{aligned}
$$

对各式求期望值，就得到中心矩与各阶原点矩的关系：

$$
\begin{aligned}
\mu_2 &= \nu_2 - 2\nu_1\nu_1 + {\nu_1}^2 \\
\mu_3 &= \nu_3 - 3\nu_2\nu_1 + 3\nu_1{\nu_1}^2 - {\nu_1}^3 \\
\mu_4 &= \nu_4 - 4\nu_3\nu_1 + 6\nu_2{\nu_1}^2 - 4\nu_1{\nu_1}^3 + {\nu_1}^4
\end{aligned}
$$

实践中，常把连续型随机变量 X 的主要取值范围均等划分成宽为 Δx 的一系列区间，各区间中点的值为 x_i，然后统计样本落在这些区间上的频率，进而计算样本的偏态和峰态。注意到 $x_i = c + i\Delta x$，即 $i = \dfrac{x_i - c}{\Delta x}$。这也是一个随机变量。它的各阶原点矩相对容易计算，涉及的数字多是整数，平方数、立方数等。进而可求 $\mu_k(i)$。最后，由 $\mu_k(x) = (\Delta x)^k \mu_k(i)$ 即可求得 X 的中心矩。

6.9 参数估计

6.9.1 个体、样本、总体、统计量与观测值

通常，为了确定一个随机变量 X 的分布，需要进行多次实验。各次实验的结果在理论上都是随机变量，称为个体，记为 X_i。整个这一组随机变量称为

X 的一个样本，记为 $\{X_1,\ldots,X_n\}$。实验次数 n 称为样本容量。某次实验的实际值 x_i 称为对应的个体 X_i 的观测值。随机变量 X 本身则被称为总体。

函数 $f(X_1,\ldots,X_n)$ 称为统计量，把观测值代入，得数值 $f(x_1,\ldots,x_n)$，称为该统计量的观测值。

6.9.2 样本平均值、样本方差与修正样本方差

最常用的统计量是样本平均值 $\overline{X}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i$，样本方差 $S^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2$，修正样本方差 $S^{*2}=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2$，

因为 $E(\overline{X})=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}E(X_i)=\dfrac{1}{n}na=a$，所以说 $\overline{X}$ 是对 a 的无偏估计。

但是，因为

$$\begin{aligned}
E(S^2) &= \frac{1}{n}\sum_{i=1}^{n}E((X_i-\overline{X})^2)\\
&= \frac{1}{n}\sum_{i=1}^{n}E(((X_i-a)-(\overline{X}-a))^2)\\
&= \frac{1}{n}\sum_{i=1}^{n}E((X_i-a)^2-2(X_i-a)(\overline{X}-a)+(\overline{X}-a)^2)\\
&= \left(\frac{1}{n}\sum_{i=1}^{n}E((X_i-a)^2)\right)-E((\overline{X}-a)^2)\\
&= \frac{1}{n}n\sigma^2-D(\overline{X})\\
&= \sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2,
\end{aligned}$$

所以，S^2 不是 σ^2 的无偏估计。很明显，偏差来自于 $D(\overline{X})$。S^2 不是以 a 为基准计算的，而是以 $\overline{X}$ 为基准。这个基准本身有方差。

因为 $S^{*2}=\dfrac{n}{n-1}S^2$，所以 $E(S^{*2})=\sigma^2$，即 S^{*2} 是 σ^2 的无偏估计。这是因为 S^{*2} 对 $\overline{X}$ 的方差造成的影响进行了恰如其分的修正。

实际上，任何一个个体 X_i 都是对 a 的无偏估计。我们之所以采用 $\overline{X}$ 来做估计，是因为 $D(\overline{X})=\dfrac{\sigma^2}{n}\leq\sigma^2=D(X_i)$。也就是说，$\overline{X}$ 作为对 a 的估计，比 X_i 更有效。

6.9.3 参数估计的置信概率与置信区间

普通的实验数据处理，都假设目标数据 X 服从正态分布 $N(a,\sigma^2)$。把 $\overline{x}$ 作为对 a 的估计，把 s^* 作为对 σ 的估计，从而把测量结果表示成 $\overline{x}\pm s^*$，并

称 s^* 为误差。忽略估计本身的偏差，按照正态分布，我们大致上可以说 X 取值在区间 $(\overline{x}-s^*, \overline{x}+s^*)$ 之内的概率为 0.683，在区间 $(\overline{x}-2s^*, \overline{x}+2s^*)$ 之内的概率为 0.954，在区间 $(\overline{x}-3s^*, \overline{x}+3s^*)$ 之内的概率为 0.997。这种区间和概率与对 a 和 σ 的估计的置信区间 (confidence interval, 或 fiducial range) 和置信概率 (confidence probability) 完全是两码事。尽管置信概率 α 一般也取 0.95 或 0.99，但是统计量作为随机变量的函数，一般不服从正态分布，所以置信区间一般不能用二倍或三倍标准差来表示。

根据大数定律，随着样本容量 n 的无限增大，$\overline{X}$ 依概率收敛于 a，即：百分之百只取 a 这一个值，不再具有随机性。当 n 为有限值时，$\overline{X}$ 也有一个分布。如果 X 服从正态分布 $N(a, \sigma^2)$，那么 $\overline{X}$ 服从正态分布 $N(a, \dfrac{\sigma^2}{n})$。忽略估计本身的偏差，按照正态分布，我们大致上可以……等一下，我们是想说 $\overline{x}$ 作为对 a 的估计有多大偏差吗？既然如此，“忽略估计本身的偏差”还有意义么？所以说，上段的逻辑在这里行不通了。当 $n > 10$（或 30）时，如果要求精度不高，我们大致可以说 $\overline{X}$ 取值在区间 $(\overline{x}-\dfrac{s^*}{\sqrt{n}}, \overline{x}+\dfrac{s^*}{\sqrt{n}})$ 之内的概率为 0.683 等等，但是如果 $n \le 10$（或 30），或者要求精度较高，那就要重新考虑了。

6.9.4 自由度为 k 的卡方分布 $\chi^2(k)$

如果总体 X 服从正态分布 $N(a, \sigma^2)$，那么统计量 $\dfrac{S^{*2}}{\sigma^2} = \dfrac{1}{n-1}\dfrac{nS^2}{\sigma^2}$ 服从 $\dfrac{1}{n-1}\chi^2(n-1)$。其中 χ^2 读作“卡方”。$\chi^2(k)$ 称为自由度为 k 的 χ^2 分布。它的分布密度是，当 $x > 0$ 时，

$$\phi(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} \exp(-\frac{x}{2}),$$

当 $x \le 0$ 时，$\phi(x) = 0$。该分布的另一个来源是：如果一组独立随机变量 $X_1, \ldots, X_k$ 都服从标准正态分布 $N(0,1)$，那么它们的平方和 ${X_1}^2 + \cdots + {X_k}^2$ 服从 $\chi^2(k)$ 分布。

显见，如果独立随机变量 X 和 Y 分别服从 $\chi^2(k_x)$ 和 $\chi^2(k_y)$ 分布，那么 $X+Y$ 服从 $\chi^2(k_x+k_y)$ 分布。

上段中的函数 $\Gamma(x)$ 定义为 $\displaystyle\int_{t>0} t^{x-1}e^{-t}\,\mathrm{d}t$。可证 $\Gamma(x) = (x-1)\Gamma(x-1)$，而 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$，$\Gamma(1) = 1$。因此，

$$\begin{aligned}\Gamma(n+\frac{1}{2}) &= (n-\frac{1}{2})(n-\frac{3}{2})\cdots\frac{1}{2}\Gamma(\frac{1}{2}) \\ &= \frac{(2n-1)!!}{2^n}\sqrt{\pi}\end{aligned}$$

而 $\Gamma(n+1) = n!$。

6.9.5 自由度为 k 的学生分布 $t(k)$

如果随机变量 X 与 Y 相互独立，X 服从 $N(0,1)$，而 Y 服从 $\chi^2(k)$，那么随机变量 $T=\dfrac{X}{\sqrt{Y/k}}$ 服从自由度为 k 的 t 分布，又称"学生"分布。英国化学家、数学家希利 • 戈塞特 (William Sealey Gosset) 在 1908 年提出小样本平均数的误差问题，但是他所在的酿酒厂禁止员工发表与酿酒工艺有关的文章，所以他只好以"学生 (student)"为笔名发表统计学论文。t 分布的分布密度为

$$\phi(x)=\frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})}\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}}.$$

如果总体 X 服从正态分布 $N(a,\sigma^2)$，那么 $\overline{X}$ 与 $\dfrac{S^{*2}}{\sigma^2}$ 相互独立。因此，综合前几段所述，$T=\dfrac{\frac{\overline{X}-a}{\sigma/\sqrt{n}}}{\sqrt{S^{*2}/\sigma^2}}=\dfrac{\overline{X}-a}{S/\sqrt{n-1}}$ 服从 $t(n-1)$ 分布。

6.9.6 期望值与方差的置信概率和置信区间

要考虑 $\overline{X}$ 作为对 a 的估计，其置信区间 $(a-\varepsilon,a+\varepsilon)$ 与置信概率 α 的关系，即 $p(|\overline{X}-a|<\varepsilon)=\alpha$，也就是 $p\left(\frac{|\overline{X}-a|}{S/\sqrt{n-1}}<\frac{\varepsilon}{S/\sqrt{n-1}}\right)=\alpha$，注意到 T 服从 $t(n-1)$ 分布，记 $t_{\frac{1-\alpha}{2}}=\dfrac{\varepsilon}{S/\sqrt{n-1}}$，上式就是 $p(|T|<t_{\frac{1-\alpha}{2}})=\alpha$。给定 α，可以查表得出 $t_{\frac{1-\alpha}{2}}$，进而得出 ε，最终得到 $p\left(|\overline{X}-a|<\frac{s}{\sqrt{n-1}}t_{\frac{1-\alpha}{2}}\right)=\alpha$。

要考虑 S^{*2} 作为对 σ^2 的估计，其置信区间与置信概率的关系，注意到 $\dfrac{(n-1)S^{*2}}{\sigma^2}=\dfrac{nS^2}{\sigma^2}$ 服从 $\chi^2(n-1)$ 分布。给定置信概率 α，查表可得 ${\chi^2}_{\frac{1+\alpha}{2}}$ 和 ${\chi^2}_{\frac{1-\alpha}{2}}$ 使得 $p(\chi^2\geq{\chi^2}_{\frac{1+\alpha}{2}})=(1+\alpha)/2$，和 $p(\chi^2\geq{\chi^2}_{\frac{1-\alpha}{2}})=(1-\alpha)/2$，进而 $p({\chi^2}_{\frac{1+\alpha}{2}}<\dfrac{nS^2}{\sigma^2}<{\chi^2}_{\frac{1-\alpha}{2}})=\alpha$，即对应于置信概率 α，σ^2 的置信区间为 $\left(\dfrac{ns^2}{{\chi^2}_{\frac{1-\alpha}{2}}},\dfrac{ns^2}{{\chi^2}_{\frac{1+\alpha}{2}}}\right)$，也可写成 $\left(\dfrac{(n-1)s^{*2}}{{\chi^2}_{\frac{1-\alpha}{2}}},\dfrac{(n-1)s^{*2}}{{\chi^2}_{\frac{1+\alpha}{2}}}\right)$，但是因为没有直接计算 s^{*2} 的简单方法，还是要先计算 s^2，所以，改用 s^{*2} 来表示 σ^2 的置信区间并没有多少实际意义。显然，对应于置信概率 α，σ 的置信区间为 $\left(\sqrt{\dfrac{n}{{\chi^2}_{\frac{1-\alpha}{2}}}}\,s,\sqrt{\dfrac{n}{{\chi^2}_{\frac{1+\alpha}{2}}}}\,s\right)$,

为什么 a 的置信区间只涉及 $t_{\frac{1-\alpha}{2}}$，而 σ^2 的置信区间会涉及 ${\chi^2}_{\frac{1+\alpha}{2}}$ 和 ${\chi^2}_{\frac{1-\alpha}{2}}$ 呢？因为 t 分布是关于原点对称的，$t_{\frac{1+\alpha}{2}}=-t_{\frac{1-\alpha}{2}}$，所以，$p(t_{\frac{1+\alpha}{2}}<T<t_{\frac{1-\alpha}{2}})$ 可以写成 $p(|T|<t_{\frac{1+\alpha}{2}})$。而 χ^2 分布恒大于零，且分布曲线没有对称轴，${\chi^2}_{\frac{1+\alpha}{2}}$ 和 ${\chi^2}_{\frac{1-\alpha}{2}}$ 之间没有简单关系，所以，无法做象 T 那样的简化。

6.10 统计推断

6.10.1 显著性水平

实践中往往需要判断某个统计量是否与某个值有显著差异。比如一批产品经过工艺处理后，该保留的特性是否保留了，该提高的指标是否提高了，等等。这样就需要人为指定一个显著性水平 α。所谓“狗咬人不是新闻，人咬狗才是新闻”，显著的东西总是少数，越是罕见就越是扎眼，即：概率越低就越是显著。所以，我们用小概率来刻画显著性水平，一般取 $\alpha = 0.05$ 或 0.01。

假设一个统计量在置信概率 α 下有相应的置信区间，那么该统计量的单次实验的观测值有大概率 α 处于该置信区间之内。如果本次实际观测值不在置信区间之内，那么，这一事实是“难以置信”的，除非前面所做的假设不成立。逻辑上有个反证法，假设 A 成立，那么必定有 B 成立，然而实验的事实是 B 不成立，那么我们就推断 A 一定不成立。

6.10.2 小概率事件的实际不可能原理

现在，我们在统计推断中引入“小概率事件的实际不可能原理”：如果事件 A 发生的概率很小，比如低于 0.05 或 0.01，那么实践上可以认为，在单次实验中 A 不会发生。据此，如果在本次实验中，“难以置信”的事实竟然发生了，那么我们不说“这次的运气不好，认了，希望下次运气能好些”，而是要推断前面所做的假设不成立。

注意到，如果置信概率 α 是大概率 0.95 或 0.99，那么“难以置信”的、“显著”的事实发生的概率就是 $1-\alpha$，是小概率 0.05 或 0.01。因此，通常就认为，置信概率与显著性水平是互补的概念，两者之和等于 1。

6.10.3 “显著性水平的取值”和“对显著性的要求”

举例来说。100 个产品中检出 17 个次品，问是否可以说这批产品的次品率为 10%。似乎可以说是，因为次品率 10% 并不意味着 100 个里面必定有 10 个，随机差异是可能的。似乎也可以说不是，因为差异为 7，好象太大了点。真正要分析这个问题，必须先选择显著性水平 α。

选 $\alpha = 0.05$。由德莫威尔-拉普拉斯定理，$p(|\frac{m}{n} - p| > \delta_{0.05}) \approx 1 - 2\Phi\left(\delta_{0.05}\sqrt{\frac{n}{pq}}\right) = 0.05$，即 $\Phi\left(\delta_{0.05}\sqrt{\frac{n}{pq}}\right) = 0.475$。查表得 $\delta_{0.05}\sqrt{\frac{n}{pq}} = 1.96$。把 $n = 100,\quad p = 0.1,\quad q = 0.9$ 代入，得 $\delta_{0.05} = 0.0588$。这批产品次品率为 $w = 0.17$，与所称的概率 $p = 0.10$ 相差 $w - p = 0.07 > 0.0588$，因此可以认为这个差异是显著的，不属于正常的随机差异，不能接受“这批产品的次品率为 10%”的说法。

如果选择显著性水平 $\alpha = 0.01$，那么 $\Phi\left(\delta_{0.01}\sqrt{\frac{n}{pq}}\right) = (1-0.01)/2 = 0.495$。查表得 $\delta_{0.01}\sqrt{\frac{n}{pq}} = 2.575$。把 n, p, q 代入，得 $\delta_{0.01} = 0.077$。这批产品次品率 w 与所称的概率 p 相差 $w - p = 0.07 < 0.077$，因此可以认为这个差异不是显著的，属于正常的随机差异，可以接受“这批产品的次品率为 10%”的说法。

可见，选择较低的显著性水平，实际上是提高了对显著性的要求，某些原本显著的，会变成不显著的。“显著性水平的取值”和“对显著性的要求”居然按相反的方向变化，这一点很让人犯晕。

以下是一些常见的统计推断问题。我们总设所选择的显著性水平为 α。

6.10.4 期望值与给定值是否有显著差异

假设总体 X 服从正态分布 $N(a, \sigma^2)$，根据一个样本观测值 $(x_i : 1 \le i \le n)$，是否可以认为 $a = a_0$？在讨论 a 的置信概率和置信区间时，我们已经谈到 $p(|T| > t_{\frac{\alpha}{2}}) = \alpha$。查表即可。

6.10.5 标准差与给定值是否有显著差异

假设总体 X 服从正态分布 $N(a, \sigma^2)$，根据一个样本观测值 $(x_i : 1 \le i \le n)$，是否可以认为 $\sigma = \sigma_0$？在讨论 σ 的置信概率和置信区间时，我们已经谈到 $p(\chi^2 \ge {\chi^2}_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ 和 $p(\chi^2 \ge {\chi^2}_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$。查表即可。

6.10.6 自由度为 (k_1, k_2) 的 F 分布 $F(k_1, k_2)$

如果独立随机变量 X 服从 $\chi^2(k_1)$，Y 服从 $\chi^2(k_2)$，那么统计量 $F = \dfrac{X/k_1}{Y/k_2}$ 的分布密度为：当 $z > 0$ 时，

$$\phi_F(z) = \frac{\Gamma(\frac{k_1+k_2}{2})}{\Gamma(\frac{k_1}{2})\Gamma(\frac{k_2}{2})} {k_1}^{\frac{k_1}{2}} {k_2}^{\frac{k_2}{2}} \frac{z^{\frac{k_1}{2}-1}}{(k_1 z + k_2)^{\frac{k_1+k_2}{2}}},$$

当 $z \le 0$ 时，$\phi_F(z) = 0$。称这个分布是自由度为 (k_1, k_2) 的 F 分布，记为 $F(k_1, k_2)$。称其中的 k_1 为第一自由度，k_2 为第二自由度。

给定实数 $0 < \alpha < 1$，通常可以查表得到 F_α 使得 $p(F \ge F_\alpha) = \alpha$。有 $F_{1-\alpha}(k_1, k_2) F_\alpha(k_2, k_1) = 1$。

如果独立总体 X 和 Y 分别服从正态分布 $N(a_x, \sigma^2)$ 和 $N(a_y, \sigma^2)$，那么统计量 $\dfrac{S_x^{*2}}{S_y^{*2}}$ 服从 $F(n_x - 1, n_y - 1)$。由 $\dfrac{S^{*2}}{\sigma^2}$ 服从 $\dfrac{1}{n-1}\chi^2(n-1)$ 分布立得。

6.10.7 两标准差是否有显著差异

假设独立总体 X 和 Y 分别服从正态分布 $N(a_x, {\sigma_x}^2)$ 和 $N(a_y, {\sigma_y}^2)$，根据样本观测值 $(x_i : 1 \leq i \leq n_x)$ 和 $(y_j : 1 \leq j \leq n_y)$，是否可以认为 $\sigma_x = \sigma_y$？先计算 $F = \dfrac{{s_x^*}^2}{{s_y^*}^2}$，查表看它是否在置信区间之内即可。

6.10.8 两标准差无差异时，两期望值是否有显著差异

同上段，如果已知 $\sigma_x = \sigma_y = \sigma$，问是否可以认为 $a_x = a_y$？因为 $X+Y$ 服从 $N(a_x + a_y, {\sigma_x}^2 + {\sigma_y}^2)$，所以 $\overline{X} - \overline{Y}$ 服从 $N\left(a_x - a_y, \left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right) \sigma^2\right)$。另一方面，$\dfrac{n_x S_x^2 + n_y S_y^2}{\sigma^2}$ 服从 $\chi^2(n_x + n_y - 2)$。因此，统计量

$$T = \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}} \frac{(\overline{X} - \overline{Y}) - (a_x - a_y)}{\sqrt{n_x S_x^2 + n_y S_y^2}}$$

服从 $t(n_x + n_y - 2)$ 分布。假设 $a_x = a_y$，根据 T 的观测值是否在置信区间之内，即可推断所做假设是否合理为真。 $\square$

6.10.9 统计分布与理论分布有无显著差异

进行 N 次独立试验，得随机变量 X 的统计分布，即：将数轴划分成 n 个区间，各区间频数为 $m_1, \ldots, m_n$。问是否可以认为该统计分布与某个理论分布 $p_1, \ldots, p_n$ 的差异是不显著的？

皮尔逊 (Pearson) χ^2 准则：差异度 $U = \displaystyle\sum_{i=1}^{n} \frac{N}{p_i}(w_i - p_i)^2$ 在 $N \to \infty$ 时趋于 $\chi^2(n - r - 1)$ 分布，其中 r 是理论分布中需要利用样本观测值去估计的未知参数的个数。

因此，只要计算 U 的观测值，再查表看它是否在置信区间之内即可。要注意的是，$N \to \infty$ 这个条件在实践中一般取 $N \geq 50$。另外，应该使每个频数 $m_i \geq 5$。如有必要，可把相邻区间合并。 $\square$

6.10.10 因素 A 的不同水平对期望值是否有显著影响

在因素 A 的 m 个不同水平 $A_1, \ldots, A_m$ 下做试验，对于 $1 \leq i \leq m$，设水平 A_i 下的试验为总体 X_i，得 X_i 的观测值 $x_{i1}, \ldots, x_{in_i}$。假定 X_i 服从 $N(a_i, \sigma^2)$，即因素 A 不影响标准差，但可能影响期望值。问：根据试验结果，是否可以认为因素 A 对期望值无显著影响，即 $a_1 = \cdots = a_m$？

当 $m = 2$ 时，如前所述，可以用与 $\overline{X}_1 - \overline{X}_2$ 相关的 t 分布来判断。如果 $m > 2$，就需要用组间平均平方和 $\overline{U}_A$ 与组内平均平方和 $\overline{U}_e$ 的比做成的 F 分布来判断。

记试验总次数为 $n = \sum_{i=1}^{m} n_i$，组平均值为 $\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$，总平均值为 $\overline{X} = \frac{1}{n}\sum_{i=1}^{m} n_i\overline{X}_i$。那么，总的离差平方和为

$$\begin{aligned}
U &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij}-\overline{X})^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n_i}[(\overline{X}_i-\overline{X})+(X_{ij}-\overline{X}_i)]^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n_i}[(\overline{X}_i-\overline{X})^2+(X_{ij}-\overline{X}_i)^2+2(\overline{X}_i-\overline{X})(X_{ij}-\overline{X}_i)] \\
&= \sum_{i=1}^{m} n_i(\overline{X}_i-\overline{X})^2+\sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2 \\
&= U_A + U_e
\end{aligned}$$

其中组间平方和 U_A 反映由 A 的不同水平而引起的差异，而 $X_{ij}-\overline{X}_i = (X_{ij}-\overline{X})-(\overline{X}_i-\overline{X})$，所以，组内平方和 U_e 反映 A 的相同水平下的随机差异。

注意到 $\frac{U}{\sigma^2} = \frac{nS^2}{\sigma^2}$ 服从 $\chi^2(n-1)$ 分布，而 $\frac{U_e}{\sigma^2} = \sum_{i=1}^{m}\frac{n_iS_i^2}{\sigma^2}$ 服从 $\chi^2(\sum_{i=1}^{m}(n_i-1)) = \chi^2(n-m)$ 分布，又可证 U_A 和 U_e 相互独立，因此，U_A 服从 $\chi^2(m-1)$ 分布。

称 $\overline{U}_A = \frac{U_A}{m-1}$ 为组间平均平方和，称 $\overline{U}_e = \frac{U_e}{n-m}$ 为组内平均平方和，那么 $F = \frac{\overline{U}_A}{\overline{U}_e}$ 服从 $F(m-1,n-m)$ 分布。据此查表即可。 □

6.10.11 因素 A 和因素 B 的不同水平对期望值是否有显著影响

因素 A 有 ℓ 个不同水平 $A_1,\ldots,A_\ell$，因素 B 有 m 个不同水平 $B_1,\ldots,B_m$，在每一个水平 (A_i,B_j) 下都做一次试验，得数据 x_{ij}。假设 X_{ij} 服从正态分布 $N(a_{ij},\sigma^2)$，即因素水平不影响标准差，但可能影响期望值。问：根据这些数据，是否可以认为因素 A 和 B 对期望值都无显著影响？

记 $\overline{X}_{i\cdot} = \frac{1}{m}\sum_{j=1}^{m} X_{ij}$，$\overline{X}_{\cdot j} = \frac{1}{\ell}\sum_{i=1}^{\ell} X_{ij}$，总的离差平方和为

$$\begin{aligned}
U &= \sum_{i=1}^{\ell}\sum_{j=1}^{m}(X_{ij}-\overline{X})^2 \\
&= \sum_{i=1}^{\ell}\sum_{j=1}^{m}[(\overline{X}_{i\cdot}-\overline{X})+(\overline{X}_{\cdot j}-\overline{X})+(X_{ij}-\overline{X}_{i\cdot}-\overline{X}_{\cdot j}+\overline{X})]^2
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{\ell}\sum_{j=1}^{m}[(\overline{X}_{i\cdot}-\overline{X})^2+(\overline{X}_{\cdot j}-\overline{X})^2+(X_{ij}-\overline{X}_{i\cdot}-\overline{X}_{\cdot j}+\overline{X})^2+2\cdots] \\
&= m\sum_{i=1}^{\ell}(\overline{X}_{i\cdot}-\overline{X})^2+\ell\sum_{j=1}^{m}(\overline{X}_{\cdot j}-\overline{X})^2+\sum_{i=1}^{\ell}\sum_{j=1}^{m}(X_{ij}-\overline{X}_{i\cdot}-\overline{X}_{\cdot j}+\overline{X})^2 \\
&= U_A+U_B+U_e
\end{aligned}
$$

注意到 $X_{ij}-\overline{X}_{i\cdot}-\overline{X}_{\cdot j}+\overline{X}=(X_{ij}-\overline{X})-(\overline{X}_{i\cdot}-\overline{X})-(\overline{X}_{\cdot j}-\overline{X})$，所以 U_e 是随机差异。

显见 $\dfrac{U}{\sigma^2}=\dfrac{\ell m S^2}{\sigma^2}$ 服从 $\chi^2(\ell m-1)$ 分布，$\dfrac{U_A}{\sigma^2}$ 服从 $\chi^2(\ell-1)$ 分布，$\dfrac{U_B}{\sigma^2}$ 服从 $\chi^2(m-1)$ 分布，又可证 U_A,U_B,U_e 相互独立，因此，$\dfrac{U_e}{\sigma^2}$ 服从 $\chi^2((\ell-1)(m-1))$ 分布。根据 $F_A=\dfrac{\overline{U}_A}{\overline{U}_e}$ 和 $F_B=\dfrac{\overline{U}_B}{\overline{U}_e}$ 即可查表推断因素 A 和 B 是否对期望值有显著影响。 □

6.10.12 两随机变量之间是否有显著的线性相关关系

设总体是二维随机变量 (X,Y)，如果它的样本 $\{(X_i,Y_i):1\le i\le n\}$ 的观测值 $\{(x_i,y_i)\}$ 满足 $y_i\approx a+bx_i,\quad (b\ne 0)$，则称 X 和 Y 之间存在线性相关关系，称 $y=a+bx$ 为线性回归 (regression) 方程。注意这里的线性相关与线性代数中一组向量的线性相关是同名不同义，不可混淆。问：根据样本观测值 $\{(x_i,y_i)\}$，是否可以认为 X 与 Y 之间没有显著的线性相关关系？

记 $\hat{y}_i=a+bx_i$。考虑 Y 的总的离差平方和

$$
\begin{aligned}
U &= \sum_{i=1}^{n}(y_i-\overline{y})^2 \\
&= \sum_{i=1}^{n}[(\hat{y}_i-\overline{y})+(y_i-\hat{y}_i)]^2 \\
&= \sum_{i=1}^{n}[(\hat{y}_i-\overline{y})^2+(y_i-\hat{y}_i)^2+2(\hat{y}_i-\overline{y})(y_i-\hat{y}_i)] \\
&= \sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \\
&= U_g+U_s
\end{aligned}
$$

其中 U_g 称为回归平方和，U_s 称为剩余平方和。因为 $y_i-\hat{y}_i=(y_i-\overline{y})-(\hat{y}_i-\overline{y})$，所以剩余平方和反映随机差异。

如果 X 与 Y 之间没有显著的线性相关关系，那么可证 $\dfrac{U}{\sigma^2}$ 服从 $\chi^2(n-1)$，$\dfrac{U_g}{\sigma^2}$ 服从 $\chi^2(1)$，而且 U_g 与 U_s 相互独立，因此，U_s 服从 $\chi^2(n-2)$。从而 $F=\dfrac{U_g}{U_s/(n-2)}$ 服从 $F(1,n-2)$ 分布。据此查表即可。 □

此问题还有另一种解法。记 $\ell_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$，$\ell_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$，$\ell_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$，那么样本相关系数

$$r = \frac{\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{\ell_{xy}}{\sqrt{\ell_{xx}\ell_{yy}}}.$$

可证 $F = \dfrac{(n-2)r^2}{1-r^2}$，即 $|r| = \sqrt{\dfrac{F}{F+n-2}}$。直接计算 r 并查 $|r|$ 的表即可。 □

6.10.13 最小二乘法

最小二乘法假设对于 $1 \leq i \leq n$，试验 Y_i 服从正态分布 $N(f(x_i), \sigma^2)$。根据极大似然原理，似然函数是这些分布密度的积。它取极大值的条件是 $U = \sum_{i=1}^{n}[y_i - f(x_i)]^2$ 取极小值。

特别是，如果设 $f(x) = a + bx$，那么应当使 $U = \sum_{i=1}^{n}(y_i - a - bx_i)^2$ 取极值，即 U 不随 a 和 b 的微小变动而变动。注意到

$$\begin{aligned}
U_{a+\Delta a, b+\Delta b} &= \sum_{i=1}^{n}[y_i - (a+\Delta a) - (b+\Delta b)x_i]^2 \\
&= \sum_{i=1}^{n}[(y_i - a - bx_i) - \Delta a - \Delta b\, x_i]^2 \\
&= \sum_{i=1}^{n}[(y_i - a - bx_i)^2 - 2(y_i - a - bx_i)\Delta a \\
&\qquad -2(y_i - a - bx_i)\Delta b\, x_i + \text{高阶小项}\,]
\end{aligned}$$

要使得 U 不随 a 和 b 的微小变动而变动，就应该使 Δa 和 Δb 的系数恒为零，即

$$\begin{cases}
\sum_{i=1}^{n}(y_i - a - bx_i) = 0 \\
\sum_{i=1}^{n}(y_i - a - bx_i)x_i = 0
\end{cases}$$

展开就可以解出 a 和 b。 □

6.10.14 回归预测

利用回归方程进行预测是指预测下一次试验 X_{n+1}。如果其观测值是 x_{n+1}，那么 Y_{n+1} 的分布如何，给定置信概率，相应的置信区间，称为预测区间，如何。

可证，当 $n\to\infty$ 时，Y_{n+1} 服从 $N(\hat{y}_{n+1},{\sigma_s}^2)$。其中 $\hat{y}_{n+1}=a+bx_{n+1}$ 是回归直线上的点，$\sigma_s=\sqrt{\dfrac{U_s}{n-2}}$ 称为剩余标准差。据此分布可求预测区间，比如 Y_{n+1} 落在 $\hat{y}_{n+1}\pm 2\sigma_s$ 内的概率为 0.954 等等。 □

6.10.15 统计对象的两个属性是否相互独立

在 n 个统计对象中，具有属性 A 的有 n_A 个，具有属性 B 的有 n_B 个，更详细的情况如下表所示

	B	$\overline{B}$	
A	n_{AB}	$n_{A\overline{B}}$	n_A
$\overline{A}$	$n_{\overline{A}B}$	$n_{\overline{AB}}$	$n_{\overline{A}}$
	n_B	$n_{\overline{B}}$	n

其中的每一个数都大于 5，$\overline{A}$ 表示 A^{C}，$\overline{B}$ 表示 B^{C}。问：统计对象具有属性 A 和具有属性 B 这两种事件是否相互独立？

如果它们相互独立，则 $p(A\cap B)=p(A)p(B)$，$p(\overline{A}\cap B)=p(\overline{A})p(B)$，$p(A\cap\overline{B})=p(A)p(\overline{B})$，$p(\overline{A}\cap\overline{B})=p(\overline{A})p(\overline{B})$。可证

$$\begin{aligned}U &= \frac{(\frac{n_{AB}}{n}-\frac{n_A}{n}\frac{n_B}{n})^2}{\frac{n_A}{n}\frac{n_B}{n}}+\frac{(\frac{n_{A\overline{B}}}{n}-\frac{n_A}{n}\frac{n_{\overline{B}}}{n})^2}{\frac{n_A}{n}\frac{n_{\overline{B}}}{n}}+\frac{(\frac{n_{\overline{A}B}}{n}-\frac{n_{\overline{A}}}{n}\frac{n_B}{n})^2}{\frac{n_{\overline{A}}}{n}\frac{n_B}{n}}+\frac{(\frac{n_{\overline{AB}}}{n}-\frac{n_{\overline{A}}}{n}\frac{n_{\overline{B}}}{n})^2}{\frac{n_{\overline{A}}}{n}\frac{n_{\overline{B}}}{n}}\\ &= \frac{(n_{AB}n_{\overline{AB}}-n_{A\overline{B}}n_{\overline{A}B})^2}{n_An_{\overline{A}}n_Bn_{\overline{B}}}\end{aligned}$$

服从 $\chi^2(1)$。查表即可。 □

Chapter 7

理财学与数学

7.1 数钱原理与测度论

7.1.1 数钱，数钱包里的钱

现在，再来看看我钱包里的这些钱。一共六张，记为 a, b, c, d, e, f，构成集合 M。其中的任意元素记为 $m \in M$。

对于集合 M，以幂集 2^M 为 σ-代数，做成可测空间 $(M, 2^M)$。我们可以给它配上计数测度 μ，做成测度空间 $(M, 2^M, \mu)$；也可以给它配上金额测度 ν，做成测度空间 $(M, 2^M, \nu)$。

金额测度 ν 在计数测度 μ 上的密度函数就是面值函数 ϕ: $M \to \mathbf{R}$，即 $\mathrm{d}\nu = \phi\,\mathrm{d}\mu$。计数测度 μ 在金额测度 ν 上的密度函数记为 ψ: $M \to \mathbf{R}$，即 $\mathrm{d}\mu = \psi\,\mathrm{d}\nu = \dfrac{1}{\phi}\,\mathrm{d}\nu$。那么，我们有

m	a	b	c	d	e	f
$\mu(\{m\})$	1	1	1	1	1	1
$\nu(\{m\})$	10	10	20	20	20	50
$\phi(m) = \dfrac{\nu(\{m\})}{\mu(\{m\})}$	10	10	20	20	20	50
$\psi(m) = \dfrac{\mu(\{m\})}{\nu(\{m\})} = \dfrac{1}{\phi(m)}$	0.1	0.1	0.05	0.05	0.05	0.02

这些钱的总金额 A 可以写成

$$A = \int_M \mathrm{d}\nu = \sum_{m \in M} \nu(\{m\}) = 10 + 10 + 20 + 20 + 20 + 50 = 130,$$

也可以写成

$$A = \int_M \phi\,\mathrm{d}\mu = \sum_{m \in M} \phi(m)\mu(\{m\})$$

$$= 10\times1+10\times1+20\times1+20\times1+20\times1+50\times1=130.$$

总张数 C 可以写成

$$C=\int_M \mathrm{d}\mu=\sum_{m\in M}\mu(\{m\})=1+1+1+1+1+1=6,$$

也可以写成

$$\begin{aligned}C &= \int_M \psi\,\mathrm{d}\nu=\sum_{m\in M}\psi(m)\nu(\{m\})\\ &= 0.1\times10+0.1\times10+0.05\times20+0.05\times20+0.05\times20+0.02\times50\\ &= 6.\end{aligned}$$

借用统计物理的术语，可以说测度 μ 和 ν 都是广延量，密度函数 ϕ 和 ψ 都是强度量。

7.1.2 数钱，数欧元、美元、日元的钱

上海世博会期间，我一天碰到三个老外，收了两张十欧元、三张二十美元、一张五十日元的外币。记它们分别为 a,b,c,d,e,f。我不知道当天中行的外汇牌价，但是，我记得有个老外，付了 a,c,f，说是可换人民币 224 元；有个老外付了 b,d，说是可换人民币 220 元；还有个老外付了 e，说是可换人民币 130 元。记“外币元测度”为 α，“可兑换人民币测度”为 β。后者在前者上的密度函数为 ϕ。

设十欧元、二十美元、五十日元当天在中行可兑换人民币 x,y,z，那么

$$\begin{cases}\beta(\{a,c,f\}) = x+y+z=224\\ \beta(\{b,d\}) = x+y=220\\ \beta(\{e\}) = y=130\end{cases}.$$

解得 $x=90,\quad y=130,\quad z=4$。

仿照上节，列表如下：

m	a	b	c	d	e	f
$\alpha(\{m\})$	10	10	20	20	20	50
$\beta(\{m\})$	90	90	130	130	130	4
$\phi(m)=\dfrac{\beta(\{m\})}{\alpha(\{m\})}$	9	9	6.5	6.5	6.5	0.08

$100\times\phi$ 就是外币现钞买入价。即，当天中行的外币现钞买入价为欧元 900.00，美元 650.00，日元 8.0000。

世博之后，我手上攒了一些欧元、美元和日元。这是我那段时间记下的外汇牌价和可兑换人民币金额：

日期	欧元牌价	美元牌价	日元牌价	手上外币可兑换人民币
20xx.4.1	900.00	650.00	8.0000	2224
20xx.5.1	850.00	640.00	7.0000	2151
20xx.6.1	950.00	700.00	9.0000	2377

不同日期的外汇牌价确定了不同的密度函数 ϕ，进而确定不同的 β 测度。设当时我有欧元、美元和日元各 x, y, z 元，可以列表如下：

手上外币 m	欧元	美元	日元
$\alpha(\{m\})$	x	y	z
$\phi_1(m)$	9.0	6.5	0.08
$\phi_2(m)$	8.5	6.4	0.07
$\phi_3(m)$	9.5	7.0	0.09

由 $\int_M \mathrm{d}\beta = \int_M \phi\,\mathrm{d}\alpha$ 可列方程组：

$$\begin{cases} 9.0\ x\ +\ 6.5\ y\ +\ 0.08\ z\ =\ 2224 \\ 8.5\ x\ +\ 6.4\ y\ +\ 0.07\ z\ =\ 2151 \\ 9.5\ x\ +\ 7.0\ y\ +\ 0.09\ z\ =\ 2377 \end{cases}.$$

解得 $x = 100, \quad y = 200, \quad z = 300$。

7.1.3 数钱，数末日之前能挣的所有的钱

设折现率为 r。如果未来各年的我能挣的钱为 $x_1, x_2, \ldots, x_n, \ldots$，那么，它们在当前的总价值，即现值，就是

$$p = \frac{1}{1+r}x_1 + \frac{1}{(1+r)^2}x_2 + \cdots + \frac{1}{(1+r)^n}x_n + \cdots.$$

各年的折现系数 $q_n = \dfrac{1}{(1+r)^n}$。实际上，考虑到其他风险，可能需要对折现系数再做调整。设最终确定的各年的折现系数为 $q_1, q_2, \ldots, q_n, \ldots$，则现值为

$$p = q_1x_1 + q_2x_2 + \cdots + q_nx_n + \cdots = \sum_{i=1}^{\infty} q_ix_i.$$

比方说，明年的一元相当于（现在的）九毛五，后年的一元相当于（现在的）九毛，这实际上就是不同的“面值”。可以设想，把世界末日之前能挣的所有的钱都放在一起，按年份划分等价类 $\{S_i : i \in \mathbf{N}\}$。同一年的钱有相同的“面值”，即 $\phi(S_i) = q_i$。同一年的钱的“数量”就是 $\mu(S_i) = x_i$。因此，这些钱的“总金额”就是现值

$$p = \int_M \mathrm{d}\nu = \int_M \phi\,\mathrm{d}\mu = \sum_{i=1}^{\infty} \phi(S_i)\mu(S_i) = \sum_{i=1}^{\infty} q_ix_i.$$

7.2 债券、股票、货币、流动性与利率

7.2.1 债券

每年年末收入一元，称为一元期末年金。它的现值记为 p_0。由于 $\mu(S_i)$ 恒等于 1。所以，

$$p_0 = \int_M \mathrm{d}\nu = \sum_{i=1}^{\infty} \phi(S_i) = \sum_{i=1}^{\infty} q_i.$$

特别是，不考虑其他风险，使用简单折现率 r 折现时，

$$p_0 = \sum_{i=1}^{\infty} \frac{1}{(1+r)^i} = \frac{1}{r}.$$

值得一提的是，如果是每年付息、最后一期付息还本的债券，那么现值与还本期限无直接关系。因为

$$\begin{aligned} p_0 &= \frac{1}{1+r} + \frac{1}{(1+r)^2} + \cdots + \frac{1}{(1+r)^m} + \frac{1}{(1+r)^{m+1}} + \cdots \\ &= \frac{1}{1+r} + \cdots + \frac{1}{(1+r)^m} + \frac{1}{(1+r)^m}\left(\frac{1}{1+r} + \cdots\right) \\ &= \frac{1}{1+r} + \cdots + \frac{1}{(1+r)^m} + \frac{1}{(1+r)^m} p_0 \end{aligned}$$

无论 m 取什么正整数，此式恒成立。因此，多数债券都采取这种付息还本方案，以简单地、直接地反映利率 r，方便投资者进行决策。

但是，利率 r 却与还本期限正相关。所谓“夜长梦多”，很多年才能还本，意味着投资者要承担更大的风险，因而会要求更高的利率。比如，中长期国债利率高于短期国债利率。

三年期国债的典型利率是 5%，所以，它的价格大约是每年利息的二十倍。也就是说，现在要付二十元，才能公平地交换到每年提现一元的权利。而银行给高端客户提供的长期信托基金，由于违约风险高于国债，客户会要求相应的风险回报，所以，典型的利率在 10% 左右。即，现在只要付十元，就能买到每年收益一元的权利。这就是（投资者认为）风险高的投资品种，只能低价销售（给投资者）的道理。

但是，投资者对投资品种的认识，是可以被“舆论导向”的。比如，控制了宣传机器的特 × 阶层，当它们想侵吞某类资产的时候，就大造舆论，说这类资产“没救了”，给谁都不会要，然后它们就可以“名正言顺”地低价买进；当它们想把手中的某类资产变现的时候，就可以大肆鼓吹，制造这种资产“越来越稀缺”，“一定会涨”，“没有风险”的假象，然后以高昂的价格销售给社会公众。一旦特 × 阶层出货完毕，民众的钱落入它们的口袋，落袋为安，舆论也就“另寻新欢”了。所以说，控制宣传机器，能带来巨大的经济利益。

当然，宣传只是最末端的教育。前端的、以及各期的教育都很重要。从小要听父母的话，上学要听老师的话，工作要听领导的话。培养听话、顺从的国民很重要。这跟养猪是一个道理。除了要解决好他们的温饱问题、住的问题、医疗问题，还要调教他们顺从，遵守秩序，和谐共处。必要时，还可以扔一些带色的皮球给他们，让他们玩，让他们咬，让他们发泄，引开他们的注意力，避免他们思考一些对猪官不利的问题。这样才会有好的猪肉，可以杀，可以吃，可以卖。

7.2.2 股票

投资股票获取权益收益与投资债券获取利息是类似的。如果一只股票的权益收益是固定的，而且全部都是现金分红，那么权益收益就是永续年金，股票价格就应该是 $P = \frac{1}{r}E$，各年除权除息之后的股价应该相差不大。

但是，有些股票不分红。投资者只能通过卖出股票而获得现金流量。如果一只股票的买入价是 P_0，那么 m 年后的卖出价应该是 $P_m = (1+r)^m P_0$，才能保本。因此，不分红的股票的市价应该是按几何级数增长的。

过高的股价会让中小投资者望而却步。所以，不分红的上市公司常常选择“股票分红”：按稍低一点的价位把权益收益折算成新增股份，分给投资者，用持股数量的按几何级数增长替换股价的按几何级数增长。

如果投资者愿意，可以将“股票分红”卖出，将它变成现金分红。实际上，即便上市公司不做股票分红，投资者也可以卖出一小部分股票，使得剩余的股票的市值与期初的基本相当，从而自己实现现金分红。

所以，无论是否分红，无论是现金分红还是股票分红，投资者的财富都是按几何级数增长的。在这个级数中，真正有意义的指标，就是上市公司的权益收益率，也就是从市场上赚取利润，并最终为投资者提供回报的能力。

当然，中国例外。在中国，目前，现金分红要缴所得税，而股票分红，卖出获得现金，只须缴印花税、支付佣金，不缴所得税。

7.2.3 货币

货币在与贵金属脱钩之前，本质上还是商品。最早的用贝壳来交易，相当于现在用各类收藏品来交易，要交易双方达成共识并不容易。后来的铜钱、银元、金币，兼有一般等价物和贵金属制品双重身份。贵金属商品市场的波动，常会影响到货币的一般等价物功能的发挥。这个矛盾有时会非常激烈。比如，满清曾经从民间大力搜集铜器，以填补铸币用铜的不足，造成的文物损失难以估量。再比如，美国历史上一度白银短缺，民众只好改用烟叶或茶叶作为支付手段进行交易。

现钞，包括纸币和硬币，一开始代表着发钞行的承诺：持此现钞可到发钞

行的任何一家分号或其代理机构兑换指定数量的贵金属货币。显见，使用现钞并不能从根本上解决上述矛盾。

直到美国政府宣布美元与黄金脱钩，凭美元不再能够兑换黄金，货币才摆脱了贵金属商品市场的纠缠，开始独立地发挥一般等价物的作用。各国的货币法实际上是改写了现钞上的承诺，其中最主要的一条就是：持此现钞用于支付不会被拒收。因此，现钞是由政府信用来担保的一般等价物凭证。

货币市场与贵金属商品市场分家之后，货币投放量不再受黄金储备的限制，很快就成了发钞行和政府干预经济的一种手段。

由于单位与个人保留的现钞通常十分有限，大部分的钱都存在银行里。因此，增加货币投放会导致银行可发放的贷款增加。这时，银行就愿意降低贷款利率，以鼓励企业吸收这些贷款。企业的资金成本降低了，也愿意扩大生产，抓住机遇多赚利润。所以，如果市场能够消化企业增加的产量，那么，“适度宽松”的货币政策可以刺激经济增长。

反之，如果经济过热，就要收缩银根，使银行可放贷的资金减少。这时，银行就会挑选那些愿意支付较高利息的企业放贷。企业的资金成本提高了，就会放弃一些扩大生产的计划。从而实现经济的软着陆，避免大起大落引发社会动荡。

7.2.4 流动性

资产，如果能在较短的时间内，比如一年之内，通过经常性的经济活动转化成现金，则称为流动资产。转化所需的时间越短，则称该资产的流动性越强。

一般来说，如果风险基本相当，那么流动性越强的资产，孳生利息的能力越低。股票的情况有点特殊。股票投资分短期投资和长期投资两种类型。如果买入股票是为了赚取短期价差，或者是为了流动性，（随时都可卖出而换成现金，）那么这些股票就是短期投资，是流动资产。如果买入股票是为了长期持有，获取股息，那就是长期投资，不是流动资产。短期股票投资的孳息能力高于其他流动资产，是因为它的风险高，而不是流动性低。

无论是企业还是个人，都希望在减值风险可以承受的前提下，尽量提高自己所拥有的资产的孳息能力。因此，当高流动性的资产过多时，人们就会卖掉过多的部分，而买进设备、家具、房产、收藏、股票等，导致这些耐用消费品和长期投资品种的价格上涨。在消费方面，当消费者手上的钱增多时，他们会扩大消费量，也会选择更高品质的消费，总之是更多的钱竞买市场上的商品，导致日用消费品价格上涨。这就是流动性泛滥导致通货膨胀的道理。

通货膨胀就是货币的实际购买力下降。在美元与黄金脱钩之前，货币对其他商品的比价也会波动，比如铜银比价的波动最终迫使清政府完全放弃了铜铸币，但是这种波动还是商品市场内部的此消彼长，除了技术进步会导致长期效

果之外，大多数波动是短期的，总是会回到合理的水平。然而脱钩之后，货币变成了由政府信用来担保的一般等价物凭证。这种依靠信用来担保的凭证对其他商品的比价是没有下限的。比如美国南北战争时期，南方政府发行的战争债券，在南方战败之后都变成了废纸。

无论在什么情况下，恶性通胀都是政府信用的瓦解，是政府经济政策的巨大失败。

有经济学家声称，温和通胀可以促进经济增长。实际上这正是美国所选择的道路。美元与黄金脱钩之后，就一直不断地贬值，以至于投资界流传这样一个笑话：睡美人一觉醒来，就拿起电话查询自己的账户，发现账上有一百亿美元。她狂喜不已。紧接着，电话中的声音说："您的免费查询时间即将用完，请支付五十亿美元，以便延续三分钟。"

温和通胀问题要从两个方面看。在生产方面，企业通常都有负债，通胀降低了资金的实际成本，因而可以促进企业扩大生产，增加市场供应。这通常是积极的，除非生产已经过剩。

在消费方面，越是穷人越无法贷款消费，而且穷人的资产少、分散、不稳定，商业金融机构为穷人提供的理财产品要收取很高的佣金。自己投资股市又没有时间、精力和渠道去关注。所以，越是穷人就越需要保留现金，也只能保留现金。而通胀，即便是温和的，也是对保留现金者的惩罚。如果对穷人没有相应的扶助和补偿，那么，温和通胀同样是"劫贫济富"。美国能够施行温和通胀，是因为美国同时采取了相配套的改善穷人福利的措施，使得穷人和中产阶级没有因为温和通胀而受到过大的伤害。

美国的长期温和通胀还造就了美国人的信贷消费习惯。借银行的钱似乎还不够，还要通过金融衍生品借世界人民的钱。最终借债太多，无力偿还，信用崩溃，造成了这次全球金融危机。所以，用温和通胀来促进经济增长是否具有可持续性，很值得商榷。

无论如何，防范恶性通胀是任何政府在经济上的头等大事。而流动性泛滥是恶性通胀的前兆，就像暴雨是洪水的前兆一样。那么，流动性为什么会泛滥呢？典型原因有三：一是央行过度发钞，二是贫富不均，三是境外热钱流入。

如果是央行发钞过度，则可以由央行来"收钞入库"，把发出去的现钞收回来束之高阁。办法主要是提高存款准备金率和由央行发行债券。

贫富不均在经济上有两方面的效应。其一，穷人虽然穷，但是如果穷人的数量庞大，那么所有穷人的资产总和也是个不小的数目。而穷人抗风险能力差，不得不让资产保持高流动性，这就限制了这些资产的孳息能力的发挥。中国政府总想把居民储蓄赶出银行，以刺激经济。但是，如果穷人缺乏保障，你还要逼着他拿保命的钱去冒险，那不是太残忍了吗？

二是，大量高流动性资产快速集中到富人手中之后，为了逐利，总是要转化成长期资产。这时候，"让哪一部分人先富起来"这个问题的答案就曝光

了。如果是私有企业的业主富起来了，那么他们通常选择扩张自己的事业，这对社会是有益的。如果是各类蛀虫富起来了，那么“不必劳就可以获”的食利型资产就会价格飞涨。

解决贫富不均问题，需要政府“鼓励竞争，保障公平”。这两者并不矛盾。如果没有最低保障，一旦失败只能跳楼，那么绝大多数人都会“好死不如赖活”地混日子，不愿冒险创业。如果市场的游戏规则不公平，垄断和特权可以上下其手，把私有企业的业主当唐僧肉，那么，市场就是有“特色”的市场，本质上是屠宰场，只对奴才和骗子有吸引力。

保障公平的政府应当遏制不劳而获的食利型资产的膨胀。比如，富人扎堆买房，抬高了房价。政府可以按住房面积实行累进税率，对合理的消费房免税，对豪宅和投资房课以重税。当然，如果政府是“为有钱的人民服务”的，那么，这办法就是与虎谋皮了。

如果是境外热钱流入，那需要分析热钱的流向。如果是进了房市，则可以限制境外机构买房；如果是进了股市，则可以趁机扩容，对整个股市实施“毒丸计划”，使热钱在股市上无利可图，从而失去对股市的兴趣。

也许有人不喜欢股市毒丸计划，认为用扩容来消化热钱，热钱固然无利可图，可是境内投资者也一样无利可图；而到了热钱退出的时候，还是要在股市上抛售，造成股价下跌，蒙受损失的还是境内投资者。

这问题要这样看：第一，如果让热钱有利可图，是不是大家就都有利可图了呢？不是！热钱利用资金优势来兴风作浪，造成过度投机的市场，搞零和游戏，在短时间内获得大量盈利，这些盈利恰恰就是境内投资者的亏损！

第二，热钱退出时的股价下跌，是不是境内投资者的亏损？也不是。热钱退出时的抛售，恰恰是境内投资者低价接盘的机会，然后股市恢复正常，这是境内投资者盈利的开始。

新股无论如何总是要发的。热钱在的时候，发的多些快些；热钱走了以后，就可以发的少些慢些。这对境内投资者来说是件好事。

信息技术的不断进步对流动性也有影响。比如，物物交换网络不需要一般等价物来做中介；各类社交和游戏网站的可用于交换现实利益的虚拟财富，冲减了对货币的需求。如果大量交易走这些渠道，被排挤出来的货币涌入其他市场，也会推高流动性。

7.2.5 利率

利率是资金使用权的价格。

前面讲过，央行增大货币投放，会导致市场利率下降；而收钞入库，会导致市场利率上升。如果这样力度还不够，达不到预期的调控效果，那么，央行还可以直接调整基准利率。这就是所谓的央行加息或者降息。

用自有资金放贷古已有之。现代银行吸收存款放贷，谋取存贷息差的业务模式发端于威尼斯。当时商人的钱如果不便携带，就存进钱庄，并支付保管费，和今天的保险柜业务差不多。后来，钱庄就发现，商人们并不要求取回当初存入的那些钱，只要是同等价值的钱就可以了。于是，钱庄就挪用储户的存款去放贷。但是这样一来，储户就要承担额外的风险。最后钱庄不得不向储户支付利息，才能招揽到存款，进而拿这些存款去放贷。这种业务模式，本质上是储户委托银行放贷，是委托理财的一种，存贷息差就是佣金。

存贷款利率曾经是完全市场化的。但是，金钱的逐利本能如果不受约束，会导致很多社会问题。所以，后来各国都加强了金融监管，由央行决定基准利率，其他银行的存贷款利率参照基准利率浮动，不得超出一定的范围。

7.3 风险、投资组合与基金

7.3.1 风险

风险是指未来的不确定性，尤其是蒙受损失的可能性。

有些人不愿冒风险，因为世上很多事是不可逆的。比如，人突发急病，却提不出钱来治病；人死之后，钱提出来了，也没用啦。企业破产、经理人降职，都是这样，要发生很容易，要恢复就很难。所以，很多人都力保平稳不出事，尽量回避风险。

有些人喜欢冒风险，因为利润高的事情，通常风险也高。利润又高、风险又低的事情一般都是稍纵即逝，很难抓住。其实，这种“很难抓住”本身就是一种风险，万一看走了眼，以为是，一抓却不是，那损失就大了。所以，追求高利润，就是要拼眼光、拼才智、拼胆量、拼勤奋，把别人想不到、不敢做、做不好的事情做好，就是要把别人眼中的千风万险一一化解。这可不是狂妄、吹牛的人可以做到的，得有真本事才行。

典型的风险有：

- 流动性风险。就是需要用钱的时候，提不出钱来怎么办。一定要提，只能降价抛售，蒙受损失。
- 通胀风险。就是货币政策的变化。
- 经济景气风险。国际国内经济形势的变化。比如全球金融危机对中国经济的影响。
- 市场风险。股市、债市、期市、房市、金市、汇市、藏市、农市等不同市场各自的波动和相互影响，比如，相互竞争资金。
- 行业风险。政府产业政策调整，或者特定产业的突发事态。比如，禽畜疫病爆发对养殖业的影响。

- 公司风险。具体公司经营环境和状况的变化。
- 代理风险。委托别人代理某项事务时，代理人不能胜任或者另谋它利而使委托人利益受损。比如，证券公司挪用客户资金余额等。
- 主力控盘风险。一只股票如果被主力控盘，那么它的价格可以背离市场，乖戾无常。有些短线炒家喜欢跟着这些主力，希望能够“主力吃肉，自己喝汤”。结果往往是，一不小心掉进汤里，成了主力碗里的肉。
- 初创风险。企业在初创时期所面临的一些特殊风险。比如，管理团队缺乏经验，队伍不稳定等。

7.3.2 机会成本与效用曲线族

任何决策都是在多项方案之间进行选择。如果风险相当，那么收益低的被排除；如果收益相当，那么风险高的被排除。最后留下的，是一组低风险低收益、高风险高收益，彼此不相上下的方案。

所谓“不相上下”，就是把收益按相应的风险打折扣，风险大的折扣大，风险小的折扣小，无风险的不折扣，如此调整之后，这些方案的“无风险收益”基本相等。这个“无风险收益”，常被称为机会成本。

在平面直角坐标系上，以横轴表示风险，纵轴表示收益，把各种投资方案按其风险和收益的大小标在相应的点上，就做成了收益-风险图。

给定机会成本，在收益-风险图上，彼此不相上下的方案排成一条向右上方加速翘起的曲线，称为效用曲线。对各种不同的机会成本，有一组基本上平行的效用曲线，称为效用曲线族。

机会成本在数值上最接近于风险最低的那个方案的收益。所以，实务上常把购买国债等低风险方案的收益视为机会成本。

7.3.3 投资组合

通常，政策或事态的某种变化对不同的投资品种有不同的影响。比如，央行加息，新发行的债券的利率也随之升高，就会有人抛掉股票来买债券，造成债市升温、股市降温。再比如，发改委提高成品油价格，增加了众多用油企业的生产成本，却扩大了采油炼油企业的利润空间。

为了使资产的价值不随政策或事态的变化而大幅波动，可以在不利品种和受益品种两方面都做适当投资，使得不利品种的减值和受益品种的升值基本相抵。这就是适当的投资组合可以降低风险的道理。

可是，如果全都投资受益品种，获得大幅升值岂不是更好？为什么要满足于“不大幅波动”呢？

因为政策或事态的变化究竟怎样发展还不知道，比如，央行是加息还是降息，还是维持不变，事前没人知道。如果普遍预期央行降息，那么股市作为受益者，很可能已经做了充分“预热”。一旦央行决定暂不降息，股市就会下跌。受益品种一下子就变成了不利品种。“大幅升值”的预期变成了“大幅减值”的结果。所以，不愿冒风险的人宁可选择“虽然升值缓慢但是平稳可靠”的投资组合。

当然，如果你真能未卜先知，算准了央行一定会如何如何，那你的确可以每一次都实现资产的大幅升值。你甚至可以借钱来操作，运用财务杠杆来获得超额的利润。可惜这个“如果”对绝大多数人，实际上是对所有的人，都不成立。

在“平稳升值”和“大幅升值”之间，还有很宽的中间地带，按不利品种和受益品种的比重不同，可以配置不同的投资组合。也就是说，在投资策略中，除了“破财消灾”，牺牲利润买平稳，和“火中取栗”，买对了大赚，买错了大赔，还可以有“适度积极”的策略：确实感觉某一变化发生的可能性很大，而市场还没有对此做出充分的反映，那么就可以适度加重对受益品种的投资，在能够承受相应损失的前提下，以积极的态度，争取尽可能大的收益。

7.3.4 效率前缘

举个简化的例子。假设只有债券和股票两个投资品种。完全投资债券收益不高，波动（即风险）也不见得小。在收益-风险图上，这个方案对应一个位置较低，离纵轴较远的点。由于债市和股市存在竞争资金的关系，在两方都做适当投资，涨跌互补，可以减小组合的波动，所以逐步增加股票比重，方案会向左上方移动，并达到风险最小值。继续增加股票比重，方案会向右上方移动，直至完全投资股票的方案。这并不是该曲线的终点，因为理论上我们可以“借钱炒股”。实践上作者绝不建议这样做。在借钱炒股这一段，风险迅速加大，而收益却越来越微薄，甚至收益不够还利息，所以这段曲线迅速趋平并下弯。

当投资组合包含更多投资品种时，所有可行方案构成收益-风险图上的一个区域，称为可行区域。可行区域的边界称为效率前缘 (efficient frontier)，也称有效边界。

如果投资组合不包含零风险品种，那么效率前缘通常类似于上述债券和股票的例子。

如果投资组合含有零风险品种，那么其中只有收益率最大的那个是有效的，称为有效零风险品种。扣除所有零风险品种之后的效率前缘称为无零效率前缘。从有效零风险品种所对应的点，向无零效率前缘做切线。该切线下方都是可行区域。因此，含有零风险品种的投资组合的效率前缘是一条直线。

“最佳”方案是效用曲线与效率前缘相切时，切点所对应的方案。

7.3.5 基金

投资组合不是一劳永逸的，需要根据形势变化不断调整各个品种的比重，称为维护。比如，人们常说投资股市要做长线，通过组合投资就可获得稳定的高收益。其实不然。任何企业都有成长期、成熟期、衰败期。大多数企业的最终结局是被兼并，或者破产。因此，股市上的投资组合，如果不维护，那么二十年、五十年、一百年后，注定会所剩无几。

教科书上说，股市长线投资收益高，那是隐含地使用了某种维护机制。比如使用股市指数计算收益。综合指数随着新股上市和差股退市不断调整，而成分指数也是定期调整、有进有出的。

有些民众缺乏时间、精力、渠道和技能来亲自维护投资组合。于是，金融机构就针对他们的各种不同的理财需求，设计相应的投资组合，组织专业人员来做维护，然后把投资组合划分成小份卖给这些民众，从中收取服务费。这就是基金。

基金集中了大量资金，可以利用规模优势，从券商处争取佣金折扣，降低交易成本；也可以摊薄研究费用。可是这些好处未必能落到普通民众手中。目前中国的基金收的服务费很高，而业绩却普遍不佳，还没有形成什么让人放心的品牌。

7.4 商品、投机与期货

7.4.1 商品

投资商品牟利的典型方式有：

利用供需双方在交易地点上的矛盾。

某商品在乙地需求强烈，却只在甲地供应充足。需方前往甲地采购，或者供方前往乙地销售，都要增加运费、损耗等成本。如果你有办法降低运费或者损耗，那么你就可以适度降价而不损失利润，你在市场上就有竞争力。

利用供需双方在交易规模上的矛盾。

需方对某商品有大规模需求，而供方却很分散，并且都只能小规模供货。需方逐一前往供方处收购，成本总和会很高。供方前往需方处销售，因为规模小，无法摊薄固定成本，所以，也没有积极性。如果你有办法以较低的成本完成这项“零购整销”的工作，那么供需双方都会非常乐意委托你来办理此事，并付你佣金。这是个三方共赢的局面。

如果把矛盾双方交换位置，供方对某商品有大规模供应，而需方却很分散，并且都只有小规模需求，那么同样存在交易成本高的问题。如果你有办法实现低成本的“整购零销”，那么也可以实现三赢。

利用供需双方在交易时间上的矛盾。

比如，农产品的收获是季节性的，但是消费需求却是持续不断的。解决这一矛盾的各种措施都有成本，如果你的措施成本更低，那么你就有竞争力。

利用供需双方在性价比上的矛盾。

类似品质的商品，如果有多家供方独立竞争，那么任何一家单独提价都会失去竞争力，进而失去市场。要领先其他竞争者，抢到更大的市场份额，主要有两个办法：一个是提高品质，一个是降低成本。

提高品质，提高品牌美誉度，需要改善管理，也需要加大投入，而提价空间又十分有限，所以，更常采用的办法是研发新品。采用新技术，提高性价比。这样，厂家可以定一个相对较高的价位，而消费者则获得更多的产品性能的提高。

当然，由于消费者对新产品新功能的认知不一定完全，不一定准确；负责监管的政府职能部门或者不称职，或者被收买而加以包庇，或者干脆就是一套人马两块牌子，本来就是一家；而媒体又不独立，什么可以说，什么不可以说，完全操纵在王八蛋的手中；这就给了无良厂商以可乘之机。它们并不把心思放在切实提高产品性能上，而是对产品略加改造后，就寻找和炮制各种噱头，夸大宣传，把稻草吹成金条，欺骗消费者，目的只是为了涨价，牟取暴利。比如臭名昭著的 ××× 推广计划，以及其他许多许多臭名昭著的计划等。

新品上市，没有竞争，厂家可以在一段时期内获得较高的利润。随着竞争者的加入，价格逐渐降低。虽然成本也可以降低，但总的来说，利润会越来越薄。研发能力强的厂家可以在适当的时候，退出价格战，再推新品上市，开始一个新的周期。

降低成本，也需要改善管理，采用新技术等。但是还有一个更常采用的办法，就是利用规模经济摊薄固定成本。

综合以上所述，不难发现，无论哪种牟利方式，最主要的因素是创新：改陆运为水运；降温保鲜减少路途损耗；组织农社合伙销售；建立网站合伙采购；改良品种和储藏技术；研究市场需求，开发新产品；等等，都是创新。引领社会发展方向，降低社会成本，推动社会进步，改善民众生活，是财富的真正源泉。

7.4.2 投机

理财上所谓的投机，是指投资于价格波动的机会。

价格围绕价值波动是个一般规律。当价格过度低于价值时，趁价格低的机会买进该资产，这就是投资于这个低价买进的机会。待价格回升到价值附近时，卖出获利。当价格过度高于价值时，借入资产，趁价格高的机会卖出，这是投资于高价卖出的机会。待价格回落到价值附近时，买进资产归还，也可获利。其中的难点在于，所谓的价值，究竟该如何评估，如何判断每天的经济事件对资产价值的影响。

市场上的适度投机可以活跃市场，及时发现价格。投机者在资产被低估时买进，被高估时卖出，避免了过度波动，稳定了市场。因此，这种情况下投机获利是完全正当的。

但是，在过度投机情况下，价格一旦下跌，投机者便纷纷抛售，落井下石，加剧价格跳水；而价格一旦上涨，投机者便纷纷追涨，火上浇油，加剧价格飞涨。这种情况下，真实的经济因素变化对价格的影响，被投机者相互之间的心理博弈、零和游戏所淹没。市场的价格发现、资源配置等功能无法正常发挥。受害的是广大中小投资者，获利的是那些凭借资金优势和其他特殊关系，操纵市场，操纵传媒的王八蛋。

典型的投机牟利方式是短线炒股和期货。

7.4.3 期货

期货在概念上，是在未来某时点，对某商品的交易权，在操作上，是标准化合约。合约上规定了商品种类、数量和质量标准、交货时点以及其他交易条件，但是没有规定价格。价格由期货市场交易产生。

在期货市场，卖出了一定数量的合约的，可以择机再买进等量的合约；买进了一定数量的合约的，可以择机再卖出等量的合约；称为对冲。没有完全对冲掉的合约称为头寸。合约到期，仍然持有卖出头寸的，要承担提供相应货物的责任；仍然持有买进头寸的，要承担接收相应货物的责任。这称为实物交割。

举例来说，假设某农村种粮合作社打算种植大豆，预计产量五百吨，按照相应的大豆期货的当前价位核算，应当是有合理利润的。于是，该合作社可以按此价位卖出五百吨的合约，锁定合理利润。到收获季节，实物交割时，合作社负责提供货物，期货交易所按照交易规则确定接收货物的买家，合作社的销售价格就是合约上的价格。这样，合作社的合理利润就得到了保障。

当然，要保障，就得放弃“横财”。如果实物交割时，现货大豆价格高于合约上的价格，合作社只能满足于合理利润，放弃“意外波动”产生的额外利润。

假设某豆油厂预计在大豆收获季节需要收购五百吨，但是相应的大豆期货合约的当前价位太高，所以暂时不买，等待时机。一个月后该合约价格滑落到一定程度，该厂核算后认为有合理利润空间，于是按此价位买入五百吨的合约，以锁定成本，锁定合理利润。到实物交割时，期货交易所按照交易规则确定提供货物的卖家，而该豆油厂的采购价格就是合约上的价格。这样，该厂的合理利润也得到了保障。

在现货交易中，成交价既是卖价也是买价，而在期货交易中，实物交割的双方按照各自的合约价位实现销售和采购，中间的价差盈亏由投机者承担。

所谓的“投机者”，是指那些不以销售或采购商品为目的，只在价格波动中谋求价差利益的投资者。他们在合约到期之前，无论是赔是赚，一定要对冲完毕，以规避实物交割的责任。相对于投机者，那些参与实物交割的投资者称为“避险者”，也称套期保值者。

避险者追求的是规避风险，投机者追求的是价差利益。上例中，合作社在决定种植大豆的时候就需要锁定卖价，这时的卖价不一定高，只是相对于种植成本，有合理利润而已。豆油厂为了安排这五百吨大豆的加工生产，也需要提前计划，这时的买价不一定低，只是相对于豆油的价格而言，有合理利润而已。这样一来，合作社的卖价有可能低于豆油厂的买价。两者之间的价差利益就由投机者获得。这是对投机者承担价格波动风险的回报，所以也是公平的，正当的。

投机者对每天的经济事件进行分析，根据形势的变化在市场上卖出或者买进合约，使得合约价格及时、恰当地反映形势的变化。这是适度投机的好处。对于市场的参与者而言，判断正确就盈利，判断错误就亏损，资金向真正睿智的投资者手中汇聚，以此实现资源配置的优化。

如果有主力资金控盘，操纵市场，大批中小投机者跟风，追涨杀跌，就会使得合约价格背离形势变化，乖戾无常。这就是过度投机。这种情况下，只有王八蛋才能获利。

7.5 股市投资

7.5.1 成长性与市盈率

如果一家公司的权益收益率逐年增长，就说这家公司有成长性。增长率越高，成长性越好。

设一只股票期初的价格是 P，持有一年期间的每股收益为 E，那么 P/E 就称为市价/盈利比率，简称市盈率。

假设某公司股利政策为全部现金分红，预期本年度的每股收益为 E，每年

增长率都是 g，取折现率为 r，那么该个股的合理价位 P 就是

$$
\begin{aligned}
P &= \frac{E}{1+r} + \frac{(1+g)E}{(1+r)^2} + \frac{(1+g)^2E}{(1+r)^3} + \cdots \\
&= \frac{E}{1+r} \times \left(1 + \frac{1+g}{1+r} + \left(\frac{1+g}{1+r}\right)^2 + \cdots\right) \\
&= \frac{E}{1+r} \times \frac{1}{1-\dfrac{1+g}{1+r}} \\
&= \frac{E}{r-g}
\end{aligned}
$$

取 $r = 0.1,\quad g = 0.02$，那么 $P = 12.5E$，合理市盈率为 12.5 倍。而不增长的公司，即 $g = 0$，有 $P = 10E$，合理市盈率为 10 倍。

假设持有 m 年后以价格 P_m 卖出，那么，我们有

$$
\begin{aligned}
P &= \frac{E}{1+r} + \cdots + \frac{(1+g)^{m-1}E}{(1+r)^m} + \frac{(1+g)^m}{(1+r)^m}\left(\frac{E}{1+r} + \frac{(1+g)E}{(1+r)^2} + \cdots\right) \\
&= \frac{E}{1+r} + \cdots + \frac{(1+g)^{m-1}E}{(1+r)^m} + \frac{(1+g)^mP}{(1+r)^m}
\end{aligned}
$$

注意到，第 m 年预期期末分红收益为 $E_m = (1+g)^mE$，而第 m 年期初合理价格为 $P_m = (1+g)^mP$，因此，增长率固定为 g 的公司，其每股收益按几何级数逐年增长，股价亦然，而市盈率 $\dfrac{P_m}{E_m} = \dfrac{P}{E} = \dfrac{1}{r-g}$ 则保持不变。

虽然现实中，企业的增长不可能如此“标准”，但是，这个例子还是定性地说明了，市盈率是对公司未来长期发展前景的量度。预期每股收益 E 是对公司当年度预期经营成果的量度。在预期本年度每股收益 E 相差不大的情况下，未来发展前景广阔、成长性好的公司，股价可以适当高些。

7.5.2 市场收益率与量价配合

一般把综合指数的收益率称为整个股市的市场收益率，记为 r_m。

对于增长率为 g 的个股，期初投入 P，当期的分红收益为 E，在期末，价格上涨到 $(1+g)P$，当期的增值收益为 gP，所以，该项投资当期的总的收益率为 $(E+gP)/P = E/P + g = r$。注意，它与公司的成长性无关。

在上一节的例子中，以 12.5 元的价位买成长的公司，和以 10 元的价位买不成长的公司，投资收益率都是 $r = 10\%$。也就是说，如果暂不考虑风险，那么在有效的股市中，股价已经包含了成长性等各种因素的影响，最终使得买任何公司的股票都得到相差不大的投资收益率，即市场收益率。

注意，公司股东的权益收益率，与二级市场上投资的投资收益率是不同的。前者是 E，是假设以每股一元的名义价格买进而获得的收益率。后者是 r，是假设以市场价 P 买进而获得的收益率。一家公司盈利能力强，E 比较

高，但是如果市场已经对此做了充分反映，P 已经提到了相应的高位，那么 r 就不会高于市场收益率了。

据说有种选股方法，就是把报纸的行情版挂在墙上，然后用飞镖去打，打到哪个是哪个。本文作者绝不推荐这种做法。因为不同的投资者对各种风险的熟悉程度和应对能力不同，不同的个股对各种风险的响应也不同，所以，投资者应该挑选合适的个股进行投资，以争取更好的业绩。

实际上，如果暂不考虑风险，那么，随着各种经济因素不断变化，上市公司的收益前景也不断地受到影响。投资者会抛售前景变差的个股，称为做空这些个股；而竞购前景变好的个股，称为做多这些个股。

关于事态变化如何影响公司前景，不同的投资者会有不同的看法。所以一般总是有人做空，有人做多。做空的一方称为空方，也称空头；做多的一方称为多方，也称多头。多空双方在同一价位上遭遇，就会达成交易，产生成交量。

成交量缩放与股价涨跌的搭配组合，称为量价配合。比如，放量涨、放量跌、无量涨、无量跌、放量胶着、无量盘整等。

如果市场对事态后果看法一致，都看空或者都看多，那么行情就会无量涨跌。涨跌到一定程度，原本占优势的一方开始减弱，而另一方则开始增强。于是，市场在这个价位上看法开始出现分歧。多空双方展开拉锯战，形成放量胶着的行情，也称在这个价位上遇到“阻击”。如果另一方暂时接济不上，行情继续原来的趋势，则称为“突破阻击”。如果行情在此价位上逐步缩量盘整，则称阻击成功。这种缩量盘整，说明市场形成了初步共识，认为该价位与个股的预期收益 E 和成长性 g 相匹配，而且换股也不能进一步提高收益率了。其他量价配合的含义不再赘述。

正是因为投资任何个股所获得的收益率都相同，才使得前景变好的上市公司的股价升高，市值增大，更容易获得贷款，也更容易增发再融资。所以说，股市通过发现价格，来引导社会资金向前景好的上市公司流动，优化实体经济的资源配置，提高资源的利用效率。注意，“资金向前景好的上市公司流动”是指贷款、首发和增发，而不是指股民在二级市场上买该公司的股票。

7.5.3 市场风险与 β 系数

现实的市场一定有波动。一段时期内的综合指数的相对均方差，称为这一时期的市场风险，记为 $s_m = \sigma_m/P_m$。

一段时期内的个股 A 的价格的相对均方差，称为这一时期内该个股的风险，记为 $s_A = \sigma_A/P_A$。平衡状态下，投资任何个股都会得到市场收益率，不同的只是风险大小，即不平衡时的波动的大小。

绝大多数个股在绝大多数情况下会随着综合指数的波动而发生类似的波动。也就是说，如果一段时期内综合指数相对稳定，那么个股价格一般都比较

稳定；如果一段时期内综合指数大起大落，那么个股价格一般都大起大落。这一关系通常被表述为

$$s_A = \alpha_A + \beta_A s_m.$$

其中 α_A 的长期平均值一般很小，在分析平衡市场时，可忽略不计；β_A 称为个股 A 的 β 系数，主要与该公司的总股本、股权结构、所在行业、企业性质、以及筹码集中度等因素有关，通常在一定时期内相对稳定。

所谓“船大经风浪”，一般来说，股本庞大、筹码分散、业务稳定的公司，股价日常波动较小。这样的个股称为蓝筹股。选三十种分属不同行业的蓝筹股做投资组合，足以逼近效率前缘，获得与市场风险相适应的市场收益率。这就是道琼斯（三十种工业）指数的意义。其他成分指数的意义类似。

7.5.4 综合指数方案

综合指数不仅仅是一个点数。它也是一种投资组合方案。许多实证研究表明，在有效的股票市场，综合指数方案通常就在效率前缘上，其具体位置和漂移路径主要由机构投资者的交易特性所决定。

机构投资者不会在股市上长期满仓操作。不满仓时抽出来的资金主要投向债券市场。最终的效果就相当于，绝大多数机构投资者有收益率基本相同的零风险品种。从该品种向股市效率前缘做切线，假设综指方案在切线上方，那么机构投资者会做多股市，导致综指方案下移；假设综指方案在切线下方，那么机构投资者会做空股市，导致综指方案上移。这就是说，在机构投资者的推动下，综指方案倾向于稳定到切点上。

但是，综指方案稳定并不意味着机构投资者满仓。不同的机构投资者有不同的经营目标和投资策略，尤其是有不同的效用曲线族。机构投资者的效率前缘是直线。厌恶风险的机构投资者的效用曲线在风险较低时就开始上翘，所以，它与效率前缘的切点就靠近零风险品种，即持有债券多、股票少。而喜好风险的机构投资者的效用曲线在风险较大时才开始上翘，所以，上述切点靠近综指方案，即持有债券少、股票多。这样，有不同风险偏好的机构投资者在效率前缘上都能找到自己的平衡位置，按照低风险低收益、高风险高收益的规律分布。

显见，机构投资者的多样性和相互制衡，对股市稳定非常重要。如果机构投资者都“统一思想，统一认识”，或者某种特殊类型的投资者“一权独大”，足以操纵市场而不受制衡，那么，股市注定（由于众所周知的原因，此处省略若干字。）

预测综合指数的走势，通常要考虑四个方面：基本面、消息面、资金面和技术面。基本面是指宏观经济前景。消息面是指近期发生的和将要发生的重要事件。资金面是指货币政策、通胀预期、股市与债市房市竞争资金的情况等。技术面是指对常用的一些技术指标的分析解读。

顺便说一下，预测个股走势，通常也要考虑四个方面：大势、公司基本面、公司消息面和技术面。大势就是综合指数的走势。公司基本面是指该公司的运营状况和发展前景。公司消息面是指近期发生的和将要发生的与该公司有关的重大事件。

7.5.5 方案漂移

当各种经济因素变化时，个股，进而投资组合，进而效率前缘，都会在收益-风险图上发生或大或小的漂移。要对漂移路线进行精确全面的分析是不现实的。在此我们只能粗略地、定性地说明一下。

给定个股或投资组合，如果有利好消息出台，未来可预期收益增加，那么，当前价格就是低价，以当前价格买进，就能获得较高收益。因此，该方案在利好消息刺激下，向左上方“跳跃”了一定距离。随着投资者纷纷做多，其价格上涨，投资收益率下降，风险增加，方案向右下方移动，最终会回到效率前缘附近。反之，如果有利空消息出台，那么，当前价格就是高价，以当前价格买进，只能获得较低收益。因此，该方案在利空消息刺激下，向右下方“跳跃”了一定距离。随着投资者纷纷做空，其价格下跌，投资收益率上升，风险减小，方案向左上方移动，最终也会回到效率前缘附近。

简言之，在正常的股市中，客观经济事态的变化造成各种方案从效率前缘上“跳开”，而投资者做空或做多的交易行为，把各种方案再推回到效率前缘上。

效率前缘的长期稳定位置，决定于国家的长期经济发展形势。当然，这种稳定是指相对的、动态的平衡，绝不是指固定不动。

如果所出台的消息对整个股市利好，那么整个效率前缘从长期稳定位置向左上方跳跃一定距离。随着投资者做多，指数上涨，收益率下降，风险增加，整个效率前缘向右下方移动，最终稳定在长期位置上。反之，如果对整个股市利空，那么整个效率前缘向右下方跳跃后再基本恢复，最终还是稳定在长期位置附近。

7.5.6 风险报酬率与央行加息降息

在投资领域，低风险低收益，高风险高收益是个一般规律。收益-风险图上，所提高的收益率除以所增加的风险，称为风险报酬率。

央行加息一方面意味着公司的资金成本增加，盈利能力下降，对股市是个利空；另一方面提高了机构投资者零风险品种的收益率，两者都导致风险报酬率降低。股市吸引力降低，机构投资者减持股票。在收益-风险图上表现为，机构投资者的效用曲线与效率前缘的切点向左移动。

反之，央行降息导致风险报酬率升高，机构投资者会增持股票。

7.5.7 挑选合适的基金

目前，中国的公众投资者对风险与收益率的关系认识不充分，在选择基金的时候片面追求收益率。这其实也不是公众的错。许多基金在销售的时候就片面鼓吹过去一段时期内的高收益率，对风险却避而不谈，或含糊其辞、一笔带过。结果是，公众抱着很高的期望买了，然后期望落空，然后就骂这些基金是骗子。这是可悲的双输。

有人说，政府加强监管，要求基金充分披露投资原则和策略，并监督其忠实执行，可以给公众投资者一些信心。但是，“商场如战场”，“兵者，诡道也”，“将在外，君令有所不受”，公众是君，基金是将，授权并承担必要的代理风险是不可避免的事。要说区别，那就是战国时期大将出征，要把家眷抵押给国君，打赢了回来加官进爵，打输了连家眷一起处斩。而基金如果打赢了，基金公司和基金经理可以大块吃肉，公众投资者也能跟着喝汤；如果打输了，基金公司和基金经理仍旧可以喝汤，公众投资者却只能割肉离场。这种局面与中国社会状况非常相似。“硕鼠硕鼠，无食我粟”，被鱼肉的草民除了考雅思，办护照，“誓将去汝，适彼乐土”之外，别无他法。

解决这个“将大欺君”的问题，需要引入竞争机制，给公众更多选择的自由和言论的自由。做误导性宣传、给投资者造成损失的基金，要受到媒体的揭露，要被公众抛弃，基金经理要失去职业前途，基金公司要破产。这样“有赏有罚，赏罚得当”，公众投资者的权益才能得到公平的保障。社会问题也是如此。（由于众所周知的原因，此处省略若干字。）

基金在销售宣传和业绩报告中应披露哪些信息，如何披露，这个问题我们不在这里谈。笔者另外有一个建议，就是希望研究机构、财经传媒、或者交易所、基金协会之类有条件有能力的组织，能够定期统计并发布机构投资者的效率前缘。这样，公众投资者就可以依据同一时期，基金在收益-风险图上的位置，来判断基金的业绩。如果基金在效率前缘上方，那就是超出一般水平；如果在下方，就是低于一般水平。

强调一下，在过去一段时期内有效的投资策略，在未来不一定仍旧有效。所以，投资者挑选基金，不仅要了解基金本身，还要了解经济形势和市场的变化，要对基金的投资理念是否能适应未来的发展变化做出独立的判断。

7.5.8 股指期货

多数机构投资者资金庞大，在短时间内集中增持或减持某只个股，会对其价格造成过度冲击，因此只能把操作时间拉长，慢慢地“吸筹”，慢慢地“出货”。这意味着大部分交易是以平衡价格进行的，如果始终全仓操作，那么这部分资金的收益率很难高过市场收益率。因此，机构投资者就逐渐分化为两类，一类专注于公司风险和行业风险，力图规避市场风险；另一类则专注于市

场风险，希望通过准确预测行情变化，从股市的整体波动中牟取利益。于是，股指期货便应运而生。

股指期货是买卖未来某时点某个股指的点数。按照每个点数折算一定金额来计算交易额。合约到期不需要实物交割。交易所按最后一段时间内的交易情况确定结算点数。卖出头寸点数高于结算点数的，获得差额，低于结算点数的，补缴差额。买入头寸点数高于结算点数的，补缴差额，低于结算点数的，获得差额。

利用股指期货，股市投资组合持有者可以卖出适当数量的合约，使得投资组合随股市整体波动的增值或减值，能够大体上被股指期货合约的盈亏所对冲，从而实现规避市场风险，套期保值的目标。

股市卖空者通过买进适当数量的合约来套期保值。所谓股市卖空者，就是趁股价高位的时机，借入股票卖出，以期在股价回落之后，再买进股票归还。这其实也是一种投资组合。不同之处在于：股市下跌时，该投资组合增值；股市上涨时，该投资组合减值。

股市卖空者如果持有适量的股指期货买进头寸，那么，无论股市如何波动，买进头寸的盈亏能够基本上与投资组合的增值减值对冲，从而保持了资产总价值的稳定。

重提那个老问题：多盈利不好吗？为什么要对冲？都对冲完了，不就成了竹篮打水一场空，白忙活了吗？

回答是：在这个不确定的世界上，多盈利的另一面就是多亏损。大赚虽好，可惜大赔赔不起啊。对冲，只是对冲掉了市场风险，并没有对冲掉上市公司的实质盈利。多数机构投资者的理财目标，就是长期地、在可接受的风险水平上，获得这种实质盈利。

股指期货市场也需要适度投机，保持交投活跃，以便及时、准确地发现价格。投机者发现合约点位过高，就会卖出合约；发现合约点位过低，就会买进合约。这在客观上起到了稳定合约点位的作用。

不过，投机牟利风险很大。机构投资者相互之间竞争激烈，各种消息真真假假、虚虚实实；同一消息究竟是利好还是利空，各种理论众说纷纭、莫衷一是；市场对某一消息究竟是已经做了充分反映，还是没有充分反映，还是已经反映过度，也不容易判断。任何失误都可能招致巨大损失。

7.5.9 做市商与对冲基金

机构投资者中也有专门利用个股的短期波动牟利的，比如做市商。如果市场交投清淡，有人想以合理价格买却没人卖，或者有人想以合理价格卖却没人买，时间长了，整个市场就会废掉。所谓做市，就是在市场人气低落，不足以支撑正常交易的时候，出面接盘，维持市场必要的活跃度。

做市之所以有利可图，首先是因为做市商一般是证券公司的自营部门，佣金全免，交易成本很低。其次是因为市场气氛冷清，有买家无卖家的时候，买家会把价格推到合理区间的高端；有卖家无买家的时候，卖家会把价格推到合理区间的低端。做市商利用资金雄厚、交易成本低廉的优势，耐心等待时机，在合理价格区间的低端买，在高端卖，赚取价差利益。注意，不要把做市与主力控盘混为一谈。

另一种炒短线的机构投资者称为对冲基金，这里的“对冲”指对冲头寸，即：做多买进的，一定要择机卖掉；做空卖出的，一定要择机买回。对冲基金是走技术分析路线的专业团队，一般由金融工程方面的行家主持，配置有高速数据通道与交易所相连，用高性能计算机和特别设计的软件系统，自动分析数据并自动下单。

程序下单在提高市场有效性的同时，也埋藏了巨大的风险。对冲基金没有能力也没有义务在经济本身出现危机的情况下，挺身而出，拯救股市。逃跑避险是投资者的本能和权利。问题是，对冲基金的能量是如此巨大，神经却又如此脆弱，以至于有些时候，股市原本可以依靠自身的调节机制避免股灾的，却由于对冲基金的落井下石而崩溃；甚至于，在经济没有实质性异常的情况下，对冲基金也会对偶发的稍大的随机波动做出过度反应，造成无缘无故的暴涨暴跌，干扰市场正常功能的发挥。后来，有些交易所规定，在行情波动超过一定幅度时，就自动屏蔽程序下单。但是，把程序下单伪装成普通客户下单以穿透屏蔽，在技术上并无难度。所以，要么个股停牌，大家都不交易；要么就得忍受对冲基金这样的“超人汉考克 (Hancock)”式的坏脾气。

对冲基金的“脆弱”来源于多个方面。首先，基金经理知识结构相似，投资理念趋同，平抑风险的手法一致。这样就容易发生“共振”，采取不约而同的行动，威胁市场的稳定。其次，技术分析根据历史预测未来，容易“水涨船高”，“久入芝兰之室而不闻其香”，把长期泡沫当成实体经济的真实水平，失去了对风险的警惕性。这也是这次全球金融危机的重要原因之一。再次，对冲基金的投机性质决定了他们喜好风险，比其他类型的机构投资者更乐于使用价外期权 (out-of-money options) 等金融工具把“正常范围”内的收益率放大，而不顾“异常情况”下的损失也被放大的事实。一旦“异常情况”发生，局面往往无法收拾。

程序下单作为技术手段无可厚非。对冲基金也是合法的投资者，封禁是不可取的。要“扬其长、避其短”还得走多元化、相互制衡的路子。多元化（由于众所周知的原因，此处省略若干字。）

7.5.10 行业平均市盈率水平

人们常说朝阳产业、夕阳产业，就是指行业平均增长率的高低。上市公司的增长率 g_A 由该公司所属行业的增长率 g_{Ai} 和该公司超出行业增长率之外的部

分 g_{Ae} 组成，即 $g_A = g_{Ai} + g_{Ae}$。

如果市场竞争充分，那么，过高的 g_{Ae} 是很难长期保持的。其他竞争企业会赶上来，争夺市场份额。过低的负的 g_{Ae} 也不会长期持续，或者该企业管理层励精图治，赶上行业平均水平；或者该企业股东不满意，更换管理层；或者该企业被重组、被收购，更换股东，总之，在市场机制有效发挥作用的情况下，资源不会被长期压制在低效率状态。因此，就长期趋势而言，企业增长率会向行业平均增长率回归。

这也就意味着，在市场机制有效发挥作用的情况下，企业市盈率会向行业平均市盈率回归。如果个股的市盈率过度偏离行业平均市盈率水平，那么，投资者应该提高警惕，仔细审查可能的原因。如果是市场有些情绪化，或者是该个股已经被主力控盘，那么，被高估的就该卖掉，被低估的就可以考虑买进。

当然，中国是个例外。中国的市场是个“有特色”的市场，所谓的“竞争”与众不同。（由于众所周知的原因，此处省略若干字。）

7.5.11 适应性和独创性

草原上既有兔子也有狼。兔子打不过狼。但是草原上濒临灭绝的却不是兔子，而是狼。兔子有兔子的活法，不是非要和狼去打斗不可。专业上讲，不同的生物都有自己的生态位。只要一个物种的结构、体型、食性等各方面特点与它的生态位相适应，那么，这个物种就能够在生存竞争中延续下去。当然，这种延续是整体的、动态的。个体的兔子总有不幸者要被狼吃掉。而且长期来看，总有物种要灭绝，也总有新物种要分化、发展出来。

用生态系统来类比市场，很容易被误解，好象人可以在经济的名义下，自甘堕落为禽兽似的。市场竞争的主体是人或法人，而人和法人都要受社会规范的约束。“在商言商”，言商人的利润的同时，也要言商人的社会责任。“商场如战场”，战场也有道德底线，也不能丧心病狂。市场需要竞争，需要激励，但是，竞争不能“不择手段”，激励不是“唯利是图”。人类社会，哪怕只是在经济方面，也不能简单地套用丛林法则。经济不是堕落的理由。要赚钱，一定要想明白赚钱的道理，不能在利益诱惑面前迷失自我。

如果把股市与草原生态环境做类比，那么，上市公司权益收益是草。做长线的投资者是食草动物。他们根据自己的口味，挑不同类型的草来吃。做短线的投资者是食肉动物。他们的盈利中有很大一部分是其他投资者的亏损。到处寻找短线主力，试图跟着跑、分一杯羹的投资者，是食腐动物。

不能说食肉动物可恨，食腐动物可憎。他们既然能够长期存在，就一定有某种理由。要点在于，任何理由都要讲究适度和平衡。生物界的丛林法则是自然适度的，食肉动物一般不会过度捕猎。而当人类要“简单地套用丛林法则”的时候，通常只是为自己的贪婪无度寻找借口，所以必然会导致失衡，害人害己。

要想在股市上立足，就得知彼知己。首先要知彼，了解别人有哪些投资策略，为什么他们能够生存，其中的合理性是什么。其次要知己，清醒地认识自己，拥有什么，不拥有什么，能做到什么，不能做到什么。然后，选择适合自己的方案试一试，积累一些经验，再逐步加入自己的独创。

市场竞争中，有“一着鲜，吃遍天”的说法。股市上，要想获得高于市场一般水平的业绩，也得有“一着鲜”。一种投资方法，无论一开始多么行之有效，一旦被公开，被广为采纳，就会迅速失去“抢在别人之前”的意义，变得无效了。由此甚至发展出一种“逆反投资策略”：如果散户大厅里面人满为患，新开户数又创新高，就抛空离场；反之，如果散户大厅内稀稀拉拉，人气涣散，就逐步建仓。做大盘如此，做个股亦然，也有“人抢我抛，人弃我收”之说。当然，不能把任何话当教条。对公司基本面的考察总是不可少的。

草原上既有群居的动物，比如羊和狼，也有独居的动物，比如兔和鹰。股市上，到处扎堆，打听小道消息，天天把炒股挂嘴上的股民，是“群居的”；自己坐家里看电脑，编程序的股民，是“独居的”。群居的股民更容易从众，独居的股民比较有条件做逆反。投资方法要与自己的性格相匹配，扬长避短，才有好的收益。

既然是“独创”，就得靠每个人自己发挥创造力。基本的大道理是一样的：准确判断个股或投资组合在消息刺激下跳跃到什么位置，抢在市场恢复平衡之前下单。具体的判断方法，需要每个人自己独创了。

7.6 实业投资

7.6.1 集团内部投资

要准确判断权益投资的风险与收益，对行业和公司的了解至关重要。这方面优势最强的情况是，成熟的集团公司为了扩大生产规模，或者延伸产业链，而投资设立子公司或者参股相关的企业。投资分厂、分店是扩大生产规模的例子。投资上游企业，为集团核心业务提供原料保障，或者投资下游企业，对核心业务的产品进行深加工，这是延伸产业链的例子。

由于集团公司不仅对本行业非常熟悉，知道如何化解各种行业风险；而且对所投企业的管理层人事任免和财务决策能够施加重大影响，又可以化解公司风险，因此，如果集团公司决策层确有能力的话，那么，这种投资形式是最有希望出业绩的。

美国杜邦公司是个典型的例子。该公司首创的一套用来分析被投资企业权益收益率，寻找提高业绩的关键因素的方法，已经进入经典财务管理教材，被称为杜邦分析法。

上市公司有扩大规模或延伸产业链的新项目，为此而增发股票，从一级半市场募集资金，用于实施该项目，对上市公司来说，正属于上述投资形式。但

是，对股民来说，就多了一层的代理风险。上市公司有没有诚信，是否把募来的钱当回事，是个让人担心的问题。中国股市，尤其如此。

股票表示持股人对该公司资产中的权益部分的相应份额拥有名义上的所有权。之所以说是“名义上的”，是因为这份资产不能独立出来自由处置。股票在表示所有权的同时，也表示这份资产必须委托该公司来运营。所以，买股票实际上是一种授权。股民通过授权，可以利用上市公司管理层的专业能力，来回避行业风险和公司风险，提高资产的孳息能力。

7.6.2 创业和创业投资

如果你自己有足够的专业能力，对行业很熟悉，有能力管理企业，那么，你就可以自己创业，来回避代理风险。虽然做老板，也要向职员授权，但是这种经营中的代理风险，比投资中的代理风险，要小得多，可控得多。

对多数年轻人来讲，专业能力可以通过勤学苦干、多经历、多思考而获得，相对比较容易积累。而资金的积累就相对比较难。一夜暴富不是经济的常态。这样，有专业能力的年轻人要创业，就要说服别人向自己授权，给自己的创业计划投资。

商场如战场。实业投资与股票投资相比，风险更大，收益率更高。如果你只能达到一般的业绩水平，那么投资人没理由非投给你不可。如果你许诺更高的业绩，那么，对投资人来讲，就意味着更大的风险。要达成交易，需要双方都付出努力：创业者要尽力证明自己的才华，投资人要尽力“慧眼识才”，既不放过真正的人才，又不落入自视清高、眼高手低的吹牛皮者的陷阱。

为了降低代理风险，投资人通常会委派更有经验的专业人员，参与企业的经营管理，至少要与创业者一起工作一段时期，这既是一种扶助，也是一种监督。

通常，企业都有从成长，到鼎盛，再到衰败的生命周期。初创时期的企业要面临核心技术的发展前景尚有争议；管理团队缺乏经验，队伍不稳定；关系网络尚未建立，商誉和信用不足；资金支持不稳定等诸多风险，不妨统称为初创风险。因此，初创企业的“死亡率”要远高于已经进入稳定成长期的企业。

前面讲过，企业的成长性越好，市盈率就越高。创业投资者专门为创业者提供初始资金，并扶助其进入稳定成长期，在企业的高成长性得到市场认可之后，趁市盈率尚高的时候，将权益转让，兑现高收益。这等于是发挥自己的聪明才智和资源优势，既帮助创业者度过初创时期的难关，分担了创业者的创业风险，又使得后续的社会投资者能够避免承担企业的初创风险，实现了多方共赢，促进了整个社会的经济发展，因此，获得相应的高收益是完全正当的。

有两个问题要强调。第一，“发挥聪明才智和资源优势”并非易事。当业界还在对一项技术的发展前景争论不休的时候，你敢投资吗？你对创业者的素质、性格、意志力、道德水准了解清楚了吗？对其未来发展前途看得准吗？更

不要说“资源优势”了，你有钱、有队伍、有关系吗？当然，创业投资公司不是个体户，而是一个专家团队，包括负责甄别技术的，负责管理顾问的，负责融资的，以及其他必要的高级人才。把这么多高级人才拢到一块儿共事，本身就很不容易。何况所做的事业还是高风险的，压力很大。所以说，创业投资，绝对不是随便什么人都能干得了，干得好的。

第二，“相应的高收益”具体是多少不由上帝说了算，而是由市场博弈来决定。于是，和所有的市场博弈一样，夸大宣传、过度包装，乃至做假帐，都有可能发生。我们不能因噎废食，不能因此就说创业投资者都是骗子，都是贪婪的、唯利是图的剥削者、寄生虫。

图正当的利是光荣的，是人生成就。图不正当的利才是唯利是图，才是邪恶的。创业投资成为一个独立发展的新兴行业，是投资业不断发展，不断进行专业化分工的必然结果。在初期出现一些鱼龙混杂的乱象是很难完全避免的。随着这个行业的发展壮大，竞争越来越充分，各方越来越成熟，相关的规则、标准会逐渐形成，关于什么是“正当”、“适当”，业界会逐渐达成共识。到那个时候，创业投资行业就会步入健康发展的轨道。

7.7 彩票是一种娱乐消费品

买彩票不是投资，而是一种娱乐消费。看电影是花钱买别人造的梦，买彩票是花钱买自己造的梦，买一种期盼的心情，买几天幻想的权利。

赌博也是一样，是昂贵的娱乐消费。能够盈利的只有开场子抽水的老板，其利润来自于为赌徒们娱乐而提供场地、设施和服务。当然，目前在中国大陆，开赌场是违法的。

普通的赌场，玩一把只是从赢家的获利中抽十分之一的水。而每一期福利彩票，政府都要拿走销售额的一半，而且赢家还要交五分之一的税。从这样的角度讲，彩票比赌博“贵得多”。可为什么还有人买呢？因为花几块钱中几千万的美梦，其娱乐效果比普通的赌博要强烈得多。最终，作为“美梦娱乐”产品，彩票的“性价比”是彩民们可以接受的。

好啦，其他诸如炒房、炒汇、炒金、炒收藏、炒老板、炒员工、炒土豆丝之类的事情，就不多说了。现在回头来看数学。

7.8 投资中的数学模型

7.8.1 均线系统

投资类资产的价格围绕价值波动的情况，可以表示为

$$\text{价格 } p = \text{价值 } v + \text{投机损益 } e.$$

其中投机损益 e 是假设投资者以该资产的价值 v 买入，以该资产的当前价格 p 卖出所产生的亏损或收益。

如果不考虑时间价值，则可假设投机损益 e 服从正态分布。如果有理由认为价值 v 在一定时期内没有显著变化，那么，对上式各项的各期数据取平均值，可得各期价格 p_i 的平均值等于价值 v。

股市上有人把五日均线、十日均线、月线、季线、年线看做对价值的估计，并根据这套均线系统的趋势变化来决定买进或卖出。但是这一类的策略实战业绩并不理想。主要原因在于，均线系统采用的是价格的历史数据，而决定实战业绩的是价格的未来数据。当前和未来的各种偶发事件不一定与历史有关。一个人过去十年没进过医院，不等于明天一定不会进医院。所以，有一种“随机游走”理论说，股价围绕价值做布朗运动。

市场上的价格能否迅速、如实地反映各种经济因素对价值的影响，称为市场的有效性。市场的有效性越高，价格就能够越快越准地调整到新的价值上。由于均线系统采用价格的历史数据，所以有一定的滞后，只有在经济因素本身就是缓慢、持续地发挥作用的情况下，其反映才是较为正确的。

7.8.2 离散线性系统

现实是，市场的有效性不是绝对的、超级的，各种经济因素的变化要传导到价格上，总需要一定的时间，而且有相对固定的模式。

比如，国际原油期货价格上涨，国内炼油企业股价下跌；一段时间后，国内化工原料价格开始上涨，炼油企业股价回升，化纤企业股价下跌；一段时间后，塑料、化纤织物价格开始上涨，化纤企业股价回升。这是上游推动型的变化。

再比如，某项经济政策导致失业率下降，一段时间后居民可支配收入增加，一段时间后消费板块上涨，一段时间后，银行板块上涨。这是下游拉动型的变化。

考虑一个简化的例子。假设价格 P 受两个主要经济因素 X 和 Y 的影响。原始数据是三个时间序列 $P(t), X(t), Y(t)$，解得多元回归方程为

$$P(t) = 0.8P(t-1) + 0.3P(t-2) + 0.5X(t-2) + 0.2Y(t-3).$$

定义状态向量 $(S_i(t))$ 为：

$$
\begin{array}{llll}
S_1(t)=Y(t-2), & S_2(t)=Y(t-1), & S_3(t)=Y(t), & S_4(t)=X(t-1),\\
S_5(t)=X(t), & S_6(t)=P(t-1), & S_7(t)=P(t).
\end{array}
$$

上述回归方程就转化为非齐次离散线性系统：

$$
\begin{cases}
S_1(t) = S_2(t-1)\\
S_2(t) = S_3(t-1)\\
S_3(t) = Y(t)\\
S_4(t) = S_5(t-1)\\
S_5(t) = X(t)\\
S_6(t) = S_7(t-1)\\
S_7(t) = 0.8S_7(t-1)+0.3S_6(t-1)+0.5S_4(t-1)+0.2S_1(t-1)
\end{cases}.
$$

理论上，给定 $P(0), P(1)$ 和 $X(t), Y(t)$，可以逐步推出 $P(t)$。许多机构常把数千个经济因素的数年乃至数十年的数据放在一起考虑，计算量非常巨大。实际发掘出来的系统不一定是常系数的，也不一定是线性的。

从这样的角度看，市场上各式各样的系统时而形成，时而瓦解；一个系统形成之后，可能促进也可能抑制另一个系统的形成或瓦解。喜欢技术分析的股民，寻找一些简单的经验系统以预测股价走势。而经济学家们则试图对稳定存在的系统，以及系统之间的关系，做出合乎逻辑的解释。

有人说，数学是人与自己的博弈，物理是人与上帝的博弈，经济是人与别人的博弈。存在牟利机会的系统被发现，常常意味着该系统将瓦解，或者被其他系统所平衡。牟利行为是一种反馈，可能是负反馈，也可能是正反馈。在经济的世界，能够自我增强的正反馈系统还是很多的，但一般都有某种机制来限制其最大偏离范围。经济危机和金融危机是失去了控制的正反馈系统。这通常被解释为经济或金融体系存在漏洞。

均线系统可以看做是离散线性系统的特例。一般的离散线性系统通过引进其他经济因素的数据，部分地解决了均线系统的滞后问题。但是，问题在于系统的各项参数，乃至系统本身，是从历史数据中发掘出来的，不一定能够延续到未来。如果想通过追究参数的变化规律，和系统之间的关系，来判断延续到未来的可靠性，那么，问题的复杂度将迅速膨胀到实际不可能解决的程度。所以，一般是参考该系统在经济学上的解释进行判断的。

好啦，其他的以后再说吧。写得不好，只要能对初学投资的人有微小的帮助，我也就开心了。

Chapter 8

微积分与相关的几何

本章内容并非教材。建议初学者在粗读一遍教材之后，再看本章内容，并在阅读中不断与教材相互参照。

本章重视数学思想，而常常忽略技术细节。比如本章大量使用“在适当条件下”这个短语，以避免技术细节淹没主要线索。看官应参照教材，了解“适当条件”的具体内容。有人说，“数学”不能这样写。而笔者认为，“数学教材”不能这样写，但本文不是“教材”。不是所有的文章都必须是教材。

8.1 绪论

8.1.1 函数的“算术”

微积分的英文是 calculus，就是演算、计算的意思。我们小学学算术，$2+3=5$, $2\times 3=6$，初中学代数，$(a+b)(a-b)=a^2-b^2$，高中学函数，$y=ae^{bx}$，基本的研究对象都是数。而微积分相当于函数的“算术”，其基本研究对象是函数。

尽管如今我们多使用计算器来做多位数乘除法，但是，了解手工乘除法的原理仍然是十分必要的。不必追求速度，但必须了解原理。

微积分，作为函数的“算术”，情况类似。对于一般的应用，无论是符号演算还是数值计算，都有成熟的微积分软件可用。但我们仍然有必要准确、全面地理解相关概念和原理。这样才是“人使用软件”，而不是“软件捉弄人”。

函数，比如说 f，有一个定义域，比如说 X，对其中的任何一个元素 $x\in X$，都有一个数 y 与之对应，记为 $y=f(x)$。x 称为函数 f 的自变量，y 称为函数 f 的因变量。如果 X 是多个集合的笛卡尔积的子集，那么，这时的函数称为多元函数。比如，四则运算可以看做是二元函数。

因为函数的因变量也是数，对任何数可以做的运算，对函数的因变量也可以做。以这种方式可以把数的运算扩展成函数的运算。比如，可以把 $a+b$ 扩展成 $f(x)+g(x)$。

我们要学习的，是函数的区别于数的“全新”的性质。

比如函数的复合运算。如果 $y=f(x)$, $z=g(y)$，那么，有新函数 $z=h(x)=g(f(x))=(g\circ f)(x)$，记为 $h=g\circ f$。

再比如可微函数的求导运算和可积函数的求原运算。给定函数 f，如果它可微，则可以求其导函数 $\dfrac{\mathrm{d}}{\mathrm{d}x}f(x)$，常简记为 f'；如果它可积，则可以求其原函数 $\displaystyle\int_a^x f(x)\,\mathrm{d}x$。我们有

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(x)\,\mathrm{d}x=f(x),$$

和所谓的微积分基本定理

$$\int_a^x \frac{\mathrm{d}}{\mathrm{d}x}f(x)\,\mathrm{d}x=f(x)\Big|_a^x=f(x)-f(a).$$

现在我们从阿基米德计算球的体积的方法入手，逐步引入相关的基本概念。

8.1.2 一个曲边三角形的面积

直角坐标系第一象限中，曲线 $y=x^2$ 从原点 $(0,0)$ 到点 (h,h^2) 是一条曲边，从 (h,h^2) 到 $(h,0)$ 是一条直边，从 $(h,0)$ 回到原点是另一条直边。求此曲边三角形的面积 $S(h)=\displaystyle\int_0^h x^2\,\mathrm{d}x$。

把从 $(0,0)$ 到 $(h,0)$ 的线段分成 n 份。两端点和中间的断点从左到右构成一个序列，

$$x_0=0,\quad x_1=\frac{h}{n},\quad x_2=\frac{2h}{n},\quad \ldots,\quad x_i=\frac{ih}{n},\quad \ldots,\quad x_n=h.$$

在 (x_0,x_1) 这一段，以 ${x_1}^2$ 为高，做一个窄条矩形。在 (x_1,x_2) 这一段，以 ${x_2}^2$ 为高，做一个窄条矩形。依此类推，一直到在 (x_{n-1},x_n) 这一段，以 ${x_n}^2$ 为高，做一个窄条矩形。所有这些窄条矩形的面积的和，记为 $U_n(h)$，显然是 $S(h)$ 的上界。有

$$\begin{aligned}U_n(h)&=\frac{h}{n}\left(\frac{h^2}{n^2}+\frac{4h^2}{n^2}+\cdots+\frac{n^2h^2}{n^2}\right)\\&=\frac{h^3}{n^3}(1+4+\cdots+n^2)\\&=\frac{h^3}{n^3}\frac{n(n+1)(2n+1)}{6}\\&=h^3(\frac{1}{3}+\frac{1}{2n}+\frac{1}{6n^2})\end{aligned}$$

在 (x_0, x_1) 这一段，以 ${x_0}^2$ 为高，做一个窄条矩形。在 (x_1, x_2) 这一段，以 ${x_1}^2$ 为高，做一个窄条矩形。依此类推，一直到在 (x_{n-1}, x_n) 这一段，以 ${x_{n-1}}^2$ 为高，做一个窄条矩形。所有这些窄条矩形的面积的和，记为 $L_n(h)$，显然是 $S(h)$ 的下界。有

$$\begin{aligned}
L_n(h) &= \frac{h}{n}\left(\frac{0h^2}{n^2} + \frac{h^2}{n^2} + \cdots + \frac{(n-1)^2h^2}{n^2}\right) \\
&= \frac{h^3}{n^3}(0 + 1 + 4 + \cdots + (n-1)^2) \\
&= \frac{h^3}{n^3}\frac{(n-1)n(2n-1)}{6} \\
&= h^3(\frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2})
\end{aligned}$$

8.1.3 极限

上例中，取 $n = 10$，有 $U_{10}(h) = h^3(\frac{1}{3} + \frac{1}{2}0.1 + \frac{1}{6}0.01)$；取 $n = 1000$，有 $U_{1000}(h) = h^3(\frac{1}{3} + \frac{1}{2}0.001 + \frac{1}{6}0.000001)$。显然，只要 n 取足够大的值，$U_n(h)$ 能够按任意的精度要求逼近 $U_\infty(h) = \frac{1}{3}h^3$，即数列 $\{U_n(h)\}$ 以 $U_\infty(h)$ 为极限，记为

$$\lim_{n\to\infty} U_n(h) = U_\infty(h) = \frac{1}{3}h^3.$$

所谓精度要求，就是给定一个正数 ε，典型取值可以是 0.01，或者 0.0001。

所谓 n 取足够大的值，就是存在一个与精度要求 ε 有关的阈值 $N_\varepsilon(h)$，对所有大于 $N_\varepsilon(h)$ 的 n，相应的 $U_n(h)$ 均符合精度要求。即，当 $n > N_\varepsilon(h)$ 时，恒有

$$|U_n(h) - U_\infty(h)| = \frac{h^3}{2n}(1 + \frac{1}{3n}) < \frac{h^3}{2n}(1+1) = \frac{h^3}{n} < \varepsilon.$$

显然，只要取 $N_\varepsilon(h) > \frac{h^3}{\varepsilon}$ 即可。比如，指定 $\varepsilon = 0.01$，则 $N_\varepsilon(h) = 100h^3$。如果 $h = 2$，则 $N_\varepsilon(2) = 800$。也就是说，$U_{800}(2)$ 是 $U_\infty(2)$ 的可接受的近似值。

不信可以试试看。$U_\infty(2) = 8/3 = 2.6667$，$U_{800}(2) = 8(0.3333 + 0.0006 + 0.0000) = 2.6712$，两者之差为 $0.0045 < 0.01$。

同理，对任意给定的正数 ε，存在相应的正整数 $M_\varepsilon(h)$，使得当 $n > M_\varepsilon(h)$ 时，恒有

$$|L_n(h) - L_\infty(h)| = \frac{h^3}{2n}\left|-1 + \frac{1}{3n}\right| < \frac{h^3}{2n}(1+1) = \frac{h^3}{n} < \varepsilon.$$

其中 $L_\infty(h) = \frac{1}{3}h^3$。显然，只要取 $M_\varepsilon(h) > \frac{h^3}{\varepsilon}$ 即可。这就是说，

$$\lim_{n\to\infty} L_n(h) = L_\infty(h) = \frac{1}{3}h^3.$$

8.1.4 夹逼定理

上例中，对任意 $n \geq 1$，恒有 $L_n(h) \leq S(h) \leq U_n(h)$，而我们又证明了 $\lim\limits_{n\to\infty} L_n(h) = \lim\limits_{n\to\infty} U_n(h) = \frac{1}{3}h^3$，所以，直观上不难确信，$S(h) = \frac{1}{3}h^3$。

如果一个待定的值，不大于一个序列中的任何项，则称该序列是该待定值的上界序列。如果一个待定的值，不小于一个序列中的任何项，则称该序列是该待定值的下界序列。如果一个待定值的上界序列和下界序列都存在极限而且两个极限的值相等，那么，该待定值就等于这个极限值。这称为夹逼定理。

微积分作为一种“算术”，按给定的精度要求，求出待定量的一个可接受的近似值当然是重要的，然而，更重要的是，微积分建立了一套分析推理方法，厘清了“对任意给定的精度要求，均存在可接受的近似值”的条件，也就是极限值存在的条件。这样，在符合这些条件的情况下，我们就说这个极限值是待定量的“精确解”。

当一个近似解可以“要多精确就有多精确”的时候，我们没有理由说它不是精确解。极限就是这样一个出身于“近似”，而又超越了“近似”的概念。求导是求极限，求积分也是求极限，说“微积分就是求极限”一点都不过分。

初学者对所谓的 ε-N 语言、以及类似的 ε-δ 语言，常会感到很别扭，知道这样说是对的，但是不知道为什么非要这么说不可。难道没有更“自然”的表述方法吗？

如果你理解了“极限出于近似而胜于近似”的道理，就知道 ε-N 语言是“最自然”的表述方法了。

朴素的极限逼近问题从古希腊人发现无理数的时候起就出现了。正方形对角线不可公度，那么该用怎样的可公度的线段来近似它呢？斐波那契数列相邻两项的比趋于黄金分割数，祖冲之割圆术求圆周率，等等都包含着极限的思想。然而，直到发明了 ε-N 和 ε-δ 语言，极限的概念才得到了严格而清晰的表达。

8.1.5 正圆锥的体积公式

在空间直角坐标系中，把正圆锥的顶点放在坐标原点，主轴放在 x 轴上。设圆锥的高为 h，底圆半径为 r。

用平行于底圆的平面把圆锥切成许许多多的小薄片，每个薄片是一个侧立的圆台，左边圆面稍小，右边圆面稍大。把每个薄片都扩大成截面为其右边圆面的圆柱，所有这些薄圆柱的体积的和，就是圆锥的体积的上界近似值。把每个薄片都缩小成截面为其左边圆面的圆柱，那么，这些薄圆柱的体积和，就是圆锥的体积的下界近似值。当这些薄片切得足够薄的时候，上界和下界近似值趋于同一个极限。

这个情景我们似曾相识。直接说结果吧，在 x 处的截面面积为 $S(x) = \dfrac{\pi r^2}{h^2}x^2$，因此，该圆锥的体积为

$$V = \int_0^h S(x)\,\mathrm{d}x = \frac{\pi r^2}{h^2}\int_0^h x^2\,\mathrm{d}x = \frac{\pi r^2}{h^2} \times \frac{1}{3}h^3 = \frac{1}{3}\pi r^2 h.$$

8.1.6 球的体积公式

把半个球扣在桌面上，设其半径为 r。取一圆柱，高和截面半径都为 r。从该圆柱中挖去一个漏斗形的圆锥，其顶点是圆柱底面的圆心，其底面是圆柱的顶面。我们来证，半球的体积等于该圆柱剩下的部分的体积。

在高度 $h < r$ 上，半球的横截面是圆，半径为 $r_h = \sqrt{r^2 - h^2}$，因此，横截面的面积为 $\pi {r_h}^2 = \pi(r^2 - h^2)$；上述圆柱剩余部分的横截面是圆环，内圆半径为 h，外圆半径为 r，因此，该圆环的面积为 $\pi r^2 - \pi h^2$。两者在任何高度上的横截面的面积都相等，按照前面讲的分割成薄片求体积的和的方法，显见两者的体积相等。

上述圆柱的体积为 $\pi r^2 r$，挖去的圆锥的体积为 $\dfrac{1}{3}\pi r^2 r$，因此，半球的体积为 $\dfrac{2}{3}\pi r^3$。一个完整的球的体积为 $\dfrac{4}{3}\pi r^3$。

8.1.7 开普勒三定律

开普勒的工作也很能体现“微积分是一种算术”的思想。他继承了导师第谷的大量观测数据，进行了十多年的艰苦计算，先后总结出了轨道定律、面积定律和周期定律，合称开普勒三定律，为牛顿提出万有引力定律奠定了坚实的基础。

轨道定律是说，行星的轨道是椭圆，太阳在椭圆的一个焦点上。

面积定律是说，行星与太阳的连线在相等的时间内扫过的面积相等。

周期定律是说，行星公转周期的平方与椭圆轨道半长轴的立方成正比。

我们只看面积定律。设在一段时间 Δt 之内，行星的位移为 $\Delta\vec{r}$。在 Δt 足够小的情况下，行星与太阳的连线，即矢径 $\vec{r}$，扫过的面积 ΔS 近似等于 $\dfrac{1}{2}\Delta\vec{r} \times \vec{r}$。此处的 $\times$ 是矢量的叉乘运算。取 $\Delta t \to 0$ 时的极限，有

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \lim_{\Delta t\to 0}\frac{\Delta S}{\Delta t} = \frac{1}{2}\lim_{\Delta t\to 0}\frac{\Delta\vec{r}}{\Delta t} \times \vec{r} = \frac{1}{2}\frac{\mathrm{d}\vec{r}}{\mathrm{d}t} \times \vec{r} = \frac{1}{2}\vec{v} \times \vec{r}.$$

这个值是一个常数，意味着 $\dfrac{\mathrm{d}}{\mathrm{d}t}\dfrac{\mathrm{d}S}{\mathrm{d}t} = 0$。由于 $\vec{v} \times \vec{v} = 0$，所以

$$\frac{\mathrm{d}}{\mathrm{d}t}(\vec{v} \times \vec{r}) = \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} \times \vec{r} + \vec{v} \times \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \vec{a} \times \vec{r} = 0.$$

这就是说，加速度 $\vec{a}$ 与矢径 $\vec{r}$ 平行。因此，引力场是中心力场。

古人云，“工欲善其事，必先利其器”。从这个例子中，我们可以感受到发明适当的数学工具对于表述和分析问题有多么重要。

8.2 微分

8.2.1 一元函数的导数

函数 $y=f(x)$ 的导函数，简称导数，代数上定义为

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

其几何意义是曲线 $y=f(x)$ 在点 $(x,f(x))$ 上的切线的斜率。

现在我们解释一下 $\dfrac{\mathrm{d}}{\mathrm{d}x}\displaystyle\int_a^x f(x)\,\mathrm{d}x=f(x)$。令 $z=g(x)=\displaystyle\int_a^x f(x)\,\mathrm{d}x$。在适当情况下，该积分可解释为曲线 $y=f(x)$ 下在区间 (a,x) 内的曲边梯形的面积。有

$$\frac{\mathrm{d}z}{\mathrm{d}x}=\lim_{\Delta x\to 0}\frac{\displaystyle\int_a^{x+\Delta x}f(x)\,\mathrm{d}x-\int_a^x f(x)\,\mathrm{d}x}{\Delta x}=\lim_{\Delta x\to 0}\frac{\displaystyle\int_x^{x+\Delta x}f(x)\,\mathrm{d}x}{\Delta x}.$$

极限内分子是一个窄条的面积，分母是该窄条的宽度，因此，该分式趋于窄条的高度，即 $f(x)$。

8.2.2 微积分基本定理

该定理说，在适当条件下，$\displaystyle\int_a^x \frac{\mathrm{d}}{\mathrm{d}x}f(x)\,\mathrm{d}x=f(x)\Big|_a^x=f(x)-f(a)$。我们来看它的一个简单的几何上的解释。

把区间 (a,x) 划分成 n 段，断点从左到右依次记为 $x_1,x_2,\dots,x_{n-1}$，并记 $x_0=a$ 和 $x_n=x$。记 $\Delta x=\max\limits_{i=1}^{n}(x_i-x_{i-1})$。那么，当 $\Delta x\to 0$ 时，$n\to\infty$。有

$$\begin{aligned}\int_a^x \frac{\mathrm{d}}{\mathrm{d}x}f(x)\,\mathrm{d}x &= \lim_{\Delta x\to 0}\sum_{i=1}^{n}f'(x_{i-1})(x_i-x_{i-1})\\ &= \lim_{\Delta x\to 0}\sum_{i=1}^{n}(f(x_i)-f(x_{i-1}))\\ &= \lim_{\Delta x\to 0}(f(x_n)-f(x_0))\\ &= f(x)-f(a)\end{aligned}$$

就是说，适当条件下，在每个点 $(x_{i-1},f(x_{i-1}))$ 上都可做相应的切线，y 沿切线所获得的增加量为 $f'(x_{i-1})(x_i-x_{i-1})$，这是 y 的实际增量 $f(x_i)-f(x_{i-1})$ 的主要部分。根据导数的定义，两者在取上述极限时趋于一致。

8.2.3 微分

注意到 $\dfrac{\mathrm{d}}{\mathrm{d}x}f(x)=\lim\limits_{\Delta x\to 0}\dfrac{\Delta}{\Delta x}f(x)$，所以，算符 $\dfrac{\mathrm{d}}{\mathrm{d}x}$ 本质上是一个完整的求极限算符。单独讨论 $\mathrm{d}x$ 或者 $\mathrm{d}y$ 是没有意义的。

说 $\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=\dfrac{\lim\limits_{\Delta x\to 0}\Delta y}{\lim\limits_{\Delta x\to 0}\Delta x}$ 是不对的，因为 $\lim\limits_{\Delta x\to 0}\Delta x=0$，以零为分母的分式根本没有意义。说 $\lim\limits_{\Delta x\to 0}\Delta y=f'(x)\lim\limits_{\Delta x\to 0}\Delta x$ 倒是没有错，但此式就是 $0=0$，实在也没有什么意义。

有意义的是，在微积分基本定理所述之情形下，

$$\lim_{\Delta x\to 0}\sum_{i=1}^{n}\Delta y_i=\lim_{\Delta x\to 0}\sum_{i=1}^{n}f'(x_i)\Delta x_i.$$

用积分符号，可以把此式写成

$$\int_{f(a)}^{f(b)}\mathrm{d}y=\int_a^b f'(x)\,\mathrm{d}x.$$

当最终的积分是不言而喻的，而在讨论的某个阶段，我们集中关注局部性质时，上式可以进一步简写为

$$\mathrm{d}y=f'(x)\,\mathrm{d}x.$$

所以，微分本质上是积分的缩写形式，是在不关注积分上下限时对积分号的省略，刻画的是在积分中使用的微小单元 $\mathrm{d}y$ 和 $\mathrm{d}x$ 之间的关系。单独的 $\mathrm{d}y$ 和 $f'(x)\,\mathrm{d}x$ 都没有意义，但是两者相等的这种关系是有意义的，因为它意味着，尽管 $f'(x)$ 是通过求极限来定义的，然而在形式上，$f'(x)$ 恰好等于两个微元之比。

在近似计算中，常有 $\Delta y\approx f'(x)\Delta x$。它与上式很像，但是，其中的 Δx 和 Δy 是实际的数值，与极限和积分无关。单独讨论 Δx 或 Δy 都是有意义的。

8.2.4 不定积分

注意到原函数 $\displaystyle\int_a^x f'(t)\,\mathrm{d}t=f(x)-f(a)$，其中的积分下限 a 可以是可积区域内的任何值，任何定积分 $\displaystyle\int_{x_0}^{x_1}f'(x)\,\mathrm{d}x=f(x_1)-f(x_0)=(f(x_1)-f(a))-(f(x_0)-f(a))$，可见 a 对定积分无关紧要。重要的仅仅是根据 $f'(x)$ 求 $f(x)$ 而已。

于是人们发明了一种更能够集中反映这种关系的表达方式：$\displaystyle\int f'(x)\,\mathrm{d}x=f(x)+c$，其中 c 表示任意常数。这种表达方式称为不定积分，也称反微分。它与 $\mathrm{d}y=f'(x)\,\mathrm{d}x$ 表达完全相同的数学实质，区别只是语义上的侧重点有所不同。

8.2.5 偏导数

对于多元函数 $w=f(x,y,z)$，如果限定 $y=y_0,\quad z=z_0$，就得到关于 x 的

一元函数 $w_1 = f(x, y_0, z_0) = f(x, y, z)\Big|_{y=y_0, z=z_0}$。对 x 求导，得

$$\frac{\mathrm{d}w_1}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x} f(x, y_0, z_0) = \frac{\mathrm{d}}{\mathrm{d}x}\left(f(x, y, z)\Big|_{y=y_0, z=z_0} \right) = g(x, y_0, z_0).$$

显然，y_0 和 z_0 是有取不同值的自由的，因此，一元函数 $g(x, y_0, z_0)$ 实际上是一个新的多元函数 $g(x, y, z)$，称为函数 $f(x, y, z)$ 关于自变量 x 的偏导函数，简称偏导数，记为

$$\frac{\partial}{\partial x} f(x, y, z) = \frac{\mathrm{d}}{\mathrm{d}x}\left(f(x, y, z)\Big|_{y,z} \right) = g(x, y, z).$$

同理可做函数 $f(x, y, z)$ 关于自变量 y 和 z 的偏导数，

$$\frac{\partial}{\partial y} f(x, y, z) = \frac{\mathrm{d}}{\mathrm{d}y}\left(f(x, y, z)\Big|_{x,z} \right), \quad \frac{\partial}{\partial z} f(x, y, z) = \frac{\mathrm{d}}{\mathrm{d}z}\left(f(x, y, z)\Big|_{x,y} \right).$$

8.2.6 线积分与全微分

行星在公转时，会受到太阳的引力。考虑行星轨道上一个微元 $\mathrm{d}\vec{r}$，该微元内引力可视为恒力 $\vec{F}$，那么从该微元的起点到终点，引力所做的功为 $\mathrm{d}W = \vec{F} \cdot \mathrm{d}\vec{r}$，其中的 $\cdot$ 表示矢量的点积，即数量积。因此，从一段轨道 γ 的起点到终点，引力所做的总功为

$$W = \int_\gamma \mathrm{d}W = \int_\gamma \vec{F} \cdot \mathrm{d}\vec{r} = \int_\gamma (F_x \,\mathrm{d}x + F_y \,\mathrm{d}y + F_z \,\mathrm{d}z).$$

这是线积分的一个典型例子。

记行星的势能为 U，那么，有

$$F_x = -\frac{\partial U}{\partial x}, \quad F_y = -\frac{\partial U}{\partial y}, \quad F_z = -\frac{\partial U}{\partial z}.$$

也就是说，引力在微元上所做的功为

$$\mathrm{d}W = -\left(\frac{\partial U}{\partial x} \mathrm{d}x + \frac{\partial U}{\partial y} \mathrm{d}y + \frac{\partial U}{\partial z} \mathrm{d}z \right) = -\,\mathrm{d}U.$$

其中的 $\mathrm{d}U = \dfrac{\partial U}{\partial x} \mathrm{d}x + \dfrac{\partial U}{\partial y} \mathrm{d}y + \dfrac{\partial U}{\partial z} \mathrm{d}z$ 称为函数 $U(x, y, z)$ 的全微分。

行星公转问题中，引力所做的功等于行星动能 K 的增加，因此有 $\mathrm{d}W = \mathrm{d}K$，进而上式变为 $\mathrm{d}K = -\,\mathrm{d}U$，即 $\mathrm{d}(K + U) = 0$。其中动能 K 与势能 U 的和就是机械能 $E = K + U$。因此，$\mathrm{d}E = 0$，这表示在行星公转问题中，机械能守恒。

8.2.7 方向导数与等值面

给定多元函数 $f(x,y,z)$，适当条件下可做全微分 $\mathrm{d}f=\dfrac{\partial f}{\partial x}\mathrm{d}x+\dfrac{\partial f}{\partial y}\mathrm{d}y+\dfrac{\partial f}{\partial z}\mathrm{d}z$。其中 $\mathrm{d}\vec{r}=\mathrm{d}x\,\vec{\imath}+\mathrm{d}y\,\vec{\jmath}+\mathrm{d}z\,\vec{k}$ 为矢量微元，可看做是过点 (x,y,z) 的曲线在该点邻域内的微小一段，相应的积分为沿曲线的线积分。

沿此曲线可求曲线长度 $s=\displaystyle\int_\gamma \mathrm{d}s$，其中 $\mathrm{d}s=|\mathrm{d}\vec{r}|=\sqrt{\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2}$ 为上述微小一段的长度。以曲线长度 s 为参数，该曲线可表述为参数方程 $x=x(s),\quad y=y(s),\quad z=z(s)$。因而函数 $f(x,y,z)=f(x(s),y(s),z(s))$ 实际上是关于 s 的一元函数，进而 $\dfrac{\mathrm{d}f}{\mathrm{d}s}$ 是有意义的。它表示动点沿曲线移动时，函数值随路程的变化率。

上述微小一段的方向就是此曲线在该处的切线的方向，其方向余弦为 $\cos\alpha=\dfrac{\mathrm{d}x}{\mathrm{d}s},\quad \cos\beta=\dfrac{\mathrm{d}y}{\mathrm{d}s},\quad \cos\gamma=\dfrac{\mathrm{d}z}{\mathrm{d}s}$。参照全微分，显有

$$\frac{\mathrm{d}f}{\mathrm{d}s}=\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}s}+\frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}s}+\frac{\partial f}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}s}=\frac{\partial f}{\partial x}\cos\alpha+\frac{\partial f}{\partial y}\cos\beta+\frac{\partial f}{\partial z}\cos\gamma.$$

这个称为函数 $f(x,y,z)$ 沿该曲线在此点的切线方向的方向导数。

选择恰当的曲线，可以使动点沿曲线移动时，函数值不变，即 $\dfrac{\mathrm{d}f}{\mathrm{d}s}=0$。比如，在地形图上，动点沿等高线移动时，高度不变。再比如，行星轨道如果是标准的圆，那么它的势能不随公转运动而改变。

给定函数 $w=f(x,y,z)$，方程 $f(x,y,z)=c$ 的解通常会构成一个曲面，称为该函数的等值面，或者更一般的，等值子流形。对等值面上的任何曲线，恒有 $\dfrac{\mathrm{d}f}{\mathrm{d}s}=0$。

8.3 张量场

8.3.1 曲线坐标系

有三个特殊的函数 $u_1(x,y,z)=x,\quad u_2(x,y,z)=y,\quad u_3(x,y,z)=z$，称为坐标投影函数。常简记为 $u_1=x,\quad u_2=y,\quad u_3=z$。

给定点 $P(x_0,y_0,z_0)$。函数 u_1 过该点的等值面为平面 $x=x_0$，它平行于 y-z 平面。同理，函数 u_2 和 u_3 过该点的等值面也是平面，分别平行于 x-z 平面和 x-y 平面。三个等值面在点 P 的邻域上恰做成一个局部的坐标框架，简称局部标架。三个坐标轴为相应的等值面的交线。本例中恰为直线。

对空间任意点 P，用 r 表示 OP 的长度，用 θ 表示 OP 与 z 轴的夹角，用 ϕ 表示过点 P 和 z 轴的平面与 x-z 平面的二面角，称三元组 (r,θ,ϕ) 为点 P 的球坐标。给空间每个点都配上球坐标，就做成空间的球坐标系。

设给定点 P 的球坐标为 (r_0,θ_0,ϕ_0)。等值面 $r=r_0$ 是球面，$\theta=\theta_0$ 是圆锥面，$\phi=\phi_0$ 是半平面。半平面与圆锥面的交线为直线。沿此直线，只有 r 变

动而 θ 和 ϕ 不变动，称为 r 坐标轴。球面与半平面的交线为经圆，沿此经圆，只有 θ 变动而 r 和 ϕ 不变动，称为 θ 坐标轴。球面与圆锥面的交线为纬圆，沿此纬圆，只有 ϕ 变动而 r 和 θ 不变动，称为 ϕ 坐标轴。因为三个坐标轴中有曲线，所以，球坐标系是曲线坐标系。

在球坐标系中，过点 P 的三个等值面和三个坐标轴在该点的邻域上也做成一个局部标架。由于三个坐标轴相互垂直，所以称这个局部标架是正交的。

一般地，适当条件下，三个函数 $u = u(x, y, z)$, $v = v(x, y, z)$, $w = w(x, y, z)$ 可以通过等值面相交来确定三个曲线坐标轴，进而建立曲线坐标系。这样的函数也称坐标函数。注意，一般的曲线坐标系中，局部标架不一定正交。

8.3.2 标量场

为了研究各种不同几何性质的“空间”，常需将“空间”抽象为几何点的集合。我们将其记为 M。给“空间”M 中的所有几何点 P，都指定相应的数值 $w = w(P)$。这定义了一个映射 $w\colon M \to \mathbf{R}$。这些指定在几何点上的数值称为标量，而这个映射称为标量场。

在本节之内，请暂时把“空间”M 理解为流形上的，与实欧氏空间中的开集同胚的一个局部。

当采用直角坐标系时，说几何点 P 的坐标是 $P(x, y, z)$，实际上是给任意点 P，指定了三个数值 $x = x(P)$, $y = y(P)$, $z = z(P)$。这定义了从“空间”M 到三维实欧氏空间 $\mathbf{R}^3$ 的子集 D 的双射 $\psi\colon M \to D$, $D \subseteq \mathbf{R}^3$，使得 $\psi(P) = (x(P), y(P), z(P))$。其逆映射为 $\psi^{-1}\colon D \to M$，可写成 $P = \psi^{-1}(x, y, z)$。这样，$w = w(P) = w(\psi^{-1}(x, y, z))$。记 $f = w \circ \psi^{-1}$，则有 $w = f(x, y, z)$，称为标量场 $w(P)$ 在直角坐标系下的表示。

换成球坐标系。说几何点 P 的坐标是 $P(r, \theta, \phi)$，这定义了从 M 到三维实欧氏空间 $\mathbf{R}^3$ 的子集 D_2 的双射 $\xi\colon M \to D_2$, $D_2 \subseteq \mathbf{R}^3$。其逆映射为 $\xi^{-1}\colon D_2 \to M$，即 $P = \xi^{-1}(r, \theta, \phi)$。于是，$w = w(P) = w(\xi^{-1}(r, \theta, \phi))$。记 $g = w \circ \xi^{-1}$，则有 $w = g(r, \theta, \phi)$，称为标量场 $w(P)$ 在球坐标系下的表示。

按球坐标的定义，有坐标变换

$$\begin{cases} x = x(r, \theta, \phi) = r \sin\theta \cos\phi \\ y = y(r, \theta, \phi) = r \sin\theta \sin\phi \\ z = z(r, \theta, \phi) = r \cos\theta \end{cases}.$$

这可看作是从 D_2 到 D 的双射，记为 $\sigma\colon D_2 \to D$，使得 $(x(P), y(P), z(P)) = \sigma(r(P), \theta(P), \phi(P))$，即 $\psi(P) = \sigma(\xi(P))$，也即 $\psi = \sigma \circ \xi$。这意味着 $\psi^{-1} = \xi^{-1} \circ \sigma^{-1}$，即 $\xi^{-1} = \psi^{-1} \circ \sigma$。这样，$g = w \circ \xi^{-1} = w \circ \psi^{-1} \circ \sigma = f \circ \sigma$。这个关系其实就是 $g(r, \theta, \phi) = f(x(r, \theta, \phi),\ y(r, \theta, \phi),\ z(r, \theta, \phi))$。

任意两个从 $\mathbf{R}^n$ 到 $\mathbf{R}$ 的函数 f 和 g，如果存在一个坐标变换 σ，使得 $g = f \circ \sigma$，则称 f 和 g 等价。所有相互等价的函数，都表示同一个标量场。

8.3.3 流形

我们用一个实例来说明流形的概念。如图 8.1 所示。 平面直角坐标系中的

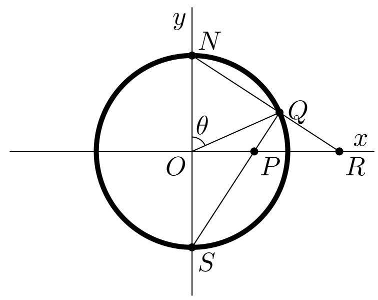

图 8.1: 南极投影和北极投影

单位圆，作为几何点的集合，记为 $\mathcal{C}$。点 $S(0,-1)$ 称为南极点，点 $N(0,1)$ 称为北极点。集合 $\mathcal{C}\setminus\{S\}$ 中任意一点 Q，与 S 的连线交 x 轴于点 P，称点 P 为点 Q 的南极投影。集合 $\mathcal{C}\setminus\{N\}$ 中任意一点 Q，与 N 的连线交 x 轴于点 R，称点 R 为点 Q 的北极投影。记 $\angle QON$ 为 θ，并以顺时针为正方向。规定 $-\pi \le \theta < \pi$。

我们在单位圆 $\mathcal{C}$ 上建立标量场 $\theta(Q)$。易见 $x_P = \tan\dfrac{\theta}{2}$，而 $x_R = \cot\dfrac{\theta}{2}$。就是说，在 $\mathcal{C}\setminus\{S\}$ 上，$\theta(Q) = f(x_P) = 2\arctan x_P$；在 $\mathcal{C}\setminus\{N\}$ 上，$\theta(Q) = g(x_R) = 2\arctan\dfrac{1}{x_R}$。因为 $x_P(Q)x_R(Q) = \tan\dfrac{\theta}{2}\cot\dfrac{\theta}{2} = 1$，所以，在 $\mathcal{C}\setminus\{N,S\}$ 上，$f(x_P(Q)) = g(x_R(Q)) = \theta(Q)$。可见 $\theta(Q)$ 确为标量场。

给个土鳖的定义：一个集合，如果能等于一系列子集的并，其中每个子集可以双射到 $\mathbf{R}^n$ 中的开集，则称该集合为 n 维流形。配上其他的条件，可定义可微流形、光滑流形、黎曼流形等。

上例中，$\mathcal{C}\setminus\{N\}$ 和 $\mathcal{C}\setminus\{S\}$ 均可双射到 $\mathbf{R}$。而且 $\mathcal{C} = (\mathcal{C}\setminus\{N\})\cup(\mathcal{C}\setminus\{S\})$，所以，单位圆是一维流形。

单位球面可以用类似的方法分别从南极和北极双射到 $\mathbf{R}^2$，两者的并则覆盖整个球面。所以，单位球面是二维流形。

8.3.4 局部标架之间的坐标变换

微积分的基本思想是，为了研究整体的性质，将整体划分成许多微小单元，研究在微小单元上的局部性质，然后把局部性质汇总起来，得到整体性质。所谓的局部性质，通常放在局部标架坐标系上进行讨论。

对前述的标量场 $w=w(P)$，考虑几何点 P 邻域内的另一个几何点 Q。记 $\Delta w=w(Q)-w(P)$，$\Delta x=x(Q)-x(P)$，等等，取极限 $Q\to P$，则有全微分

$$\mathrm{d}w=\frac{\partial w}{\partial x}\,\mathrm{d}x+\frac{\partial w}{\partial y}\,\mathrm{d}y+\frac{\partial w}{\partial z}\,\mathrm{d}z=\frac{\partial f}{\partial x}\,\mathrm{d}x+\frac{\partial f}{\partial y}\,\mathrm{d}y+\frac{\partial f}{\partial z}\,\mathrm{d}z.$$

在局部微小范围内，$(\,\mathrm{d}x,\mathrm{d}y,\mathrm{d}z)$ 可以看做是局部标架坐标系中点 Q 的坐标，而点 P 是局部标架的坐标原点。$\dfrac{\partial w}{\partial x}=\dfrac{\partial f}{\partial x}$，$\dfrac{\partial w}{\partial y}=\dfrac{\partial f}{\partial y}$，$\dfrac{\partial w}{\partial z}=\dfrac{\partial f}{\partial z}$ 在局部微小范围内可看做常数。因此，该全微分在局部微小范围内就是线性函数 $W=aX+bY+cZ$。

换成球坐标系，有全微分

$$\mathrm{d}w=\frac{\partial w}{\partial r}\,\mathrm{d}r+\frac{\partial w}{\partial \theta}\,\mathrm{d}\theta+\frac{\partial w}{\partial \phi}\,\mathrm{d}\phi=\frac{\partial g}{\partial r}\,\mathrm{d}r+\frac{\partial g}{\partial \theta}\,\mathrm{d}\theta+\frac{\partial g}{\partial \phi}\,\mathrm{d}\phi.$$

在局部微小范围内，这是线性函数 $W=a'X'+b'Y'+c'Z'$。

可以把球坐标系的三个坐标函数分别作为标量场来处理，有

$$\begin{cases}\mathrm{d}r\;=\;\dfrac{\partial r}{\partial x}\,\mathrm{d}x+\dfrac{\partial r}{\partial y}\,\mathrm{d}y+\dfrac{\partial r}{\partial z}\,\mathrm{d}z\\ \mathrm{d}\theta\;=\;\dfrac{\partial \theta}{\partial x}\,\mathrm{d}x+\dfrac{\partial \theta}{\partial y}\,\mathrm{d}y+\dfrac{\partial \theta}{\partial z}\,\mathrm{d}z\\ \mathrm{d}\phi\;=\;\dfrac{\partial \phi}{\partial x}\,\mathrm{d}x+\dfrac{\partial \phi}{\partial y}\,\mathrm{d}y+\dfrac{\partial \phi}{\partial z}\,\mathrm{d}z\end{cases}.$$

$(\,\mathrm{d}r,\mathrm{d}\theta,\mathrm{d}\phi)$ 是球坐标系的局部标架坐标系中给定点 Q 的坐标。从 $(\,\mathrm{d}x,\mathrm{d}y,\mathrm{d}z)$ 到 $(\,\mathrm{d}r,\mathrm{d}\theta,\mathrm{d}\phi)$ 是原点相同的两个局部标架坐标系之间的坐标变换。在局部微小范围内，这实际上就是线性变换

$$\begin{cases}X'\;=\;a_{11}X+a_{12}Y+a_{13}Z\\ Y'\;=\;a_{21}X+a_{22}Y+a_{23}Z\\ Z'\;=\;a_{31}X+a_{32}Y+a_{33}Z\end{cases}.$$

记这个线性变换的矩阵为 A。

8.3.5 逆变矢量和协变矢量

把上述坐标变换代入方程 $aX+bY+cZ=a'X'+b'Y'+c'Z'$，得

$$\begin{aligned}&(a-(a_{11}a'+a_{21}b'+a_{31}c'))X\\ +\;&(b-(a_{12}a'+a_{22}b'+a_{32}c'))Y\\ +\;&(c-(a_{13}a'+a_{23}b'+a_{33}c'))Z=0\end{aligned}$$

此式中的 X,Y,Z 为自由变量，因此必有

$$\begin{cases}a\;=\;a_{11}a'+a_{21}b'+a_{31}c'\\ b\;=\;a_{12}a'+a_{22}b'+a_{32}c'\\ c\;=\;a_{13}a'+a_{23}b'+a_{33}c'\end{cases}.$$

用曲线坐标系的符号来表示，这组方程就是

$$\begin{cases} \dfrac{\partial w}{\partial x} = \dfrac{\partial r}{\partial x}\dfrac{\partial w}{\partial r} + \dfrac{\partial \theta}{\partial x}\dfrac{\partial w}{\partial \theta} + \dfrac{\partial \phi}{\partial x}\dfrac{\partial w}{\partial \phi} \\ \dfrac{\partial w}{\partial y} = \dfrac{\partial r}{\partial y}\dfrac{\partial w}{\partial r} + \dfrac{\partial \theta}{\partial y}\dfrac{\partial w}{\partial \theta} + \dfrac{\partial \phi}{\partial y}\dfrac{\partial w}{\partial \phi} \\ \dfrac{\partial w}{\partial z} = \dfrac{\partial r}{\partial z}\dfrac{\partial w}{\partial r} + \dfrac{\partial \theta}{\partial z}\dfrac{\partial w}{\partial \theta} + \dfrac{\partial \phi}{\partial z}\dfrac{\partial w}{\partial \phi} \end{cases}.$$

由于上述推理中直角坐标系与球坐标系只是作为两种不同的曲线坐标系而出现，并没有哪个更特殊，因此，交换两者，同理易得

$$\begin{cases} \dfrac{\partial w}{\partial r} = \dfrac{\partial x}{\partial r}\dfrac{\partial w}{\partial x} + \dfrac{\partial y}{\partial r}\dfrac{\partial w}{\partial y} + \dfrac{\partial z}{\partial r}\dfrac{\partial w}{\partial z} \\ \dfrac{\partial w}{\partial \theta} = \dfrac{\partial x}{\partial \theta}\dfrac{\partial w}{\partial x} + \dfrac{\partial y}{\partial \theta}\dfrac{\partial w}{\partial y} + \dfrac{\partial z}{\partial \theta}\dfrac{\partial w}{\partial z} \\ \dfrac{\partial w}{\partial \phi} = \dfrac{\partial x}{\partial \phi}\dfrac{\partial w}{\partial x} + \dfrac{\partial y}{\partial \phi}\dfrac{\partial w}{\partial y} + \dfrac{\partial z}{\partial \phi}\dfrac{\partial w}{\partial z} \end{cases}.$$

也就是

$$\begin{cases} a' = b_{11}a + b_{12}b + b_{13}c \\ b' = b_{21}a + b_{22}b + b_{23}c \\ c' = b_{31}a + b_{32}b + b_{33}c \end{cases}.$$

记这个线性变换的矩阵为 $B = (A^T)^{-1}$。其中的符号 M^T 表示矩阵 M 的转置，M^{-1} 表示 M 的逆。有 $A = (B^{-1})^T = (B^T)^{-1}$。不熟悉矩阵的看官只要知道 A 和 B 可以相互决定就可以了。

在进行局部标架坐标系变换时，和 (X, Y, Z) 一样，按照上述矩阵 A 进行变换的矢量，称为逆变矢量，常记为 (X^1, X^2, X^3)，或简记为 X^i；和 (a, b, c) 一样，按照矩阵 B 进行变换的矢量，称为协变矢量，常记为 (X_1, X_2, X_3)，或简记为 X_i。简记法中的 i 称为指标。写在右上角的，称为逆变指标；写在右下角的，称为协变指标。

注意，$(X^1, X^2, X^3) = (X, Y, Z)$，而 $(X_1, X_2, X_3) = (a, b, c)$，两者毫无关系，连函数关系或变换关系也没有。除非在特定的问题中，用某种方式把两者联系起来。比如射影几何里的对射变换，或者黎曼几何里的指标上升算符与指标下降算符。

8.3.6 矢量的并积与张量

给定两个逆变矢量 (X, Y, Z) 和 (U, V, W)，称 9-元组 $(XU, XV, XW, YU, YV, YW, ZU, ZV, ZW)$ 为两者的并积。因为 $X' = a_{11}X + a_{12}Y + a_{13}Z$，并且 $U' = a_{11}U + a_{12}V + a_{13}W$，所以，

$$\begin{aligned} X'U' &= (a_{11}X + a_{12}Y + a_{13}Z)(a_{11}U + a_{12}V + a_{13}W) \\ &= a_{11}a_{11}XU + a_{11}a_{12}XV + a_{11}a_{13}XW \end{aligned}$$

$$+a_{12}a_{11}YU + a_{12}a_{12}YV + a_{12}a_{13}YW$$
$$+a_{13}a_{11}ZU + a_{13}a_{12}ZV + a_{13}a_{13}ZW$$

其他 8 个分量 $X'V', X'W', Y'U', Y'V', Y'W', Z'U', Z'V', Z'W'$ 情况类似。用 X^i 表示 (X, Y, Z)，用 Y^j 表示 (U, V, W)，那么，上述并积就可记为 X^iY^j。

在进行局部标架坐标系变换的时候，任何两个逆变矢量的并积都按照上述规律做线性变换。任意多个同类型并积的和也是如此。如果一个 9 -元组按照同两个逆变矢量的并积一样的规律进行坐标变换，则称为二逆张量，常简记为 T^{ij}。

给定两个协变矢量 (a, b, c) 和 (p, q, r)，称 9-元组 $(ap, aq, ar, bp, bq, br, cp, cq, cr)$ 为两者的并积。因为 $a' = b_{11}a + b_{12}b + b_{13}c$，并且 $p' = b_{11}p + b_{12}q + b_{13}r$，所以，

$$\begin{aligned} a'p' &= (b_{11}a + b_{12}b + b_{13}c)(b_{11}p + b_{12}q + b_{13}r) \\ &= b_{11}b_{11}ap + b_{11}b_{12}aq + b_{11}b_{13}ar \\ &\quad + b_{12}b_{11}bp + b_{12}b_{12}bq + b_{12}b_{13}br \\ &\quad + b_{13}b_{11}cp + b_{13}b_{12}cq + b_{13}b_{13}cr \end{aligned}$$

其他 8 个分量 $a'q', a'r', b'p', b'q', b'r', c'p', c'q', c'r'$ 情况类似。用 X_i 和 Y_j 分别表示 (a, b, c) 和 (p, q, r)，那么，上述并积就可记为 X_iY_j。

在进行局部标架坐标系变换的时候，任何两个协变矢量的并积都按照上述规律做线性变换。任意多个同类型并积的和也是如此。如果一个 9 -元组按照同两个协变矢量的并积一样的规律进行坐标变换，则称为二协张量，常简记为 T_{ij}。

给定一个协变矢量 (a, b, c) 和一个逆变矢量 (X, Y, Z)，称 9-元组 $(aX, aY, aZ, bX, bY, bZ, cX, cY, cZ)$ 为两者的并积。因为 $a' = b_{11}a + b_{12}b + b_{13}c$，并且 $X' = a_{11}X + a_{12}Y + a_{13}Z$，所以，

$$\begin{aligned} a'X' &= (b_{11}a + b_{12}b + b_{13}c)(a_{11}X + a_{12}Y + c_{13}Z) \\ &= b_{11}a_{11}aX + b_{11}a_{12}aY + b_{11}a_{13}aZ \\ &\quad + b_{12}a_{11}bX + b_{12}a_{12}bY + b_{12}a_{13}bZ \\ &\quad + b_{13}a_{11}cX + b_{13}a_{12}cY + b_{13}a_{13}cZ \end{aligned}$$

其他 8 个分量 $a'Y', a'Z', b'X', b'Y', b'Z', c'X', c'Y', c'Z'$ 情况类似。用 X_i 表示 (a, b, c)，用 Y^j 表示 (X, Y, Z)，那么，上述并积就可记为 X_iY^j。

在进行局部标架坐标系变换的时候，一个协变矢量与一个逆变矢量的并积总是按照上述规律做线性变换。任意多个同类型并积的和也是如此。如果一个 9 -元组按照一个协变矢量与一个逆变矢量的并积一样的规律进行坐标变换，则称为一协一逆张量，常简记为 T_i^j。

同理可以定义三个逆变矢量的并积 $X^iY^jZ^k$ 与三逆张量 T^{ijk}，两个逆变矢量一个协变矢量的并积 $X^iY^jZ_k$ 与一协二逆张量 T_k^{ij} 等等。r 协 s 逆张量也称 (r,s) 型张量，常简记为 $T_{i_1i_2\cdots i_r}^{j_1j_2\cdots j_s}$。定义在 n 维欧氏空间上的 (r,s) 型张量有 n^{r+s} 个分量。因此也称 $r+s$ 阶张量。比如三协张量、三逆张量、一协二逆张量和二协一逆张量都是三阶张量。

矢量是一阶张量。标量是零阶张量。

8.3.7 张量的并积

设有二协张量 $(T_{11},T_{12},T_{13},T_{21},T_{22},T_{23},T_{31},T_{32},T_{33})$。按定义，在局部标架坐标系变换下，该张量的 9 个分量的变换规律为

$$\begin{aligned}T'_{11} =\ & b_{11}b_{11}T_{11}+b_{11}b_{12}T_{12}+b_{11}b_{13}T_{13}+b_{12}b_{11}T_{21}+b_{12}b_{12}T_{22}+b_{12}b_{13}T_{23}\\ & +b_{13}b_{11}T_{31}+b_{13}b_{12}T_{32}+b_{13}b_{13}T_{33}\end{aligned}$$

后 8 个分量 T'_{12} 等情况类似，兹不赘述。设另有逆变矢量 (X^1,X^2,X^3)，在上述坐标变换下的变换规律为 $X^{1\prime}=a_{11}X^1+a_{12}X^2+a_{13}X^3$，$X^{2\prime}$ 和 $X^{3\prime}$ 兹不赘述。那么，有

$$\begin{aligned}T'_{11}X^{1\prime} &= (b_{11}b_{11}T_{11}+b_{11}b_{12}T_{12}+\cdots)(a_{11}X^1+a_{12}X^2+a_{13}X^3)\\ &= b_{11}b_{11}a_{11}T_{11}X^1+b_{11}b_{11}a_{12}T_{11}X^2+b_{11}b_{11}a_{13}T_{11}X^3\\ &\quad +b_{11}b_{12}a_{11}T_{12}X^1+b_{11}b_{12}a_{12}T_{12}X^2+b_{11}b_{12}a_{13}T_{12}X^3\\ &\quad +b_{11}b_{13}a_{11}T_{13}X^1+b_{11}b_{13}a_{12}T_{13}X^2+b_{11}b_{13}a_{13}T_{13}X^3\\ &\quad +b_{12}b_{11}a_{11}T_{21}X^1+b_{12}b_{11}a_{12}T_{21}X^2+b_{12}b_{11}a_{13}T_{21}X^3\\ &\quad +b_{12}b_{12}a_{11}T_{22}X^1+\cdots \text{ 后面各项类似，兹不赘述。}\end{aligned}$$

后 26 个分量兹不赘述。称这个按照该规律进行坐标变换的 27-元组为该二协张量与该逆变矢量的并积，简记为 $T_{ij}X^k$。

同理可定义任意两个张量的并积。一个 (r,s) 型张量 $T_{i_1\cdots i_r}^{j_1\cdots j_s}$ 与一个 (p,q) 型张量 $T_{k_1\cdots k_p}^{\ell_1\cdots \ell_q}$ 的并积是 $T_{i_1\cdots i_r}^{j_1\cdots j_s}T_{k_1\cdots k_p}^{\ell_1\cdots \ell_q}$。也就是说，一个 m 阶张量与一个 n 阶张量的并积是一个 $m+n$ 阶张量。

8.3.8 张量的缩并

前面讲过标量场的全微分满足 $W = aX+bY+cZ = a'X'+b'Y'+c'Z'$。注意协变矢量 (a,b,c) 与逆变矢量 (X,Y,Z) 的并积有九个分量，分别是 $aX,aY,aZ,bX,bY,bZ,cX,cY,cZ$。选其中特定的三个 aX、bY 和 cZ 来做和，就得到一个不随坐标变换而变化的量，即标量。

这并不是巧合。用列矩阵 U 表示协变矢量 (a,b,c)，用列矩阵 V 表示逆变矢量 (X,Y,Z)，那么，有 $U' = BU$ 和 $V' = AV$。因为 $A = (B^T)^{-1}$，所以，$W = U'^T V' = (U^T B^T)(AV) = U^T(B^T A)V = U^T V$。

使用指标，可把上式表示为 $W = X_1X^1 + X_2X^2 + X_3X^3$，简记为 $W = X_iX^i$。注意简记法，并积是 X_iX^j，而这里是 X_iX^i。同一个张量的一个协变指标和一个逆变指标采用相同的符号，就表示对该符号的所有可能取值的那些分量求和。

实际上，对任何一协一逆张量 T_i^j，都可以从九个分量 $T_1^1, T_1^2, T_1^3, T_2^1, T_2^2, T_2^3, T_3^1, T_3^2, T_3^3$ 中挑出特定的三个分量 T_1^1, T_2^2, T_3^3 求和，而得到一个零阶张量 $T = T_1^1 + T_2^2 + T_3^3$，简记为 $T = T_i^i$。

再看一个例子。对二协一逆张量 T_{ij}^k，可以把 27 个分量分成三组，每组的指标 j 都相同。显见每组有九个分量。在一组之内，挑出特定的三个分量求和 $T_j = T_{1j}^1 + T_{2j}^2 + T_{3j}^3$，这样，三个组产生三个和，恰好构成一个协变矢量的三个分量。也就是说，一个二协一逆张量“收缩合并”成了一个协变矢量，简记为 $T_j = T_{ij}^i$。

如果把二协一逆张量 T_{ij}^k 的 27 个分量分成三组，每组的指标 i 都相同，那么，从每组内部挑出特定的三个分量求和 $T_i = T_{i1}^1 + T_{i2}^2 + T_{i3}^3$，也可得到一个协变矢量。简记为 $T_i = T_{ij}^j$。一般来说，T_{ij}^j 和 T_{ij}^i 是两个不同的协变矢量。

对一个张量做类似的处理，称为该张量的“缩并”运算。所谓“缩”，是指分量个数少了；所谓“并”，是指挑选特定的一些分量求和。

详细来说，给定一个既有逆变指标，又有协变指标的 (r,s) 型张量 $T_{i_1\cdots i_r}^{j_1\cdots j_s}$，它有 n^{r+s} 个分量。选择一个协变指标 i_p 和一个逆变指标 j_q。把所有分量分成 n^{r+s-2} 组，使得每一组有 n^2 个分量，它们除选定的两个指标外，其他指标都相同。对每一组，从 n^2 个分量中选择 i_p 与 j_q 相等的 n 个分量求和。每一组得到一个和，而这 n^{r+s-2} 个和恰构成一个 $(r-1,s-1)$ 型张量。就是说，对一个 (r,s) 型张量做缩并，得到一个 $(r-1,s-1)$ 型张量。简记为 $T_{i_1\cdots i_{p-1}i_{p+1}\cdots i_r}^{j_1\cdots j_{q-1}j_{q+1}\cdots j_s} = T_{i_1\cdots i_{p-1}ki_{p+1}\cdots i_r}^{j_1\cdots j_{q-1}kj_{q+1}\cdots j_s}$。

8.3.9 张量理论中的简记法

简记法的基本精神上文已有介绍，本节只补充几点。

在一个张量的指标中，如果同一个符号，既作为协变指标出现一次，又作为逆变指标出现一次，则称为哑指标。张量的一个指标，如果不是哑指标，则称为自由指标。哑指标表示求和，自由指标表示分量。值得一提的是，g_{ii} 中的指标 i，是自由指标，因此 g_{ii} 表示三个分量 g_{11}、g_{22} 和 g_{33}。

有些量符合简记法的运算规则，但是不符合张量的坐标变换规则，因此不是张量。表达式本身并不反映这种区别，看官们要自己做到心中有数。

记 $\partial^1 = \mathrm{d}x,\ \partial^2 = \mathrm{d}y,\ \partial^3 = \mathrm{d}z$，和 $\partial_1 = \dfrac{\partial}{\partial x},\ \partial_2 = \dfrac{\partial}{\partial y},\ \partial_3 = \dfrac{\partial}{\partial z}$。那么，全微分就可以记为 $\mathrm{d}w = \partial^i \partial_i w$。或者写成算符形式 $\mathrm{d} = \partial^i \partial_i$。

标量场不依赖于坐标系的选择，这一性质可表示为 $\mathrm{d} = \partial^i \partial_i = \partial^{i'} \partial_{i'}$，在前述例子中，$i$ 取 x, y, z，而 i' 取 r, θ, ϕ。

坐标变换关系 $\mathrm{d}r = \dfrac{\partial r}{\partial x}\mathrm{d}x + \dfrac{\partial r}{\partial y}\mathrm{d}y + \dfrac{\partial r}{\partial z}\mathrm{d}z$ 可以简写为 $\partial^{1'} = \partial_i^{1'} \partial^i$。整个逆变矢量坐标变换关系可写为 $\partial^{j'} = \partial_i^{j'} \partial^i$。

另一方面，$\dfrac{\partial w}{\partial r} = \dfrac{\partial x}{\partial r}\dfrac{\partial w}{\partial x} + \dfrac{\partial y}{\partial r}\dfrac{\partial w}{\partial y} + \dfrac{\partial z}{\partial r}\dfrac{\partial w}{\partial z}$ 可表示为 $\partial_{1'} = \partial_{1'}^i \partial_i$。整个协变矢量坐标变换关系可写为 $\partial_{j'} = \partial_{j'}^i \partial_i$。

坐标变换矩阵 A 和 B 之间的关系 $A^{-1} = B^T$可以写成 $\partial_i^{j'} \partial_{j'}^k = \delta_i^k$，其中 δ_i^k 是克罗奈克 (Kronecker) 符号，当 $i = k$ 时，其值为 1，否则为 0。这一关系还可写成 $\partial_{i'}^j \partial_j^{k'} = \delta_{i'}^{k'}$。

张量应该满足的坐标变换规则可写为 $T_{i'_1 \cdots i'_r}^{j'_1 \cdots j'_s} = \partial_{i'_1}^{i_1} \cdots \partial_{i'_r}^{i_r} \partial_{j_1}^{j'_1} \cdots \partial_{j_s}^{j'_s} T_{i_1 \cdots i_r}^{j_1 \cdots j_s}$。

有些文献记 $\partial_x = \dfrac{\partial}{\partial x},\ \partial_y = \dfrac{\partial}{\partial y},\ \partial_z = \dfrac{\partial}{\partial z}$，或者 $\partial_{x^i} = \dfrac{\partial}{\partial x^i}$。我们把后者进一步简化成了 ∂_i。注意 ∂_i 与 ∂_x 的联系与区别。

8.3.10 识别定理与张量场

识别定理：如果 n^{r+s}-元组 $T_{i_1 \cdots i_r}^{j_1 \cdots j_s}$ 对任意 r 个逆变矢量 $X^{i_1}, Y^{i_2}, \ldots, Z^{i_r}$ 和 s 个协变矢量 $X_{j_1}, Y_{j_2}, \ldots, Z_{j_s}$ 均能使得 $T_{i_1 \cdots i_r}^{j_1 \cdots j_s} X^{i_1} \cdots Z^{i_r} X_{j_1} \cdots Z_{j_s}$ 是标量，那么 $T_{i_1 \cdots i_r}^{j_1 \cdots j_s}$ 是 (r, s) 型张量。因为，在坐标变换下，

$$
\begin{aligned}
& T_{i'_1 \cdots i'_r}^{j'_1 \cdots j'_s} X^{i'_1} \cdots Z^{i'_r} X_{j'_1} \cdots Z_{j'_s} \\
= \ & T_{i'_1 \cdots i'_r}^{j'_1 \cdots j'_s} \partial_{i_1}^{i'_1} X^{i_1} \cdots \partial_{i_r}^{i'_r} Z^{i_r} \partial_{j'_1}^{j_1} X_{j_1} \cdots \partial_{j'_s}^{j_s} Z_{j_s} \\
= \ & T_{i'_1 \cdots i'_r}^{j'_1 \cdots j'_s} \partial_{i_1}^{i'_1} \cdots \partial_{i_r}^{i'_r} \partial_{j'_1}^{j_1} \cdots \partial_{j'_s}^{j_s} X^{i_1} \cdots Z^{i_r} X_{j_1} \cdots Z_{j_s} \\
= \ & T_{i_1 \cdots i_r}^{j_1 \cdots j_s} X^{i_1} \cdots Z^{i_r} X_{j_1} \cdots Z_{j_s}.
\end{aligned}
$$

由于式中的 $r + s$ 个矢量是任意的，所以最后两式相等意味着各项的系数相等，即 $T_{i'_1 \cdots i'_r}^{j'_1 \cdots j'_s} \partial_{i_1}^{i'_1} \cdots \partial_{i_r}^{i'_r} \partial_{j'_1}^{j_1} \cdots \partial_{j'_s}^{j_s} = T_{i_1 \cdots i_r}^{j_1 \cdots j_s}$，正符合张量的定义。 □

给流形上一定区域内所有的几何点，都指定相应的 n^{r+s}-元组 $T_{i_1 \cdots i_r}^{j_1 \cdots j_s}$，如果它们能够按识别定理的条件构成张量，则称在该区域上定义了张量场 $T_{i_1 \cdots i_r}^{j_1 \cdots j_s}$。

8.3.11 度规张量与黎曼几何

球面上两点之间最短的距离是过这两点的大圆的劣弧的弧长。可见距离、弧长对刻画几何性质很重要。并非所有的几何学都可以定义距离，但是，定义了距离的几何学有非常丰富的内容，而且也是进一步研究其他几何学的基础。

如果在一种几何中定义了两点之间的距离，那就是说距离是一种几何性质，与选择哪个坐标系来表达没有关系。也就是说，距离是标量。

欧氏空间中，曲线的一段微元的长度是 $\mathrm{d}s = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}$。然而，在相对论中我们要用到闵科夫斯基空间，简称闵氏空间，其中曲线微元的“长度” $\mathrm{d}s$ 满足 $\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$。

一般地，定义曲线微元的长度 $\mathrm{d}s$ 满足 $\mathrm{d}s^2 = g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j$。因为 $\mathrm{d}s^2$ 是标量，所以按识别定理，g_{ij} 是二协张量，称为度规张量。

如果度规张量 g_{ij} 满足 $g_{ij} = g_{ji}$，则称为黎曼度规。定义了黎曼度规的流形，称为黎曼流形。研究黎曼流形的几何学，称为黎曼几何。

8.4 求导法则与积分法

8.4.1 函数的有界性、单调性、连续性和可微性

如果函数值不大于某一给定值，则称该函数有上界。如果函数值不小于某一给定值，则称该函数有下界。既有上界，又有下界，则称有界。

当自变量的值逐步增大时，函数值逐步增大，称函数单调增。当自变量的值逐步增大时，函数值逐步减小，称函数单调减。

如果函数在某点的值等于在该点的极限，则称该函数在该点连续。如果函数在一个区域的所有点上都连续，则称该函数在该区域上连续。

如果函数在某点的微分（或全微分）可定义，则称该函数在该点可微。如果函数在一个区域的所有点上都可微，则称该函数在该区域上可微。

对于几何或物理上的问题，如果只用一个函数不足以完全刻画和分析时，可以用多个函数相互配合来做。这就是流形的基本思想。举个极端的例子。水平直线可以用函数 $y = c$ 来刻画，可是竖直直线 $x = c$ 就无法用函数来表示。这时可以使用参数方程 $\ell(t): \quad x = c, \quad y = t$。而水平直线显然有参数方程 $\ell(t): \quad x = t, \quad y = c$。这样，直线，无论是水平的还是竖直的，都有了统一的处理方法。

8.4.2 中值定理

中值定理：如果函数 $f(x)$ 在区间 $[a, b]$ 上连续，而 d 在 $f(a)$ 和 $f(b)$ 之间，那么，一定存在 $c \in [a, b]$ 使得 $f(c) = d$。

中值定理与欧氏平面的性质有关。如果在点 $e \in (a, b)$ 上，$f(e^-) \to -\infty$ 而 $f(e^+) \to \infty$，由于欧氏平面没有无穷远点，直线是两端“开”的，所以这样的函数 $f(x)$ 被认为是“断”的，不连续的。而在射影平面上，直线两端是用无穷远点“粘合”起来的。整个直线与圆同胚。只要定义 $f(e) = \infty$，刚才的函数

$f(x)$ 就是连续的。所以，在射影平面上，从起点 $(a, f(a))$ 出发，绕道 (e, ∞)，到达终点 $(b, f(b))$，这段连续曲线可以与直线 $y = d$ 不相交。

欧氏平面与射影平面，以及与复平面，有许多有趣的差异。欧氏平面同胚于无边缘的圆盘，复平面同胚于球面，两者都是双面的，而射影平面是单面的。一张矩形的纸带，两端粘合，就是环带，有上下两个边缘。让两个边缘都收缩成一点，就同胚于球面。还是这张纸带，把一端扭转 180 度，再与另一端粘合，就是茂比乌斯带。它只有一个边缘。让这个边缘收缩为一点，就同胚于射影平面。

8.4.3 求导的倒数法则

因为 $\dfrac{1}{f(x+\Delta x)} - \dfrac{1}{f(x)} = \dfrac{f(x) - f(x+\Delta x)}{f(x+\Delta x)f(x)}$，按导数的定义易见 $\left(\dfrac{1}{f(x)}\right)' = -\dfrac{f(x)'}{f(x)^2}$。这个称为倒数法则。

例如，设 $y = \dfrac{1}{x}$，有 $\dfrac{1}{x+\Delta x} - \dfrac{1}{x} = -\dfrac{\Delta x}{(x+\Delta x)x}$，易见 $y' = -\dfrac{1}{x^2}$。

8.4.4 求导的幂法则

由二项式定理，$(x+\Delta x)^n = x^n + nx^{n-1}\Delta x + \dfrac{n(n-1)}{2}x^{n-2}(\Delta x)^2 + \cdots$，由导数的定义立得 $(x^n)' = nx^{n-1}$。

另一方面，由倒数法则，$\left(\dfrac{1}{x^n}\right)' = -\dfrac{nx^{n-1}}{x^{2n}} = -n\dfrac{1}{x^{n+1}}$。

实际上，对任何实数 a，均有 $(x^a)' = ax^{a-1}$。这称为幂法则。（证明思路见以下的反函数法则）。

8.4.5 求导与积分的线性法则

给定两个可微函数 $f(x)$ 和 $g(x)$，有 $(f+g)' = f' + g'$。

给定可微函数 $f(x)$ 和常数 c，有 $(cf)' = cf'$。

给定函数 $f(x)$ 和 $g(x)$，设 a 和 b 是不同时为零的两个实数，那么，称 $af(x)+bg(x)$ 为 f 和 g 的线性组合。有 $(af+bg)' = af' + bg'$。所以，求导的线性法则可以表述为：函数的线性组合的导数等于函数的导数的同一线性组合。

与此对应的是积分的线性法则

$$\int_{x_0}^{x_1} (af(x) + bg(x))\,\mathrm{d}x = a\int_{x_0}^{x_1} f(x)\,\mathrm{d}x + b\int_{x_0}^{x_1} g(x)\,\mathrm{d}x.$$

8.4.6 莱布尼兹乘法法则与分部积分法

给定函数 $f(x)$ 和 $g(x)$，因为，

$$f(x+\Delta x)g(x+\Delta x)$$

$$
\begin{aligned}
&\approx (f(x)+f'(x)\Delta x)(g(x)+g'(x)\Delta x)\\
&= f(x)g(x)+(f'(x)g(x)+f(x)g'(x))\Delta x+f'(x)g'(x)(\Delta x)^2
\end{aligned}
$$

按导数的定义，显见 $(fg)'=f'g+fg'$。这称为（莱布尼兹）乘法法则。

对此式两边取积分，有

$$
\int_{x_0}^{x_1}(f(x)g(x))'\,\mathrm{d}x=\int_{x_0}^{x_1}f'(x)g(x)\,\mathrm{d}x+\int_{x_0}^{x_1}f(x)g'(x)\,\mathrm{d}x,
$$

即

$$
\int_{x_0}^{x_1}f(x)g'(x)\,\mathrm{d}x=f(x)g(x)\Big|_{x_0}^{x_1}-\int_{x_0}^{x_1}f'(x)g(x)\,\mathrm{d}x.
$$

这称为分部积分法。如果该式左边的积分不易直接求，而右边的积分易求，那么分部积分法就是个很好用的技巧，尤其是涉及到指数函数、对数函数、三角函数、双曲三角函数等超越函数的时候。

8.4.7 求导的除法法则

结合倒数法则与乘法法则，有 $(\frac{f}{g})'=(f\frac{1}{g})'=f'\frac{1}{g}-f\frac{g'}{g^2}=\frac{f'g-fg'}{g^2}$。这称为除法法则。

另一种证法是：记 $h=\frac{f}{g}$，则 $f=hg$，有 $f'=h'g+hg'$，即 $f'g=h'g^2+fg'$。 □

8.4.8 求导的链式法则与积分的换元法

对于函数 $y=f(x)$，在适当条件下，当 $\Delta x\to 0$ 时，有 $\Delta y\to 0$。这样，对于函数 $z=g(y)=g(f(x))=h(x)$，有

$$
\begin{aligned}
h'(x) &= \frac{\mathrm{d}z}{\mathrm{d}x}=\lim_{\Delta x\to 0}\frac{\Delta z}{\Delta x}\\
&= \lim_{\Delta x\to 0}\frac{\Delta z}{\Delta y}\frac{\Delta y}{\Delta x}\\
&= \lim_{\Delta y\to 0}\frac{\Delta z}{\Delta y}\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}\\
&= g'(y)f'(x)\\
&= g'(f(x))f'(x)
\end{aligned}
$$

这称为求导的链式法则。

该法则又可写成 $\frac{\mathrm{d}z}{\mathrm{d}x}=\frac{\mathrm{d}z}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x}$。形式上仿佛两个 $\mathrm{d}y$ 可以相消一样。这是个在一元情况下的巧合。

多元情况下，$\partial_{i'}=\partial^i_{i'}\partial_i$ 称为求偏导的链式法则。注意这里的两个坐标系不必同维。比如，单位球面上的标量场 $w=f(x,y,z)=g(\theta,\phi)$。有

$x = x(\theta, \phi), \quad y = y(\theta, \phi), \quad z = z(\theta, \phi)$，进而

$$\begin{cases} \dfrac{\partial g}{\partial \theta} = \dfrac{\partial x}{\partial \theta}\dfrac{\partial f}{\partial x} + \dfrac{\partial y}{\partial \theta}\dfrac{\partial f}{\partial y} + \dfrac{\partial z}{\partial \theta}\dfrac{\partial f}{\partial z} \\ \dfrac{\partial g}{\partial \phi} = \dfrac{\partial x}{\partial \phi}\dfrac{\partial f}{\partial x} + \dfrac{\partial y}{\partial \phi}\dfrac{\partial f}{\partial y} + \dfrac{\partial z}{\partial \phi}\dfrac{\partial f}{\partial z} \end{cases}.$$

求积分时，由于 $z = g(y) = h(x)$，即 $\mathrm{d}z = g'(y)\,\mathrm{d}y = h'(x)\,\mathrm{d}x$，所以有

$$\begin{aligned} \int_{x_0}^{x_1} h'(x)\,\mathrm{d}x &= \int_{x_0}^{x_1} g'(f(x))f'(x)\,\mathrm{d}x \\ &= \int_{x_0}^{x_1} g'(f(x))\,\mathrm{d}(f(x)) \\ &= \int_{f(x_0)}^{f(x_1)} g'(y)\,\mathrm{d}y \end{aligned}$$

其中最后一步是代入 $y = f(x)$。注意，积分上下限、被积函数和积分微元都要改变。这种操作称为积分的换元法。

实际上该积分还可以换元成 z，即该积分等于 $\displaystyle\int_{z_0}^{z_1} \mathrm{d}z = z_1 - z_0 = g(y_1) - g(y_0) = h(x_1) - h(x_0)$。

8.4.9 求导的反函数法则

如果函数 $y = f(x)$ 是双射，那么反函数 $x = f^{-1}(y)$ 存在。由于 $\mathrm{d}y = f'(x)\,\mathrm{d}x$，所以 $\mathrm{d}x = \dfrac{1}{f'(f^{-1}(y))}\mathrm{d}y$，即 $(f^{-1}(y))' = \dfrac{1}{f'(f^{-1}(y))}$。这称为反函数法则。

例如，$y = x^3$ 在 $x \neq 0$ 处有反函数。那么，$\mathrm{d}y = 3x^2\,\mathrm{d}x$，进而 $\mathrm{d}x = \dfrac{1}{3x^2}\,\mathrm{d}y = \dfrac{1}{3y^{2/3}}\,\mathrm{d}y$。所以，$\dfrac{\mathrm{d}x}{\mathrm{d}y} = (y^{1/3})' = \dfrac{1}{3}y^{-2/3}$。同理易证，对任意有理数 a，因而对任意实数 a，有 $(x^a)' = ax^{a-1}$。

8.4.10 求导的隐函数法则

方程 $F(x, y) = 0$ 可以用来表达 $y = f(x)$ 的关系，称为隐函数。记 $z = F(x, y) = F(x, f(x))$，有 $\dfrac{\mathrm{d}z}{\mathrm{d}x} = \dfrac{\partial F}{\partial x} + \dfrac{\partial F}{\partial y}\dfrac{\mathrm{d}y}{\mathrm{d}x}$。因为 z 恒等于零，所以 $\dfrac{\mathrm{d}z}{\mathrm{d}x} = 0$，即 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = -\dfrac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}$。这称为隐函数法则。

例如，令 $F(x, y) = x^2 + y^2 - 1$，有 $\dfrac{\partial F}{\partial x} = 2x, \quad \dfrac{\partial F}{\partial y} = 2y$，因而 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = -\dfrac{x}{y} = -\dfrac{x}{\sqrt{1 - x^2}}$。这充分利用了问题中的对称性，比直接从 $y = \sqrt{1 - x^2}$ 求导在计算上有所简化。

下面我们以三元为例，说明多元隐函数法则。在适当条件下，方程组

$$\begin{cases} F(u,v,w,x,y,z) = 0 \\ G(u,v,w,x,y,z) = 0 \\ H(u,v,w,x,y,z) = 0 \end{cases}.$$

可以用来表达 $x = x(u,v,w), \quad y = y(u,v,w), \quad z = z(u,v,w)$ 的多元隐函数关系。因为，六个变量三个约束方程，一般来说，只有三个变量是自由的。不妨选为 u,v,w，那么另外三个变量 x,y,z 就是 u,v,w 的函数。

上述三个方程左右两边都对 u 求偏导，则有

$$\begin{cases} \dfrac{\partial F}{\partial u} + \dfrac{\partial F}{\partial x}\dfrac{\partial x}{\partial u} + \dfrac{\partial F}{\partial y}\dfrac{\partial y}{\partial u} + \dfrac{\partial F}{\partial z}\dfrac{\partial z}{\partial u} = 0 \\ \dfrac{\partial G}{\partial u} + \dfrac{\partial G}{\partial x}\dfrac{\partial x}{\partial u} + \dfrac{\partial G}{\partial y}\dfrac{\partial y}{\partial u} + \dfrac{\partial G}{\partial z}\dfrac{\partial z}{\partial u} = 0 \\ \dfrac{\partial H}{\partial u} + \dfrac{\partial H}{\partial x}\dfrac{\partial x}{\partial u} + \dfrac{\partial H}{\partial y}\dfrac{\partial y}{\partial u} + \dfrac{\partial H}{\partial z}\dfrac{\partial z}{\partial u} = 0 \end{cases}.$$

在适当条件下可以解出 $\dfrac{\partial x}{\partial u}, \quad \dfrac{\partial y}{\partial u}, \quad \dfrac{\partial z}{\partial u}$。同理可解出另外六个偏导数。也就是说，$\partial^j_{i'}$ 的九个分量均可用 F,G,H 的偏导数来表达。这实际上是九个线性方程九个未知数的问题。

如果没有方程 H，六个变量两个约束方程，那就有四个变量是独立的，不妨选为 u,v,w,x。则两个方程对 u 求偏导，得

$$\begin{cases} \dfrac{\partial F}{\partial u} + \dfrac{\partial F}{\partial y}\dfrac{\partial y}{\partial u} + \dfrac{\partial F}{\partial z}\dfrac{\partial z}{\partial u} = 0 \\ \dfrac{\partial G}{\partial u} + \dfrac{\partial G}{\partial y}\dfrac{\partial y}{\partial u} + \dfrac{\partial G}{\partial z}\dfrac{\partial z}{\partial u} = 0 \end{cases}.$$

在适当条件下也可以解出 $\dfrac{\partial y}{\partial u}, \quad \dfrac{\partial z}{\partial u}$。同理可求 y 和 z 对 v,w,x 的偏导。这是八个线性方程八个未知数的问题。

如果只有一个方程 F，那就有五个变量是独立的，不妨选为 u,v,w,x,y，方程两边对 u 求偏导，有

$$\frac{\partial F}{\partial u} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial u} = 0.$$

易见 $\dfrac{\partial z}{\partial u} = -\dfrac{\dfrac{\partial F}{\partial u}}{\dfrac{\partial F}{\partial z}}$。其他四个偏导数类似。这是五个线性方程五个未知数的问题。

反之，如果是六个变量四个约束方程 F,G,H,I，那就只有两个自由变量，不妨选为 u,v。四个方程都对 u 求偏导，有

$$\begin{cases} \dfrac{\partial F}{\partial u} + \dfrac{\partial F}{\partial w}\dfrac{\partial w}{\partial u} + \dfrac{\partial F}{\partial x}\dfrac{\partial x}{\partial u} + \dfrac{\partial F}{\partial y}\dfrac{\partial y}{\partial u} + \dfrac{\partial F}{\partial z}\dfrac{\partial z}{\partial u} = 0 \\ \cdots \text{（另外三个方程类似，兹不赘述。）} \end{cases}.$$

可以解出 $\dfrac{\partial w}{\partial u}$， $\dfrac{\partial x}{\partial u}$， $\dfrac{\partial y}{\partial u}$， $\dfrac{\partial z}{\partial u}$。同理可解出另外四个偏导数。这是八个线性方程八个未知数的问题。

如果是六个变量五个约束方程 F, G, H, I, J，那就只有一个自由变量，不妨选为 u。那么，五个方程都对 u 求导数，有

$$\begin{cases} \dfrac{\partial F}{\partial u} + \dfrac{\partial F}{\partial v}\dfrac{\mathrm{d}v}{\mathrm{d}u} + \dfrac{\partial F}{\partial w}\dfrac{\mathrm{d}w}{\mathrm{d}u} + \dfrac{\partial F}{\partial x}\dfrac{\mathrm{d}x}{\mathrm{d}u} + \dfrac{\partial F}{\partial y}\dfrac{\mathrm{d}y}{\mathrm{d}u} + \dfrac{\partial F}{\partial z}\dfrac{\mathrm{d}z}{\mathrm{d}u} = 0 \\ \cdots \text{（另外四个方程类似，兹不赘述。）} \end{cases}.$$

可以解出五个导数 v', w', x', y', z'。这是五个线性方程五个未知数的问题。

这些就是多元隐函数法则的基本思想。

8.5 射影几何的初步概念

8.5.1 曲线的参数方程与弧长

中学已经学过，平面上圆的参数方程为 $\mathcal{C}(\theta):\quad x = x_0 + r\cos\theta,\quad y = y_0 + r\sin\theta$。空间中直线的参数方程是 $\ell(t):\quad x = x_0 + t\cos\alpha,\quad y = y_0 + t\cos\beta,\quad z = z_0 + t\cos\gamma$，其中三个余弦是直线的方向余弦，满足 $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$。易见，参数方程的中心思想是，从 $\mathbf{R}$ 上的一段区间，映射到一维流形上的一段曲线。这样，参数就可以被看做是一维流形上的曲线坐标，称为参数坐标。进而，参数方程就是坐标变换关系。

以下除非特别说明，我们将用 τ、θ、ζ 等希腊字母来表示参数坐标。如果 $y = f(\tau)$，那么我们将把 $\dfrac{\mathrm{d}y}{\mathrm{d}\tau} = f'$ 记为 $\dot{y}$，或者 $\dot{f}$。

设曲线 $\Gamma(\tau)$ 的参数方程是 $x^i = x^i(\tau)$。那么，有 $\mathrm{d}x^i = \dot{x}^i\,\mathrm{d}\tau$，曲线微元的弧长为 $\mathrm{d}s^2 = g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j = g_{ij}\dot{x}^i\dot{x}^j\,\mathrm{d}\tau^2$，所以，一段曲线的长度是 $s = \displaystyle\int_{\tau_0}^{\tau_1}\sqrt{g_{ij}\dot{x}^i\dot{x}^j}\,\mathrm{d}\tau$。

对上述平面上的圆，有 $\dot{x} = -r\sin\theta,\quad \dot{y} = r\cos\theta$，所以，$\mathrm{d}s = r\,\mathrm{d}\theta$，圆的周长为 $s = \displaystyle\int_0^{2\pi} r\,\mathrm{d}\theta = 2\pi r$。

对上述空间中的直线，有 $\dot{x} = \cos\alpha,\quad \dot{y} = \cos\beta,\quad \dot{z} = \cos\gamma$，而 $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$，所以，$\mathrm{d}s = \mathrm{d}\tau$，一段直线的长度为 $s = \displaystyle\int_{\tau_0}^{\tau_1}\mathrm{d}\tau = \tau_1 - \tau_0$。

8.5.2 曲线的切线与曲面的切面

给定三维空间中一个标量场 $w = w(P)$，在直角坐标系下的表示为 $w = f(x,y,z)$。它的等值面为 $f(x,y,z) = c$。设该等值面上有一条曲线 $\Gamma(\tau):\quad x = x(\tau),\quad y = y(\tau),\quad z = z(\tau)$。把等值面方程对 τ 求导，得

$$\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\tau} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}\tau} + \frac{\partial f}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}\tau} = 0.$$

即 $\dot{x}^i\partial_i=0$。其中 $\dot{x}^i$ 是曲线的性质，表示曲线在点 $P(x,y,z)$ 处的切线；而 ∂_i 是曲面的性质，表示曲面在点 $P(x,y,z)$ 处的切面。方程 $\dot{x}^i\partial_i=0$ 表示“曲面上曲线的切线在该曲面的切面上”。

8.5.3 齐次坐标与射影平面

如果有另一个标量场 $u=u(P)$ 在直角坐标系下的表示为 $u=\rho(x,y,z)$，而且在点 $P(x,y,z)$ 处不为零，即 $\rho(x,y,z)\neq 0$，那么在上节所述情形下，$\rho\dot{x}^i$ 表示相同的切线，$\rho\partial_i$ 表示相同的切面。

当 $\rho(x_P,y_P,z_P)\neq 0$ 表示任意非零值时，称 $\rho\dot{x}^i$ 为该切线的齐次坐标，称 $\rho\partial_i$ 为该切面的齐次坐标。而对于一个给定的具体值 $\rho_0\neq 0$，称 $\rho_0\dot{x}^i$ 和 $\rho_0\partial_i$ 为相应的解析坐标。有时也说，$\rho_0\dot{x}^i$ 和 $\rho_1\dot{x}^i$，其中 $\rho_0\neq\rho_1$，表示两个不同的解析点，但它们是同一个几何点的两种解析表示。

在点 P 的邻域内，以点 P 为球心做一个微小的球面。上述曲面在点 P 处的切面与该球面交于一个大圆，而上述曲线在点 P 处的切线是该大圆的一条直径。该切线与该球面的两个交点是该直径的两个端点。

取 $\rho=-1$，两个端点交换位置，但是切线仍旧是原来的切线。也就是说，把两个端点理解为“粘合”在一起成为一个“点”，那么，“点”与切线就一一对应。整个球面在“对径粘合”之后，就称为射影平面。射影平面上的“点”与粘合前的球面的直径一一对应，而“直线”则与大圆一一对应。

易见，射影平面上的点有齐次坐标 $\rho\dot{x}^i$，直线有齐次坐标 $\rho\partial_i$。某点在某直线上，或者该直线过该点，统一地称为该点与该直线相接合，表达式为 $\dot{x}^i\partial_i=0$。

8.5.4 点列与线束

当曲面上一个动点沿光滑闭合曲线巡行一周时，动点所在处的切线也随着转动，并最终返回原位。这可以进一步对应到射影平面上点的移动。比如，球面上动点绕赤道走一圈，对应于射影平面上的点沿直线走两圈。

曲面的几何性质可以通过其上的光滑闭合曲线来研究，因而也就可以通过射影平面上的“点的序列”，简称点列，来研究。

同理，当曲面上一个动点沿光滑闭合曲线巡行一周时，动点所在处的切面也随着转动，并最终返回原位。这可以进一步对应到射影平面上直线的移动。比如，球面上动点绕赤道走一圈，对应于射影平面上的直线“横着”扫过平面两次。这样，曲面的几何性质也可以通过射影平面上的“直线的序列”，简称线束，来研究。

8.5.5 曲面的球面映射法与高斯曲率

高斯发明的球面映射法与上节的方法有两点不同。首先，高斯使用单位球面，而不是射影平面。其次，光滑曲面上任何一点的足够微小的邻域，都有正反两面。沿曲面上给定点处的切面的法线，高斯给它指定一个正方向，并使用单位球面的平行于法线正方向的有向直径的前部端点，作为曲面上给定点在球面上的像点，由此把曲面映射到球面上。

因为这样的球面未做“对径粘合”，所以，很多时候更加直观。比如，如果所述曲面就是球面，那么球面上动点绕赤道走一圈，该动点的高斯球面像也沿大圆走一圈。

给定光滑曲面上一点，在其微小邻域内做一个简单回路，如果我们设想用右手握住曲面在该点的法线，竖起拇指，使拇指所指方向为法线正方向，那么，取四指所指向的回路方向为回路的正方向。记回路所包围的面积为 S。如果动点沿回路正方向巡回一周，称它所围的有向面积为 S；如果沿相反方向巡回一周，称它所围的有向面积为 $-S$。

上段所述情况下，该回路的高斯球面像是单位球面上的一个回路。记其有向面积为 T。像回路的有向面积与原回路的有向面积的比，在原回路趋于给定点时的极限，称为高斯曲率，也称主曲率，记为 $K = \lim\limits_{|S|\to 0} \dfrac{T}{S}$。

有趣的是，简单回路的高斯球面像不一定保持原来的方向。如果能保持，称曲面在该点是正曲率的；如果变成了相反方向，称曲面在该点是负曲率的。这两种情况下，球面映射在该点的充分小邻域内是双射。

如果高斯曲率为零，那么，球面映射在该点邻域内就不是双射。比如，整个平面的高斯球面像就是一个点。

这就意味着简单回路的高斯球面像也不一定是简单回路。比如，轮胎上离车轴远的外缘为正曲率，离车轴近的内缘为负曲率。两者之间的零曲率临界区是两个圆，一边一个，不妨称之为临界圆。在临界圆上某点的邻域内，做一个简单回路，那么，该回路的球面像类似于数字 8，当动点在正曲率区域时，其像做成一个保持原方向的回路；当动点在负曲率区域时，其像做成一个反方向的回路。原简单回路与临界圆有两个交点，它们的球面像重合，就是 8 字中间那个点。动点沿原简单回路巡行一周，像点就在单位球面上“写一个 8 字”。

8.5.6 自然对射

对球面来讲，给定一条直径，存在且只存在一个大圆，其所在平面与给定直径正交；反之，给定一个大圆，存在且只存在一条直径，与大圆所在的平面正交。称这个从直径到相应的大圆，或者从大圆到相应的直径的一一映射为自然对射。

按射影平面的对径粘合球面模型，如果一个点的齐次坐标是 (x^1, x^2, x^3)，

那么，这三个数与相应的球面直径的方向余弦成正比；如果一条直线的齐次坐标是 (ξ_1, ξ_2, ξ_3)，那么，这三个数与相应的大圆所在平面的法线的方向余弦成正比。显然，自然对射下，直径的方向恰好就是对应的平面的法线方向。因此，自然对射保持点与对应的直线的齐次坐标不变。

8.5.7 射影变换与对偶原理

每种几何学都有自己的主变换群。射影几何的主变换群是射影变换群，它保持点与直线的接合关系不变。也就是说，如果点 x^i 与直线 ξ_i 相接合，即有 $x^i\xi_i = 0$，而射影变换把 x^i 变到 y^i，把 ξ_i 变到 ζ_i，那么，一定有 $y^i\zeta_i = 0$。

射影几何的所有命题都是以点与直线的接合关系为基础的。在自然对射下，任何一个接合关系 $x^i\xi_i = 0$，都可以做重新解释：把点的坐标 x^i 解释为自然对射后的直线的坐标，而把直线的坐标 ξ_i 解释为自然对射后的点的坐标。这样，射影几何中的任何命题都可以通过自然对射变成一个新的命题，称为原命题的对偶命题。比如，“存在且只存在一条直线与给定的相异两点接合”，与“存在且只存在一个点与给定的相异两直线接合”就是相互对偶的命题。

有时对偶命题恰与原命题一致，称这样的命题为自对偶命题。

8.6 外积与霍奇星

8.6.1 射影几何中三点共线的条件

给定标量场 $w = w(P)$，在某个曲线坐标系下有全微分 $\mathrm{d}w = \partial^i\partial_i w$。设在该标量场过点 P 的等值面上，有三条过点 P 的光滑曲线 α, β, γ，它们在点 P 处的切线彼此不同。但是，三条切线都在等值面上点 P 处的切面上。即：$\dot{\alpha}^i\partial_i = 0,\quad \dot{\beta}^i\partial_i = 0,\quad \dot{\gamma}^i\partial_i = 0$，并且 $\dot{\alpha}^i$, $\dot{\beta}^i$, $\dot{\gamma}^i$ 之间，存在一个约束关系。因为两条切线足以确定切面，第三条切线既然要在切面上，那就不能是任意的了。这就是射影几何中三点共线的条件。

求此条件的思路是这样的。∂_i 的三个分量不全为零，不失一般性，取 $\partial_1 \neq 0$，记 $a = \dfrac{\partial_2}{\partial_1},\quad b = \dfrac{\partial_3}{\partial_1}$。解二元一次方程组

$$\begin{cases} \dot{\alpha}^1 + a\dot{\alpha}^2 + b\dot{\alpha}^3 = 0 \\ \dot{\beta}^1 + a\dot{\beta}^2 + c\dot{\beta}^3 = 0 \end{cases}.$$

再把解出的 a, b 代入 $\dot{\gamma}^1 + a\dot{\gamma}^2 + b\dot{\gamma}^3 = 0$，就得到三点共线的条件。

学过线性代数的，可以把 $\dot{\alpha}^i\partial_i = 0,\quad \dot{\beta}^i\partial_i = 0,\quad \dot{\gamma}^i\partial_i = 0$ 看做是关于三个

未知数 ∂_i 的齐次线性方程，它有非零解的充要条件是系数行列式为零，即

$$\begin{vmatrix} \dot\alpha^1 & \dot\alpha^2 & \dot\alpha^3 \\ \dot\beta^1 & \dot\beta^2 & \dot\beta^3 \\ \dot\gamma^1 & \dot\gamma^2 & \dot\gamma^3 \end{vmatrix} = 0.$$

8.6.2 射影直线上的基底

由于

$$\begin{vmatrix} \dot\alpha^1 & \dot\beta^1 & \dot\gamma^1 \\ \dot\alpha^2 & \dot\beta^2 & \dot\gamma^2 \\ \dot\alpha^3 & \dot\beta^3 & \dot\gamma^3 \end{vmatrix} = \begin{vmatrix} \dot\alpha^1 & \dot\alpha^2 & \dot\alpha^3 \\ \dot\beta^1 & \dot\beta^2 & \dot\beta^3 \\ \dot\gamma^1 & \dot\gamma^2 & \dot\gamma^3 \end{vmatrix},$$

如果后者为零，那么前者也为零，因此，三点共线的充要条件是，存在不全为零的三个数 a, b, c，使得三个方程 $a\dot\alpha^i + b\dot\beta^i + c\dot\gamma^i = 0$ 成立，即：三个逆变矢量 $\dot\alpha^i$, $\dot\beta^i$, $\dot\gamma^i$ 线性相关。

给定两点 $\dot\alpha^i$ 和 $\dot\beta^i$ 相异，那么 $c \neq 0$。因为，假设 $c = 0$，那么 a, b 不同时为零，并且 $a\dot\alpha^i + b\dot\beta^i = 0$。不失一般性，取 $a \neq 0$，有 $\dot\alpha^i = -\dfrac{b}{a}\dot\beta^i$，即 $\dot\alpha^i$ 和 $\dot\beta^i$ 是同一个几何点，与该两点相异矛盾。

既然 $c \neq 0$，那么 $\dot\gamma^i = -\dfrac{a}{c}\dot\alpha^i - \dfrac{b}{c}\dot\beta^i$。取 $a \neq 0, \quad b = 0$，那么点 $\dot\gamma^i$ 就是点 $\dot\alpha^i$；取 $a = 0, \quad b \neq 0$，那么点 $\dot\gamma^i$ 就是点 $\dot\beta^i$。所以，在点 $\dot\alpha^i$ 和点 $\dot\beta^i$ 相异的前提下，上述形式的 $\dot\gamma^i$ 可以表示连线上的任意一点。也就是说，相异两点 $\dot\alpha^i$ 和 $\dot\beta^i$ 连线上的任意一点 $\dot\gamma^i$，均可写成 $\dot\gamma^i = \lambda\dot\alpha^i + \mu\dot\beta^i$ 的形式，其中 λ, μ 不同时为零。

在此图景下，称相异两点 $\dot\alpha^i$ 和 $\dot\beta^i$ 的连线是一个一维射影空间，称 $\dot\alpha^i$ 和 $\dot\beta^i$ 为该空间中的基矢，两者共同构成该空间的一个基底。λ 和 μ 是矢量 $\dot\gamma^i$ 在相应基矢上的分量，两者共同构成该空间中的点在给定基底下的齐次坐标 (λ, μ)。

8.6.3 外积

在上节所述情境下，记 $\vec{e}_1 = (\dot\alpha^i), \quad \vec{e}_2 = (\dot\beta^i), \quad \vec{r} = (\dot\gamma^i)$，其中 $(\dot\alpha^i)$ 是三元组 $(\dot\alpha^1, \dot\alpha^2, \dot\alpha^3)$ 的简写。并且记 $\lambda^1 = \lambda, \quad \lambda^2 = \mu$，那么 $\dot\gamma^i = \lambda\dot\alpha^i + \mu\dot\beta^i$ 就可以改写成 $\vec{r} = \lambda^1\vec{e}_1 + \lambda^2\vec{e}_2$，简写为 $\vec{r} = \lambda^\nu\vec{e}_\nu$。注意，指标 i, j 等取 $1, 2, 3$，而指标 ν 只取 $1, 2$。

射影直线上给定基底 $\vec{e}_\nu$，规定基矢之间的外积为 $\vec{e}_1 \wedge \vec{e}_1 = \vec{e}_2 \wedge \vec{e}_2 = 0, \quad \vec{e}_2 \wedge \vec{e}_1 = -\vec{e}_1 \wedge \vec{e}_2$，而 $\vec{e}_1 \wedge \vec{e}_2$ 没有进一步的定义，运算中保持这个形式即可。或许正因如此，外积被认为是一种“形式积”，指“形式上的乘积”。

那么，任意两点 $\vec{a} = a^\nu\vec{e}_\nu$ 和 $\vec{b} = b^\nu\vec{e}_\nu$ 的外积就是

$$\vec{a} \wedge \vec{b} = (a^1\vec{e}_1 + a^2\vec{e}_2) \wedge (b^1\vec{e}_1 + b^2\vec{e}_2)$$

$$
\begin{aligned}
&= (a^1b^2 - a^2b^1)\vec{e}_1 \wedge \vec{e}_2 \\
&= \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix} \vec{e}_1 \wedge \vec{e}_2
\end{aligned}
$$

显见，外积满足反交换律：$\vec{a} \wedge \vec{b} = -\vec{b} \wedge \vec{a}$。

射影直线上，三个或更多个点的外积为零。比如，$\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_1 = -(\vec{e}_1 \wedge \vec{e}_1) \wedge \vec{e}_2 = 0$。

习惯了张量分量的看官可能想知道 $\vec{e}_1 \wedge \vec{e}_2$ 的分量是什么，回答是：它没有分量，它只是“形式上的乘积”，就这个样子放着就可以了。

可是，如果一种运算不能最终得出数值，那又有什么用呢？回答是：“最终”我们将通过霍奇星运算来得出数值。

但是现在，我们先看看射影平面上的情况。

8.6.4 射影空间中的平面

四维空间中有标量场 $w = w(P)$，其全微分为 $\mathrm{d}w = \partial^i \partial_i w$，这里 i 取 $1, 2, 3, 4$。过点 P 的曲线 $\gamma(\tau)$ 在点 P 处的切线在该标量场的等值超曲面在点 P 处的切超平面上的条件是 $\dot{\gamma}^i \partial_i = 0$。四个切线都在该切超平面上的条件是 $\partial_i w$ 有非零解，即行列式 $|\dot{\alpha}, \dot{\beta}, \dot{\gamma}, \dot{\zeta}| = 0$。这就是射影空间中四点共面的条件。

具体来说，三维射影空间中，任意点 γ 的齐次坐标由一个非零的任意系数 ρ 和四个不全为零的数 $\gamma^1, \gamma^2, \gamma^3, \gamma^4$ 所构成，记为 $\rho(\gamma^1, \gamma^2, \gamma^3, \gamma^4)$。该空间中，存在且只存在一个平面与给定的互异三点 $\alpha^i, \beta^i, \gamma^i$ 接合。该平面上任意一点 ζ^i 均可表成给定的互异三点的线性组合，即存在不全为零的三个数 a, b, c，使得 $\zeta^i = a\alpha^i + b\beta^i + c\gamma^i$。这与四点共面的条件是一致的。

在此情境下，称给定的互异三点为三个基矢，它们共同构成该平面的基底。记 $\vec{e}_1 = (\alpha^i), \quad \vec{e}_2 = (\beta^i), \quad \vec{e}_3 = (\gamma^i), \quad \vec{r} = (\zeta^i)$，并且记 $\lambda^1 = a, \quad \lambda^2 = b, \quad \lambda^3 = c$，则上式可写为 $\vec{r} = \lambda^1\vec{e}_1 + \lambda^2\vec{e}_2 + \lambda^3\vec{e}_3$，简写为 $\vec{r} = \lambda^\nu \vec{e}_\nu$。注意，这里的 ν 取 $1, 2, 3$，而 i 取 $1, 2, 3, 4$。

规定上述基矢之间的外积为 $\vec{e}_1 \wedge \vec{e}_1 = \vec{e}_2 \wedge \vec{e}_2 = \vec{e}_3 \wedge \vec{e}_3 = 0, \quad \vec{e}_2 \wedge \vec{e}_1 = -\vec{e}_1 \wedge \vec{e}_2, \quad \vec{e}_3 \wedge \vec{e}_2 = -\vec{e}_2 \wedge \vec{e}_3, \quad \vec{e}_3 \wedge \vec{e}_1 = -\vec{e}_1 \wedge \vec{e}_3$，而 $\vec{e}_1 \wedge \vec{e}_2, \quad \vec{e}_2 \wedge \vec{e}_3, \quad \vec{e}_1 \wedge \vec{e}_3$ 没有进一步的定义。

这样，任意两个点 $\vec{a} = a^\nu \vec{e}_\nu$ 和 $\vec{b} = b^\nu \vec{e}_\nu$ 的外积就是

$$
\begin{aligned}
\vec{a} \wedge \vec{b} &= (a^1\vec{e}_1 + a^2\vec{e}_2 + a^3\vec{e}_3) \wedge (b^1\vec{e}_1 + b^2\vec{e}_2 + b^3\vec{e}_3) \\
&= (a^1b^2 - a^2b^1)\vec{e}_1 \wedge \vec{e}_2 + (a^2b^3 - a^3b^2)\vec{e}_2 \wedge \vec{e}_3 + (a^1b^3 - a^3b^1)\vec{e}_1 \wedge \vec{e}_3 \\
&= \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix} \vec{e}_1 \wedge \vec{e}_2 + \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix} \vec{e}_2 \wedge \vec{e}_3 + \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix} \vec{e}_1 \wedge \vec{e}_3
\end{aligned}
$$

显见，外积满足反交换律：$\vec{a} \wedge \vec{b} = -\vec{b} \wedge \vec{a}$。

任意三个点 $\vec{a}=a^\nu\vec{e}_\nu$，$\vec{b}=b^\nu\vec{e}_\nu$ 和 $\vec{c}=c^\nu\vec{e}_\nu$ 的外积是

$$\begin{aligned}
\vec{a}\wedge\vec{b}\wedge\vec{c} = &\left(\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}\vec{e}_1\wedge\vec{e}_2 + \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}\vec{e}_2\wedge\vec{e}_3 + \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}\vec{e}_1\wedge\vec{e}_3\right) \\
&\wedge(c^1\vec{e}_1 + c^2\vec{e}_2 + c^3\vec{e}_3) \\
= &\left(c^3\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix} + c^1\begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix} - c^2\begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}\right)\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3 \\
= &\begin{vmatrix} a^1 & a^2 & a^3 \\ b^1 & b^2 & b^3 \\ c^1 & c^2 & c^3 \end{vmatrix}\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3
\end{aligned}$$

易知，射影平面上，四个或更多个点的外积为零。

8.6.5 射影空间上的情况

五维空间中有标量场 $w=w(P)$，其等值超曲面在点 P 处有切超平面。五个切线都在切超平面上的条件是 $\partial_i w$ 有非零解，即五个切线的各分量做成的行列式为零。在点 P 的邻域内，以点 P 为球心，做一个微小超球并对径粘合，就做成四维射影空间。切超平面则是三维射影空间。因此，也说五点共“体”的条件是上述行列式为零。

具体讲，四维射影空间中任意点 γ 可记为 $\rho(\gamma^1,\gamma^2,\gamma^3,\gamma^4,\gamma^5)$，其中 ρ 为任意非零值，其他 γ^i 不全为零。该四维空间中，存在且只存在一个三维射影空间与给定的互异四点 $\alpha^i,\beta^i,\gamma^i,\zeta^i$ 相接合。该三维空间上任何一点 η^i，均可表为给定互异四点的线性组合，即存在不全为零的四个数 a,b,c,d，使得 $\eta^i=a\alpha^i+b\beta^i+c\gamma^i+d\zeta^i$。

在此情境下，称给定的互异四点为四个基矢，它们共同构成该空间的基底。记 $\vec{e}_1=(\alpha^i)$ 等，则任意点可表为 $\vec{r}=\lambda^\nu\vec{e}_\nu$。这里 ν 取 $1,2,3,4$，而 i 取 $1,2,3,4,5$。

用类似的方式规定基矢之间的外积，则任意两点 $\vec{a}=a^\nu\vec{e}_\nu$ 和 $\vec{b}=b^\nu\vec{e}_\nu$ 的外积就是

$$\begin{aligned}
\vec{a}\wedge\vec{b} = &(a^1\vec{e}_1+a^2\vec{e}_2+a^3\vec{e}_3+a^4\vec{e}_4)\wedge(b^1\vec{e}_1+b^2\vec{e}_2+b^3\vec{e}_3+b^4\vec{e}_4) \\
= &\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}\vec{e}_1\wedge\vec{e}_2 + \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}\vec{e}_2\wedge\vec{e}_3 + \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}\vec{e}_1\wedge\vec{e}_3 \\
&+\begin{vmatrix} a^1 & a^4 \\ b^1 & b^4 \end{vmatrix}\vec{e}_1\wedge\vec{e}_4 + \begin{vmatrix} a^2 & a^4 \\ b^2 & b^4 \end{vmatrix}\vec{e}_2\wedge\vec{e}_4 + \begin{vmatrix} a^3 & a^4 \\ b^3 & b^4 \end{vmatrix}\vec{e}_3\wedge\vec{e}_4
\end{aligned}$$

显见，外积满足反交换律：$\vec{a}\wedge\vec{b}=-\vec{b}\wedge\vec{a}$。

任意三点 $\vec{a}=a^\nu\vec{e}_\nu$，$\vec{b}=b^\nu\vec{e}_\nu$ 和 $\vec{c}=c^\nu\vec{e}_\nu$ 的外积就是

$$\begin{aligned}\vec{a}\wedge\vec{b}\wedge\vec{c}=&\begin{vmatrix}a^1&a^2&a^3\\b^1&b^2&b^3\\c^1&c^2&c^3\end{vmatrix}\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3+\begin{vmatrix}a^1&a^2&a^4\\b^1&b^2&b^4\\c^1&c^2&c^4\end{vmatrix}\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_4\\&+\begin{vmatrix}a^1&a^3&a^4\\b^1&b^3&b^4\\c^1&c^3&c^4\end{vmatrix}\vec{e}_1\wedge\vec{e}_3\wedge\vec{e}_4+\begin{vmatrix}a^2&a^3&a^4\\b^2&b^3&b^4\\c^2&c^3&c^4\end{vmatrix}\vec{e}_2\wedge\vec{e}_3\wedge\vec{e}_4\end{aligned}$$

任意四点 $\vec{a}=a^\nu\vec{e}_\nu$，$\vec{b}=b^\nu\vec{e}_\nu$，$\vec{c}=c^\nu\vec{e}_\nu$ 和 $\vec{d}=d^\nu\vec{e}_\nu$ 的外积就是

$$\begin{aligned}&\vec{a}\wedge\vec{b}\wedge\vec{c}\wedge\vec{d}\\=&\left(\begin{vmatrix}a^1&a^2&a^3\\b^1&b^2&b^3\\c^1&c^2&c^3\end{vmatrix}\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3+\begin{vmatrix}a^1&a^2&a^4\\b^1&b^2&b^4\\c^1&c^2&c^4\end{vmatrix}\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_4\right.\\&\left.+\begin{vmatrix}a^1&a^3&a^4\\b^1&b^3&b^4\\c^1&c^3&c^4\end{vmatrix}\vec{e}_1\wedge\vec{e}_3\wedge\vec{e}_4+\begin{vmatrix}a^2&a^3&a^4\\b^2&b^3&b^4\\c^2&c^3&c^4\end{vmatrix}\vec{e}_2\wedge\vec{e}_3\wedge\vec{e}_4\right)\\&\wedge(d^1\vec{e}_1+d^2\vec{e}_2+d^3\vec{e}_3+d^4\vec{e}_4)\\=&\left(d^4\begin{vmatrix}a^1&a^2&a^3\\b^1&b^2&b^3\\c^1&c^2&c^3\end{vmatrix}-d^3\begin{vmatrix}a^1&a^2&a^4\\b^1&b^2&b^4\\c^1&c^2&c^4\end{vmatrix}+d^2\begin{vmatrix}a^1&a^3&a^4\\b^1&b^3&b^4\\c^1&c^3&c^4\end{vmatrix}-d^1\begin{vmatrix}a^2&a^3&a^4\\b^2&b^3&b^4\\c^2&c^3&c^4\end{vmatrix}\right)\\&\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3\wedge\vec{e}_4\\=&\begin{vmatrix}a^1&a^2&a^3&a^4\\b^1&b^2&b^3&b^4\\c^1&c^2&c^3&c^4\\d^1&d^2&d^3&d^4\end{vmatrix}\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3\wedge\vec{e}_4\end{aligned}$$

五个或更多个点的外积为零。

8.6.6 排列的逆序数与奇偶性

在一组互异的正整数的排列中，如果一个较大的数排在一个较小的数之前，则称这两个数构成一个逆序对。排列中的逆序对的个数称为该排列的逆序数。比如，排列 $1,2,3,4$ 的逆序数为零，而 $1,2,4,3$ 的逆序数为 1，因为 $(4,3)$ 是一个逆序对。$1,2,4$ 的逆序数为零，而 $4,1,2$ 的逆序数为 2，因为 $(4,1)$ 和 $(4,2)$ 是两个逆序对。易见 $4,2,1$ 的逆序数为 3。

排列的逆序数的奇偶性称为排列的奇偶性。如果排列 σ 为奇，记 $\operatorname{sgn}(\sigma)=1$，否则，记 $\operatorname{sgn}(\sigma)=0$。比如，$\operatorname{sgn}(1,2,3,4)=0$，而 $\operatorname{sgn}(4,2,1)=1$。

8.6.7 n 维 k 度外积空间

一般地，在 n 维射影空间中，存在且只存在一个 $n-1$ 维射影子空间与给定的互异 n 点接合。该 $n-1$ 维射影子空间中的任何一点均可表为给定的互异 n 点的线性组合。称这互异的 n 个点为基矢，记为 $\vec{e}_\nu$。它们共同构成该子空间的基底。任意点可表为 $\vec{r} = \lambda^\nu \vec{e}_\nu$。

给定 n 维射影空间和互异的 n 个点 $\vec{e}_\nu$。

- 称所有标量 a 构成 n 维 0 度外积空间。它没有基底，或者说它的基底就是标量 1。它是 0 维的。
- 称所有矢量 $\vec{r}$ 构成 n 维 1 度外积空间。它的基底包括 n 个基矢 $\vec{e}_\nu$。它是 n 维的。
- 称所有 $\vec{e}_\mu \wedge \vec{e}_\nu$ 的线性组合构成 n 维 2 度外积空间。该空间是 $\dfrac{n(n-1)}{2}$ 维的，因为有 $\vec{e}_\mu \wedge \vec{e}_\mu = 0$ 和 $\vec{e}_\mu \wedge \vec{e}_\nu = -\vec{e}_\nu \wedge \vec{e}_\mu$。
- 依此类推，称所有 $\vec{e}_{\mu_1} \wedge \cdots \wedge \vec{e}_{\mu_k}$ 的线性组合构成 n 维 k 度外积空间。从互异的 n 个点 $\vec{e}_1, \ldots, \vec{e}_n$ 中选出 k 个，把它们按指标从小到大的顺序排列，这样的一次选择只会产生一个排列，因此，n 维 k 度外积空间是 $C_n^k = \dfrac{n!}{(n-k)!k!}$ 维的。它的独立基矢可以记为 $\vec{e}_{\mu_1} \wedge \cdots \wedge \vec{e}_{\mu_k}$，其中 $1 \le \mu_1 < \cdots < \mu_k \le n$。
- 称 $\vec{e}_1 \wedge \cdots \wedge \vec{e}_n$ 的所有标量倍构成 n 维 n 度外积空间。它有唯一的基矢 $\vec{e}_1 \wedge \cdots \wedge \vec{e}_n$。因为这个基矢是唯一的，$n$ 维 n 度外积空间中的点在这个基矢上的分量，在坐标变换下没有与其他分量"协调变化"的可能，因此该分量是个标量。进而，该空间是 0 维的。

8.6.8 外积的外积的交换特性与双线性

设有 n 维 k 度外积 $\omega = \sum\limits_{1\le\mu_1<\cdots<\mu_k\le n} a_{\mu_1\cdots\mu_k} \vec{e}_{\mu_1} \wedge \cdots \wedge \vec{e}_{\mu_k}$ 和 n 维 ℓ 度外积 $\eta = \sum\limits_{1\le\nu_1<\cdots<\nu_\ell\le n} b_{\nu_1\cdots\nu_\ell} \vec{e}_{\nu_1} \wedge \cdots \wedge \vec{e}_{\nu_\ell}$，则两者的外积定义为

$$\omega \wedge \eta = \sum_{\substack{1\le\mu_1<\cdots<\mu_k\le n\\ 1\le\nu_1<\cdots<\nu_\ell\le n}} a_{\mu_1\cdots\mu_k} b_{\nu_1\cdots\nu_\ell} \vec{e}_{\mu_1} \wedge \cdots \wedge \vec{e}_{\mu_k} \wedge \vec{e}_{\nu_1} \wedge \cdots \wedge \vec{e}_{\nu_\ell},$$

显然，在最一般的情况下，这是个 n 维 $k+\ell$ 度外积。

注意到，

$$\begin{aligned} &\omega \wedge \eta \\ =& \sum_{\substack{1\le\mu_1<\cdots<\mu_k\le n\\ 1\le\nu_1<\cdots<\nu_\ell\le n}} a_{\mu_1\cdots\mu_k} b_{\nu_1\cdots\nu_\ell} \vec{e}_{\mu_1} \wedge \cdots \wedge \vec{e}_{\mu_k} \wedge \vec{e}_{\nu_1} \wedge \cdots \wedge \vec{e}_{\nu_\ell} \end{aligned}$$

$$
\begin{aligned}
&= (-1)^k \sum_{\substack{1\le\mu_1<\cdots<\mu_k\le n\\ 1\le\nu_1<\cdots<\nu_\ell\le n}} a_{\mu_1\cdots\mu_k} b_{\nu_1\cdots\nu_\ell} \vec{e}_{\nu_1}\wedge\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k}\wedge\vec{e}_{\nu_2}\wedge\cdots\wedge\vec{e}_{\nu_\ell}\\
&= (-1)^{2k} \sum_{\substack{1\le\mu_1<\cdots<\mu_k\le n\\ 1\le\nu_1<\cdots<\nu_\ell\le n}} a_{\mu_1\cdots\mu_k} b_{\nu_1\cdots\nu_\ell} \vec{e}_{\nu_1}\wedge\vec{e}_{\nu_2}\wedge\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k}\wedge\vec{e}_{\nu_3}\wedge\cdots\wedge\vec{e}_{\nu_\ell}\\
&= \cdots\\
&= (-1)^{\ell k} \sum_{\substack{1\le\mu_1<\cdots<\mu_k\le n\\ 1\le\nu_1<\cdots<\nu_\ell\le n}} a_{\mu_1\cdots\mu_k} b_{\nu_1\cdots\nu_\ell} \vec{e}_{\nu_1}\wedge\cdots\wedge\vec{e}_{\nu_\ell}\wedge\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k}\\
&= (-1)^{\ell k}\eta\wedge\omega.
\end{aligned}
$$

这称为外积的外积的交换特性。

注意到，对任意常数 c 和 d，

$$
\begin{aligned}
c\omega\wedge d\eta &= \sum_{\substack{1\le\mu_1<\cdots<\mu_k\le n\\ 1\le\nu_1<\cdots<\nu_\ell\le n}} cda_{\mu_1\cdots\mu_k} b_{\nu_1\cdots\nu_\ell} \vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k}\wedge\vec{e}_{\nu_1}\wedge\cdots\wedge\vec{e}_{\nu_\ell}\\
&= cd\omega\wedge\eta.
\end{aligned}
$$

这称为外积的外积的双线性。

8.6.9 霍奇星

霍奇星运算作用在 n 维 k 度外积空间中的点上，目标是使得该空间的任何一个基矢 $\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k}$ 均满足

$$
(\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k})\wedge(*(\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k}))=\vec{e}_1\wedge\cdots\wedge\vec{e}_n.
$$

例如，4 维 2 度外积空间上，

- 对于基矢 $\vec{e}_1\wedge\vec{e}_2$，因为 $(\vec{e}_1\wedge\vec{e}_2)\wedge(\vec{e}_3\wedge\vec{e}_4)=\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3\wedge\vec{e}_4$，所以 $*(\vec{e}_1\wedge\vec{e}_2)=\vec{e}_3\wedge\vec{e}_4$。
- 对于基矢 $\vec{e}_2\wedge\vec{e}_4$，因为 $(\vec{e}_2\wedge\vec{e}_4)\wedge(\vec{e}_1\wedge\vec{e}_3)=-\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3\wedge\vec{e}_4$，所以 $*(\vec{e}_2\wedge\vec{e}_4)=-\vec{e}_1\wedge\vec{e}_3$。

一般地，给定 n 维 k 度外积空间的一个基矢 $\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k}$。该基矢可以看做是从互异的 n 个点 $\vec{e}_1,\ldots,\vec{e}_n$ 中选出 k 个，把它们按指标从小到大的顺序做外积而得到。那么，那些没被选中的 $n-k$ 个点，按指标从小到大的顺序做外积就得到 $\vec{e}_{\nu_1}\wedge\cdots\wedge\vec{e}_{\nu_{n-k}}$。规定霍奇星运算 $*$ 为

$$
*(\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k})=(-1)^{\mathrm{sgn}(\mu_1,\ldots,\mu_k,\nu_1,\ldots,\nu_{n-k})}\vec{e}_{\nu_1}\wedge\cdots\wedge\vec{e}_{\nu_{n-k}}.
$$

同时规定霍奇星运算是线性的。这样，由于 n 维 k 度外积空间中的任何点都是形如 $\vec{e}_{\mu_1}\wedge\cdots\wedge\vec{e}_{\mu_k}$ 的基矢的线性组合，所以，该空间中任意点的霍奇星等于各基矢的霍奇星的同一线性组合。

易见，$*^2=1$。

另外规定，$*1=\vec{e}_1\wedge\cdots\wedge\vec{e}_n$。易见，$*(\vec{e}_1\wedge\cdots\wedge\vec{e}_n)=1$。

8.6.10 n 阶行列式的拉普拉斯展开

我们已经看到，由矢量 $\vec{a}_1,\ldots,\vec{a}_n$ 做成的 n 阶行列式可以被定义为

$$|\vec{a}_1,\ldots,\vec{a}_n\,| = *(\vec{a}_1\wedge\cdots\wedge\vec{a}_n).$$

以 4 阶行列式为例。

$$\begin{vmatrix} a^1 & a^2 & a^3 & a^4 \\ b^1 & b^2 & b^3 & b^4 \\ c^1 & c^2 & c^3 & c^4 \\ d^1 & d^2 & d^3 & d^4 \end{vmatrix} = *(\vec{a}\wedge\vec{b}\wedge\vec{c}\wedge\vec{d}) = *((\vec{a}\wedge\vec{b})\wedge(\vec{c}\wedge\vec{d}))$$

$$\begin{aligned}
= *\Bigg(\Bigg(&\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}\vec{e}_1\wedge\vec{e}_2 + \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}\vec{e}_2\wedge\vec{e}_3 + \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}\vec{e}_1\wedge\vec{e}_3 \\
&+ \begin{vmatrix} a^1 & a^4 \\ b^1 & b^4 \end{vmatrix}\vec{e}_1\wedge\vec{e}_4 + \begin{vmatrix} a^2 & a^4 \\ b^2 & b^4 \end{vmatrix}\vec{e}_2\wedge\vec{e}_4 + \begin{vmatrix} a^3 & a^4 \\ b^3 & b^4 \end{vmatrix}\vec{e}_3\wedge\vec{e}_4\Bigg) \\
&\wedge\Bigg(\begin{vmatrix} c^1 & c^2 \\ d^1 & d^2 \end{vmatrix}\vec{e}_1\wedge\vec{e}_2 + \begin{vmatrix} c^2 & c^3 \\ d^2 & d^3 \end{vmatrix}\vec{e}_2\wedge\vec{e}_3 + \begin{vmatrix} c^1 & c^3 \\ d^1 & d^3 \end{vmatrix}\vec{e}_1\wedge\vec{e}_3 \\
&+ \begin{vmatrix} c^1 & c^4 \\ d^1 & d^4 \end{vmatrix}\vec{e}_1\wedge\vec{e}_4 + \begin{vmatrix} c^2 & c^4 \\ d^2 & d^4 \end{vmatrix}\vec{e}_2\wedge\vec{e}_4 + \begin{vmatrix} c^3 & c^4 \\ d^3 & d^4 \end{vmatrix}\vec{e}_3\wedge\vec{e}_4\Bigg)\Bigg) \\
= *\Bigg(&\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}\begin{vmatrix} c^3 & c^4 \\ d^3 & d^4 \end{vmatrix}\vec{e}_1\wedge\vec{e}_2\wedge\vec{e}_3\wedge\vec{e}_4 + \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}\begin{vmatrix} c^1 & c^4 \\ d^1 & d^4 \end{vmatrix}\vec{e}_2\wedge\vec{e}_3\wedge\vec{e}_1\wedge\vec{e}_4 \\
&+ \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}\begin{vmatrix} c^2 & c^4 \\ d^2 & d^4 \end{vmatrix}\vec{e}_1\wedge\vec{e}_3\wedge\vec{e}_2\wedge\vec{e}_4 + \begin{vmatrix} a^1 & a^4 \\ b^1 & b^4 \end{vmatrix}\begin{vmatrix} c^2 & c^3 \\ d^2 & d^3 \end{vmatrix}\vec{e}_1\wedge\vec{e}_4\wedge\vec{e}_2\wedge\vec{e}_3 \\
&+ \begin{vmatrix} a^2 & a^4 \\ b^2 & b^4 \end{vmatrix}\begin{vmatrix} c^1 & c^3 \\ d^1 & d^3 \end{vmatrix}\vec{e}_2\wedge\vec{e}_4\wedge\vec{e}_1\wedge\vec{e}_3 + \begin{vmatrix} a^3 & a^4 \\ b^3 & b^4 \end{vmatrix}\begin{vmatrix} c^1 & c^2 \\ d^1 & d^2 \end{vmatrix}\vec{e}_3\wedge\vec{e}_4\wedge\vec{e}_1\wedge\vec{e}_2\Bigg) \\
= *\Bigg(&\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}\left((-1)^{\mathrm{sgn}(1,2,3,4)}\begin{vmatrix} c^3 & c^4 \\ d^3 & d^4 \end{vmatrix}\right)(\vec{e}_1\wedge\vec{e}_2)\wedge(*(\vec{e}_1\wedge\vec{e}_2)) \\
&+ \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}\left((-1)^{\mathrm{sgn}(2,3,1,4)}\begin{vmatrix} c^1 & c^4 \\ d^1 & d^4 \end{vmatrix}\right)(\vec{e}_2\wedge\vec{e}_3)\wedge(*(\vec{e}_2\wedge\vec{e}_3))
\end{aligned}$$

$\cdots$ （其他四项类似，兹不赘述。）

$$\begin{aligned}
= &\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}\begin{vmatrix} c^3 & c^4 \\ d^3 & d^4 \end{vmatrix} + \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}\begin{vmatrix} c^1 & c^4 \\ d^1 & d^4 \end{vmatrix} - \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}\begin{vmatrix} c^2 & c^4 \\ d^2 & d^4 \end{vmatrix} \\
&+ \begin{vmatrix} a^1 & a^4 \\ b^1 & b^4 \end{vmatrix}\begin{vmatrix} c^2 & c^3 \\ d^2 & d^3 \end{vmatrix} - \begin{vmatrix} a^2 & a^4 \\ b^2 & b^4 \end{vmatrix}\begin{vmatrix} c^1 & c^3 \\ d^1 & d^3 \end{vmatrix} + \begin{vmatrix} a^3 & a^4 \\ b^3 & b^4 \end{vmatrix}\begin{vmatrix} c^1 & c^2 \\ d^1 & d^2 \end{vmatrix}
\end{aligned}$$

一般地，n 阶行列式

$$|\vec{a}_1,\ldots,\vec{a}_n| = *(\vec{a}_1\wedge\cdots\wedge\vec{a}_n)$$

$$= *((\vec{a}_1 \wedge \cdots \wedge \vec{a}_k) \wedge (\vec{a}_{k+1} \wedge \cdots \wedge \vec{a}_n))$$

$\vec{a}_1 \wedge \cdots \wedge \vec{a}_k$ 在基矢 $\vec{e}_{\mu_1} \wedge \cdots \wedge \vec{e}_{\mu_k}$ 上的分量称为该 n 阶行列式的 k 阶子式。设与该基矢“互补”的基矢为 $\vec{e}_{\nu_1} \wedge \cdots \wedge \vec{e}_{\nu_{n-k}}$，那么 $\vec{a}_{k+1} \wedge \cdots \wedge \vec{a}_n$ 在这个基矢上的分量就称为刚才那个 k 阶子式的余子式。余子式乘上符号因子 $(-1)^{\mathrm{sgn}(\mu_1,\ldots,\mu_k,\nu_1,\ldots,\nu_{n-k})}$，就称为刚才那个 k 阶子式的代数余子式。

n 阶行列式等于前 k 行的 C_n^k 个 k 阶子式与相应的代数余子式的乘积之和。这就是 n 阶行列式的拉普拉斯展开。

8.6.11 空间中直线的 Plücker 极坐标和轴坐标

三维射影空间中相异两点 $\vec{a}$ 和 $\vec{b}$ 确定且只确定一条直线。如果两直线相交，那么两直线共面。从两直线上各取两点，使这四点互异，则两直线共面，当且仅当四点共面。于是，按上节所述，

$$\begin{aligned}
&|\vec{a},\vec{b},\vec{c},\vec{d}| = *(\vec{a} \wedge \vec{b} \wedge \vec{c} \wedge \vec{d}) \\
= &\begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}\begin{vmatrix} c^3 & c^4 \\ d^3 & d^4 \end{vmatrix} + \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}\begin{vmatrix} c^1 & c^4 \\ d^1 & d^4 \end{vmatrix} - \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix}\begin{vmatrix} c^2 & c^4 \\ d^2 & d^4 \end{vmatrix} \\
&+ \begin{vmatrix} a^1 & a^4 \\ b^1 & b^4 \end{vmatrix}\begin{vmatrix} c^2 & c^3 \\ d^2 & d^3 \end{vmatrix} - \begin{vmatrix} a^2 & a^4 \\ b^2 & b^4 \end{vmatrix}\begin{vmatrix} c^1 & c^3 \\ d^1 & d^3 \end{vmatrix} + \begin{vmatrix} a^3 & a^4 \\ b^3 & b^4 \end{vmatrix}\begin{vmatrix} c^1 & c^2 \\ d^1 & d^2 \end{vmatrix}
\end{aligned}$$

如果取

$$p_{12} = \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix}, \quad p_{23} = \begin{vmatrix} a^2 & a^3 \\ b^2 & b^3 \end{vmatrix}, \quad p_{13} = \begin{vmatrix} a^1 & a^3 \\ b^1 & b^3 \end{vmatrix},$$
$$p_{14} = \begin{vmatrix} a^1 & a^4 \\ b^1 & b^4 \end{vmatrix}, \quad p_{42} = -\begin{vmatrix} a^2 & a^4 \\ b^2 & b^4 \end{vmatrix}, \quad p_{34} = \begin{vmatrix} a^3 & a^4 \\ b^3 & b^4 \end{vmatrix}$$

那么，这六个数可以刻画点 $\vec{a}$ 和点 $\vec{b}$ 的连线，称为该直线的 Plücker 极坐标。

注意到 $(\vec{a} \wedge \vec{b}) \wedge (\vec{a} \wedge \vec{b}) = 0$ 恒成立，所以，

$$P = p_{12}p_{34} + p_{13}p_{42} + p_{14}p_{23} = 0.$$

由于这些坐标是齐次坐标，而且又受该恒等式的约束，所以，这些坐标只有 4 个独立参数。

对偶地，空间中两平面确定且只确定一条直线。两直线相交意味着四面共点。同理可得六个子行列式 $q_{12}, q_{34}, q_{13}, q_{42}, q_{14}, q_{23}$，称为直线的 Plücker 轴坐标。

轴坐标有约束关系

$$Q = q_{12}q_{34} + q_{13}q_{42} + q_{14}q_{23} = 0.$$

设上述两种方法刻画同一直线，那么

$$a^i\zeta_i = 0, \quad b^i\zeta_i = 0, \quad a^i\eta_i = 0, \quad b^i\eta_i = 0$$

用这些关系可以证明：

$$p_{12} : p_{13} : p_{14} : p_{23} : p_{42} : p_{34} = q_{34} : q_{42} : q_{23} : q_{14} : q_{13} : q_{12}$$

因此，同一直线的 Plücker 极坐标与轴坐标有变换关系：

$$\rho\, p_{ik} = \frac{\partial Q}{\partial q_{ik}}, \qquad \sigma\, q_{ik} = \frac{\partial P}{\partial p_{ik}}.$$

8.6.12 矢量的内积

两个矢量 $\vec{a}$ 和 $\vec{b}$ 的内积可以定义为

$$\vec{a} \cdot \vec{b} = *(\vec{a} \wedge (*\vec{b})) = a^1b^1 + \cdots + a^nb^n.$$

8.6.13 极矢量、轴矢量和内积运算

3 维 1 度外积空间中的矢量称为极矢量，它的基底是 $\vec{e}_x, \vec{e}_y, \vec{e}_z$。两个极矢量的外积是 3 维 2 度外积空间中的元素，而该空间恰好也是 3 维的，这使得该元素在很多方面表现得就像极矢量一样。但是，它当然不是极矢量。比如，设 $\vec{r}$ 和 $\vec{s}$ 是极矢量，做宇称变换，有 $P(\vec{r}) = -\vec{r}$ 和 $P(\vec{s}) = -\vec{s}$，但是，$P(\vec{r} \wedge \vec{s}) = \vec{r} \wedge \vec{s}$，它在宇称变换下并不像极矢量一样改变方向。因此，称 3 维 2 度外积空间中的元素为轴矢量。

一般的外积空间的基底 $\vec{e}_{\mu_1} \wedge \cdots \wedge \vec{e}_{\mu_k}$，通常选择 $1 \le \mu_1 < \cdots < \mu_k \le n$，这样，对下标的处理比较系统化。但是对 3 维 2 度外积空间，规定手性会更加方便，因此，我们选择基底 $\vec{e}_x \wedge \vec{e}_y, \vec{e}_y \wedge \vec{e}_z, \vec{e}_z \wedge \vec{e}_x$，使得极矢量的外积符合右手系的定向规则，即

$$\vec{e}_x \wedge \vec{e}_y \sim \vec{e}_z, \quad \vec{e}_y \wedge \vec{e}_z \sim \vec{e}_x, \quad \vec{e}_z \wedge \vec{e}_x \sim \vec{e}_y.$$

这样就可以把 3 维 2 度外积空间中的矢量 $\vec{A} = P\vec{e}_y \wedge \vec{e}_z + Q\vec{e}_z \wedge \vec{e}_x + R\vec{e}_x \wedge \vec{e}_y$ 改写成 $\vec{A} = P\vec{e}_x + Q\vec{e}_y + R\vec{e}_z$，看起来就像极矢量一样。如果不区分轴矢量和极矢量，那么极矢量的外积也被称为矢量的叉乘运算，即：两个矢量的叉乘得到一个新的矢量。

两个极矢量 $\vec{a}$ 和 $\vec{b}$ 的内积可以定义为

$$\vec{a} \cdot \vec{b} = *(\vec{a} \wedge (*\vec{b})) = a_xb_x + a_yb_y + a_zb_z.$$

一个极矢量 $\vec{a}$ 和一个轴矢量 $\vec{A}$ 的内积定义为

$$\vec{a} \cdot \vec{A} = *(\vec{a} \wedge \vec{A}) = a_xA_x + a_yA_y + a_zA_z.$$

两个轴矢量 $\vec{A}$ 和 $\vec{B}$ 的内积定义为

$$\vec{A}\cdot\vec{B} = *(\vec{A}\wedge(*\vec{B})) = A_x B_x + A_y B_y + A_z B_z.$$

8.7 微分形式与广义斯托克斯定理

8.7.1 n 维 k 度微分形式

在前面一开始引入微分的概念的时候，我们就说，（一元情况下，）微分本质上是积分的缩写形式，是在不关注积分上下限时对积分号的省略。多元情况下，基本上，n 维 k 度微分形式就是 n 维流形中的 k 维子流形 M 上的张量场的积分的缩写形式，常记为 ω。而相应的积分则记为 $\int_M \omega$。

比如，空间中的力场 $\vec{F}(x,y,z)$ 沿路径 $\gamma(\tau)$ 所做的功为

$$W = \int_\gamma \vec{F}\cdot \mathrm{d}\vec{s} = \int_\gamma F_x\,\mathrm{d}x + F_y\,\mathrm{d}y + F_z\,\mathrm{d}z,$$

其中的 $\vec{F}\cdot\mathrm{d}\vec{s}$ 和 $F_x\,\mathrm{d}x + F_y\,\mathrm{d}y + F_z\,\mathrm{d}z$ 都可以称作 3 维 1 度微分形式。

再比如，空间中不可压缩流体的流速场 $\vec{v}(x,y,z)$ 通过一片曲面 Σ 的流量为

$$Q = \iint_\Sigma \vec{v}\cdot\mathrm{d}\vec{S} = \iint_\Sigma v_z\,\mathrm{d}x\,\mathrm{d}y + v_x\,\mathrm{d}y\,\mathrm{d}z + v_y\,\mathrm{d}z\,\mathrm{d}x,$$

其中的 $\vec{v}\cdot\mathrm{d}\vec{S}$ 和 $v_z\,\mathrm{d}x\,\mathrm{d}y + v_x\,\mathrm{d}y\,\mathrm{d}z + v_y\,\mathrm{d}z\,\mathrm{d}x$ 都可称作 3 维 2 度微分形式。

再比如，空间中的密度场 $\rho(x,y,z)$ 在一块区域 Ω 内的质量为

$$M = \iiint_\Omega \rho\,\mathrm{d}V = \iiint_\Omega \rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z,$$

其中的 $\rho\,\mathrm{d}V$ 和 $\rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ 都可称作 3 维 3 度微分形式。

我们把标量场 $f(x,y,z)$ 本身称为 3 维 0 度微分形式，相应的积分是 $\int_{\{P_0,P_1\}} f(x,y,z)$。这个积分没有 $\mathrm{d}x$ 或者 $\mathrm{d}y$ 等的部分，它计算的是标量场在一维流形两端点上的“有向取值”的和。所谓“有向取值”，是指对有向曲线的起点 P_0，“有向取值”等于标量场在该点上的取值的相反数，即 $-f(P_0)$，对有向曲线的终点 P_1，“有向取值”等于标量场在该点上的取值，即 $f(P_1)$。这样，$\int_{\{P_0,P_1\}} f(x,y,z) = f(P_1) - f(P_0)$。

8.7.2 微分形式与矢量外积的关系

对于上一小节的做功问题，$\mathrm{d}\vec{s}$ 是极矢量，即 $\mathrm{d}\vec{s} = \mathrm{d}x\vec{e}_x + \mathrm{d}y\vec{e}_y + \mathrm{d}z\vec{e}_z$，所以有

$$\omega = \vec{F}\cdot\mathrm{d}\vec{s} = *(\vec{F}\wedge(*(\mathrm{d}\vec{s})))$$

$$
\begin{aligned}
&= *((F_x\vec{e}_x + F_y\vec{e}_y + F_z\vec{e}_z) \wedge (*(\,\mathrm{d}x\vec{e}_x + \,\mathrm{d}y\vec{e}_y + \,\mathrm{d}z\vec{e}_z))) \\
&= *((F_x\vec{e}_x + F_y\vec{e}_y + F_z\vec{e}_z) \wedge (\,\mathrm{d}x(*\vec{e}_x) + \,\mathrm{d}y(*\vec{e}_y) + \,\mathrm{d}z(*\vec{e}_z))) \\
&= *((F_x\,\mathrm{d}x + F_y\,\mathrm{d}y + F_z\,\mathrm{d}z)\vec{e}_x \wedge \vec{e}_y \wedge \vec{e}_z) \\
&= F_x\,\mathrm{d}x + F_y\,\mathrm{d}y + F_z\,\mathrm{d}z
\end{aligned}
$$

对于流量问题，$\mathrm{d}\vec{S}$ 是轴矢量，即 $\mathrm{d}\vec{S} = \mathrm{d}x\,\mathrm{d}y\,\vec{e}_x \wedge \vec{e}_y + \mathrm{d}z\,\mathrm{d}x\,\vec{e}_z \wedge \vec{e}_x + \mathrm{d}y\,\mathrm{d}z\,\vec{e}_y \wedge \vec{e}_z$，所以，有

$$
\begin{aligned}
\omega &= \vec{v} \cdot \mathrm{d}\vec{S} = *(\vec{v} \wedge \mathrm{d}\vec{S}) \\
&= *((v_x\vec{e}_x + v_y\vec{e}_y + v_z\vec{e}_z) \wedge (\,\mathrm{d}x\,\mathrm{d}y\vec{e}_x \wedge \vec{e}_y + \,\mathrm{d}z\,\mathrm{d}x\vec{e}_z \wedge \vec{e}_x + \,\mathrm{d}y\,\mathrm{d}z\vec{e}_y \wedge \vec{e}_z)) \\
&= *((v_z\,\mathrm{d}x\,\mathrm{d}y + v_x\,\mathrm{d}y\,\mathrm{d}z + v_y\,\mathrm{d}z\,\mathrm{d}x)\vec{e}_x \wedge \vec{e}_y \wedge \vec{e}_z) \\
&= v_z\,\mathrm{d}x\,\mathrm{d}y + v_x\,\mathrm{d}y\,\mathrm{d}z + v_y\,\mathrm{d}z\,\mathrm{d}x
\end{aligned}
$$

对于质量问题，有 $\mathrm{d}V = *((\,\mathrm{d}x\vec{e}_x) \wedge (\,\mathrm{d}y\vec{e}_y) \wedge (\,\mathrm{d}z\vec{e}_z)) = *(\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,\vec{e}_x \wedge \vec{e}_y \wedge \vec{e}_z) = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$，所以 $\omega = \rho\,\mathrm{d}V = \rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$。

如果把 $\mathrm{d}x\vec{e}_x$ 简记为 $\mathrm{d}\vec{x}$，那么，做功问题可简写为

$$
\begin{aligned}
\omega &= *(\vec{F} \wedge (*(\,\mathrm{d}\vec{s}))) \\
&= *(\vec{F} \wedge (*\,\mathrm{d}\vec{x}) + \vec{F} \wedge (*\,\mathrm{d}\vec{y}) + \vec{F} \wedge (*\,\mathrm{d}\vec{z})) \\
&= F_x\,\mathrm{d}x + F_y\,\mathrm{d}y + F_z\,\mathrm{d}z
\end{aligned}
$$

流量问题可简写为

$$
\begin{aligned}
\omega &= *(\vec{v} \wedge \mathrm{d}\vec{S}) \\
&= *(\vec{v} \wedge (\,\mathrm{d}\vec{x} \wedge \mathrm{d}\vec{y} + \mathrm{d}\vec{y} \wedge \mathrm{d}\vec{z} + \mathrm{d}\vec{z} \wedge \mathrm{d}\vec{x})) \\
&= v_z\,\mathrm{d}x\,\mathrm{d}y + v_x\,\mathrm{d}y\,\mathrm{d}z + v_y\,\mathrm{d}z\,\mathrm{d}x
\end{aligned}
$$

质量问题可简写为 $\omega = *(\rho\,\mathrm{d}\vec{x} \wedge \mathrm{d}\vec{y} \wedge \mathrm{d}\vec{z}) = \rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$。

这样，我们可以扩展微分形式的外延，把只含有 $\mathrm{d}\vec{x}$ 或 $\mathrm{d}\vec{y}$ 等的一次式称为 1 度微分形式，含有 $\mathrm{d}\vec{x} \wedge \mathrm{d}\vec{y}$ 之类的二次外积式称为 2 度微分形式，含有 $\mathrm{d}\vec{x} \wedge \mathrm{d}\vec{y} \wedge \mathrm{d}\vec{z}$ 之类的三次外积式称为 3 度微分形式。在这种意义上，我们可以说，一个 n 维 k 度微分形式 ω 与一个 n 维 ℓ 度微分形式 η 的外积 $\omega \wedge \eta$，在最一般的情况下，是一个 n 维 $k+\ell$ 度微分形式。

与矢量的外积的交换特性一样，微分形式的外积有 $\omega \wedge \eta = (-1)^{k\ell}\eta \wedge \omega$。

任何函数 $f(x, y, z)$ 在可微区域内的任何一个微小局部，都可视为常数，因此，微分形式的外积对函数乘是双线性的，即 $(f\omega) \wedge (g\eta) = fg\omega \wedge \eta$。

实际上，函数 f 是 0 度微分形式。函数乘 $f\omega$ 就是外积 $f \wedge \omega$，它是 $k+0 = k$ 度微分形式，所以，函数乘不增加度数。在交换时，$f \wedge \omega = (-1)^{k0}\omega \wedge f =$

$\omega \wedge f$，所以，函数乘是可交换的。对于双线性，有

$$\begin{aligned}(f\omega)\wedge(g\eta) &= f\wedge(\omega\wedge g)\wedge\eta \\ &= f\wedge(g\wedge\omega)\wedge\eta \\ &= fg\omega\wedge\eta.\end{aligned}$$

注意，ω 也可以是 0 度微分形式，即 $f\wedge g = fg$。

8.7.3 微分形式的微分

上节用矢量外积对微分形式进行“解释”，只是为了帮助看官了解新旧知识之间的衔接和过渡。实际上，微分形式的外积表达方式可以完全用微分算符来定义。

n 维 k 度微分形式构成的空间是以 $\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}$ 为基底的外积空间。设 ω 是该空间中的一个元素，它在该基底上展开为 $\omega = T_{i_1\cdots i_k}\,\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}$。微分算符 d 把它变换成 n 维 $k+1$ 度微分形式

$$\mathrm{d}\omega = \partial_j T_{i_1\cdots i_k}\,\mathrm{d}x^j\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}.$$

对任何微分形式 ω，有 $\mathrm{d}(\mathrm{d}\omega)=0$，简记为 $\mathrm{d}^2=0$。因为

$$\begin{aligned}\mathrm{d}(\mathrm{d}\omega) &= \mathrm{d}(\partial_i T_{i_1\cdots i_k}\,\mathrm{d}x^i\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}) \\ &= \partial_{ij}T_{i_1\cdots i_k}\,\mathrm{d}x^j\wedge\mathrm{d}x^i\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k},\end{aligned}$$

因为 $T_{i_1\cdots i_k}$ 可微，有 $\partial_{ij}T_{i_1\cdots i_k} = \partial_{ji}T_{i_1\cdots i_k}$，而 $\mathrm{d}x^j\wedge\mathrm{d}x^i = -\mathrm{d}x^i\wedge\mathrm{d}x^j$，所以上式为零。 □

如果 ω 是 k 度微分形式，η 是 ℓ 度的，那么

$$\begin{aligned}&\mathrm{d}(\omega\wedge\eta) \\ =\ &\mathrm{d}(T_{i_1\cdots i_k}\,\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}\wedge S_{j_1\cdots j_\ell}\,\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_\ell}) \\ =\ &\mathrm{d}(T_{i_1\cdots i_k}S_{j_1\cdots j_\ell}\,\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}\wedge\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_\ell}) \\ =\ &(\mathrm{d}(T_{i_1\cdots i_k})S_{j_1\cdots j_\ell} + T_{i_1\cdots i_k}\,\mathrm{d}(S_{j_1\cdots j_\ell}))\,\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k} \\ &\wedge\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_\ell}) \\ =\ &(\partial_m T_{i_1\cdots i_k}\,\mathrm{d}x^m S_{j_1\cdots j_\ell} + T_{i_1\cdots i_k}\partial_n S_{j_1\cdots j_\ell}\,\mathrm{d}x^n) \\ &\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}\wedge\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_\ell}) \\ =\ &(\partial_m T_{i_1\cdots i_k}\,\mathrm{d}x^m\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k})\wedge(S_{j_1\cdots j_\ell}\,\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_\ell}) \\ &+(-1)^k(T_{i_1\cdots i_k}\,\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k})\wedge(\partial_n S_{j_1\cdots j_\ell}\,\mathrm{d}x^n\wedge\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_\ell}) \\ =\ &\mathrm{d}\omega\wedge\eta + (-1)^k\omega\wedge\mathrm{d}\eta\end{aligned}$$

其中的 $(-1)^k$ 是 $\mathrm{d}x^n$ 穿越 $\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}$ 造成的。

易见，如果 ω 是 k 度微分形式，那么有 $\mathrm{d}(\mathrm{d}\omega\wedge\eta)=(-1)^{k+1}\mathrm{d}\omega\wedge\mathrm{d}\eta$，和 $\mathrm{d}(\omega\wedge\mathrm{d}\eta)=\mathrm{d}\omega\wedge\mathrm{d}\eta$。

8.7.4 微分形式与完全反对称张量

注意到，$\partial_k h\partial_j g\partial_i f$ 恰为一个三协张量，它是三个协变矢量的并积。

在前面讲过的质量问题中，给定空间中的密度场 $\rho(x,y,z)$，一块区域 Ω 内的质量为 $M=\iiint_\Omega\rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$。现在我们又说 3 维 3 度微分形式是 $T_{ijk}\,\mathrm{d}x^i\wedge\mathrm{d}x^j\wedge\mathrm{d}x^k$，那么，这两种表述到底是如何联系在一起的呢？

让我们耐心一点，把 $T_{ijk}\,\mathrm{d}x^i\wedge\mathrm{d}x^j\wedge\mathrm{d}x^k$ 展开，有

$$\begin{aligned}
&T_{ijk}\,\mathrm{d}x^i\wedge\mathrm{d}x^j\wedge\mathrm{d}x^k\\
=\;&T_{xyz}\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z+T_{xzy}\,\mathrm{d}x\wedge\mathrm{d}z\wedge\mathrm{d}y+T_{yxz}\,\mathrm{d}y\wedge\mathrm{d}x\wedge\mathrm{d}z\\
&+T_{yzx}\,\mathrm{d}y\wedge\mathrm{d}z\wedge\mathrm{d}x+T_{zxy}\,\mathrm{d}z\wedge\mathrm{d}x\wedge\mathrm{d}y+T_{zyx}\,\mathrm{d}z\wedge\mathrm{d}y\wedge\mathrm{d}x\\
=\;&T_{xyz}\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z-T_{xzy}\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z-T_{yxz}\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z\\
&+T_{yzx}\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z+T_{zxy}\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z-T_{zyx}\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z\\
=\;&(T_{xyz}-T_{xzy}-T_{yxz}+T_{yzx}+T_{zxy}-T_{zyx})\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z
\end{aligned}$$

如果记 $\varepsilon^{xyz}=\varepsilon^{yzx}=\varepsilon^{zxy}=1,\quad\varepsilon^{xzy}=\varepsilon^{yxz}=\varepsilon^{zyx}=-1$，其他分量为零，则

$$T_{ijk}\,\mathrm{d}x^i\wedge\mathrm{d}x^j\wedge\mathrm{d}x^k=(\varepsilon^{ijk}T_{ijk})\,\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z.$$

其中的 ε^{ijk} 称为三阶完全反对称张量，也称行列式张量。这样，$\rho=\varepsilon^{ijk}T_{ijk}$ 是个标量。

对三维空间中的极矢量 $\vec a$、$\vec b$ 和 $\vec c$，$\vec a\wedge\vec b$ 表示两个矢量张起的平行四边形的有向面积，而 $\vec a\wedge\vec b\wedge\vec c$ 表示三个矢量张起的平行六面体的有向体积。多数讲三重积分 $M=\iiint_\Omega\rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ 的教材，把 $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ 说成是长、宽、高分别为 $\mathrm{d}x$、$\mathrm{d}y$、$\mathrm{d}z$ 的长方体微元的体积，仿佛把这个"体积"写成 $\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}z$ 也无所谓似的。其实不然，准确地说，$\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ 是 $\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z$，即 $(\mathrm{d}x\vec e_x)\wedge(\mathrm{d}y\vec e_y)\wedge(\mathrm{d}z\vec e_z)$，是"有向体积"微元，交换位置会导致体积的"方向"，即手性，的改变。可以说，$\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ 是 $\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z$ 的简写。

8.7.5 坐标变换下体积微元的变换

上例中，在曲线坐标系 (u,v,w) 下，密度场 $\rho(P)$ 的表示为 $\rho(u,v,w)$。求质量时用到的微分形式为 $\rho(u,v,w)\,\mathrm{d}u\wedge\mathrm{d}v\wedge\mathrm{d}w$。换成坐标系 (x,y,z)，则该微分形式变换成

$$\rho(u,v,w)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$$

$$
\begin{aligned}
&= \rho(u,v,w)\,\mathrm{d}u \wedge \mathrm{d}v \wedge \mathrm{d}w \\
&= \rho(u(x,y,z),v(x,y,z),w(x,y,z))\,\mathrm{d}u(x,y,z) \wedge \mathrm{d}v(x,y,z) \wedge \mathrm{d}w(x,y,z) \\
&= \rho'(x,y,z)(\partial_i u\,\mathrm{d}x^i) \wedge (\partial_j v\,\mathrm{d}x^j) \wedge (\partial_k w\,\mathrm{d}x^k) \\
&= \rho'(x,y,z)(\partial_i u \partial_j v \partial_k w\,\mathrm{d}x^i \wedge \mathrm{d}x^j \wedge \mathrm{d}x^k) \\
&= \rho'(x,y,z)(\varepsilon^{ijk}\partial_i u \partial_j v \partial_k w)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z
\end{aligned}
$$

从 $\rho(u,v,w)$ 到 $\rho'(x,y,z)$ 被认为是标量场的坐标变换。从 $\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$ 到 $(\varepsilon^{ijk}\partial_i u\partial_j v\partial_k w)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ 被认为是体积微元的坐标变换。而 $\varepsilon^{ijk}\partial_i u\partial_j v\partial_k w$ 正是所谓的雅可比行列式。

8.7.6 势能定理与梯度

给定空间标量场 $w(P) = f(x,y,z)$，有 $\mathrm{d}f = \partial_x f\,\mathrm{d}x + \partial_y f\,\mathrm{d}y + \partial_z f\,\mathrm{d}z$。现有一个空间分段可微曲线 $\Gamma(t) : (x(t), y(t), z(t))$，起点为 $(x(t_0), y(t_0), z(t_0))$，终点为 $(x(t_1), y(t_1), z(t_1))$。将该段曲线分成大量微小的小段，对每一小段皆有 $\Delta f \approx \partial_x f \Delta x + \partial_y f \Delta y + \partial_z f \Delta z$。所有小段求和，令最长的小段的长度趋于零，则和式变成积分，有

$$\int_{\{t_0,t_1\}} f(t) = f(t_1) - f(t_0) = \int_{\Gamma(t)} \partial_x f\,\mathrm{d}x + \partial_y f\,\mathrm{d}y + \partial_z f\,\mathrm{d}z,$$

其中 $\int_{\{t_0,t_1\}} f(t)$ 是对零度微分形式 $f(t)$ 的积分。上式是对微积分基本定理的一般化。由于势能是最典型的例子，所以此式称为势能定理。

记 $\operatorname{grad} f = \partial_x f \vec{e}_x + \partial_y f \vec{e}_y + \partial_z f \vec{e}_z$，称为标量场 f 的梯度。有 $\mathrm{d}f = \operatorname{grad} f \cdot \mathrm{d}\vec{s}$，这是势能定理的微分表达方式。上式又可写成

$$\int_{\{t_0,t_1\}} f(t) = \int_{\Gamma(t)} \operatorname{grad} f \cdot \mathrm{d}\vec{s},$$

所以，势能定理又称梯度定理。

注意到 $\Gamma(t_0)$ 和 $\Gamma(t_1)$ 是曲线 $\Gamma(t)$ 的两个端点，是该曲线的“零维边界”。因此，上式实际上是

$$\int_{\partial\Gamma} f = \int_{\Gamma} \mathrm{d}f,$$

其中的 $\partial\Gamma$ 表示 Γ 的边界。

8.7.7 格林定理

考虑 2 维 1 度微分形式 $\omega = P(x,y)\,\mathrm{d}x + Q(x,y)\,\mathrm{d}y$。因为

$$\mathrm{d}(P\,\mathrm{d}x) = \mathrm{d}P \wedge \mathrm{d}x, \quad \mathrm{d}(Q\,\mathrm{d}y) = \mathrm{d}Q \wedge \mathrm{d}y,$$

所以

$$
\begin{aligned}
\mathrm{d}\omega &= \mathrm{d}(P\,\mathrm{d}x + Q\,\mathrm{d}y) \\
&= \mathrm{d}P \wedge \mathrm{d}x + \mathrm{d}Q \wedge \mathrm{d}y \\
&= (\partial_x P\,\mathrm{d}x + \partial_y P\,\mathrm{d}y) \wedge \mathrm{d}x + (\partial_x Q\,\mathrm{d}x + \partial_y Q\,\mathrm{d}y) \wedge \mathrm{d}y \\
&= \partial_y P\,\mathrm{d}y \wedge \mathrm{d}x + \partial_x Q\,\mathrm{d}x \wedge \mathrm{d}y \\
&= (\partial_x Q - \partial_y P)\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}
$$

这是格林定理的微分表达方式。注意，式中的 $\mathrm{d}x\,\mathrm{d}y$ 是 $\mathrm{d}x \wedge \mathrm{d}y$ 的简写。

上式实际上可以分成 $\mathrm{d}(P\,\mathrm{d}x)$ 和 $\mathrm{d}(Q\,\mathrm{d}y)$ 两个独立的部分。先看

$$\mathrm{d}(P(x,y)\,\mathrm{d}x) = \partial_y P(x,y)\,\mathrm{d}y \wedge \mathrm{d}x.$$

对于平面上象椭圆这样的简单区域 D，我们用网孔长 Δx、宽 Δy 的细网，把区域 D 划分成很多矩形微元。边缘上的曲边三角形或梯形，对积分的影响在极限情况下趋于零，可以忽略不计。那么，上式表明，对区域 D 内一矩形网孔，设其左下角为 $P_0(x_0,y_0)$，按偏导数的定义，近似地有

$$P(x_0, y_0+\Delta y)\Delta x - P(x_0,y_0)\Delta x \approx (\partial_y P(x,y))\Big|_{(x_0,y_0)} \Delta y \Delta x.$$

固定 x_0，对一竖排网孔求和，有

$$P(x_0, y_{\sup}(x_0))\Delta x - P(x_0, y_{\inf}(x_0))\Delta x \approx \sum_{y=y_{\inf}(x_0)}^{y_{\sup}(x_0)} (\partial_y P(x,y))\Big|_{x=x_0} \Delta y \Delta x.$$

其中的 $y_{\inf}(x_0)$ 表示这一列最低的 y 值，而 $y_{\sup}(x_0)$ 表示这一列最高的 y 值。从左到右对所有的竖排再求和，有

$$
\begin{aligned}
&\sum_{x=x_{\inf}}^{x_{\sup}} P(x, y_{\sup}(x))\Delta x - \sum_{x=x_{\inf}}^{x_{\sup}} P(x, y_{\inf}(x))\Delta x \\
\approx &\sum_{x=x_{\inf}}^{x_{\sup}} \sum_{y=y_{\inf}(x)}^{y_{\sup}(x)} (\partial_y P(x,y))\Delta y \Delta x.
\end{aligned}
$$

极限情形下，该式就是积分

$$\int_{x_{\inf}}^{x_{\sup}} P(x, y_{\sup}(x))\,\mathrm{d}x - \int_{x_{\inf}}^{x_{\sup}} P(x, y_{\inf}(x))\,\mathrm{d}x = \iint_D (\partial_y P(x,y))\,\mathrm{d}x\,\mathrm{d}y.$$

如果规定沿区域的边界行进时，使得左手边为区域内部，这个行进方向为边界的正方向，那么上式第一项是反方向的，调整后恰为

$$-\int_{\partial D} P(x,y)\,\mathrm{d}x = \iint_D \partial_y P(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

注意到这里的“边界的正方向”恰与 $(\mathrm{d}x\vec{e}_x)\wedge(\mathrm{d}y\vec{e}_y)$ 的方向相同，所以

$$\iint_D \partial_y P(x,y)\,\mathrm{d}x\,\mathrm{d}y = -\iint_D \partial_y P(x,y)\,\mathrm{d}y\wedge\mathrm{d}x = -\iint_D \mathrm{d}P\wedge\mathrm{d}x.$$

综合上两式，得

$$\int_{\partial D} P\,\mathrm{d}x = \iint_D \mathrm{d}P\wedge\mathrm{d}x.$$

记 $\omega = P\,\mathrm{d}x$，有 $\mathrm{d}\omega = \mathrm{d}P\wedge\mathrm{d}x$，所以该式实为

$$\int_{\partial D}\omega = \iint_D \mathrm{d}\omega.$$

再看 $\mathrm{d}(Q\,\mathrm{d}y) = \partial_x Q\,\mathrm{d}x\,\mathrm{d}y$。按偏导数的定义，对上述网孔有近似关系

$$Q(x_0+\Delta x, y_0)\Delta y - Q(x_0,y_0)\Delta y \approx (\partial_x Q(x,y))\Big|_{(x_0,y_0)}\Delta x\Delta y.$$

固定 y_0，对一横行网孔求和，有

$$Q(x_{\sup}(y_0),y_0)\Delta y - Q(x_{\inf}(y_0),y_0)\Delta y \approx \sum_{x=x_{\inf}(y_0)}^{x_{\sup}(y_0)}(\partial_x Q(x,y))\Big|_{y=y_0}\Delta x\Delta y.$$

其中 $x_{\sup}(y_0)$ 表示这一行最大的 x 值，$x_{\inf}(y_0)$ 表示这一行最小的 x 值。从下到上对所有的横行再求和，有

$$\begin{aligned}&\sum_{y=y_{\inf}}^{y_{\sup}} Q(x_{\sup}(y),y)\Delta y - \sum_{y=y_{\inf}}^{y_{\sup}} Q(x_{\inf}(y),y)\Delta y\\ \approx&\sum_{y=y_{\inf}}^{y_{\sup}}\sum_{x=x_{\inf}(y)}^{x_{\sup}(y)}(\partial_x Q(x,y))\Delta x\Delta y.\end{aligned}$$

极限情形下，该式就是积分

$$\int_{y_{\inf}}^{y_{\sup}} Q(x_{\sup}(y),y)\,\mathrm{d}y - \int_{y_{\inf}}^{y_{\sup}} Q(x_{\inf}(y),y)\,\mathrm{d}y = \iint_D(\partial_x Q(x,y))\,\mathrm{d}x\,\mathrm{d}y.$$

此式第二项积分方向与边界方向相反，调整后得

$$\int_{\partial D} Q\,\mathrm{d}x = \iint_D \partial_x Q\,\mathrm{d}x\,\mathrm{d}y = \iint_D \mathrm{d}Q\wedge\mathrm{d}y.$$

记 $\eta = Q\,\mathrm{d}y$，有 $\mathrm{d}\eta = \mathrm{d}Q\wedge\mathrm{d}y$，所以该式实为

$$\int_{\partial D}\eta = \iint_D \mathrm{d}\eta.$$

由微分算符的线性性质，有 $\mathrm{d}(\omega+\eta) = \mathrm{d}\omega + \mathrm{d}\eta$，所以

$$\int_{\partial D}(\omega+\eta) = \iint_D \mathrm{d}(\omega+\eta) = \iint_D(\mathrm{d}\omega + \mathrm{d}\eta).$$

把 $\omega = P\,\mathrm{d}x, \quad \eta = Q\,\mathrm{d}y$ 代入，得

$$\int_{\partial D} (P\,\mathrm{d}x + Q\,\mathrm{d}y) = \iint_D (\partial_x Q - \partial_y P)\,\mathrm{d}x\,\mathrm{d}y.$$

这就是格林定理的积分表达方式。

记 $\zeta = P\,\mathrm{d}x + Q\,\mathrm{d}y$，则格林定理实际上是说

$$\int_{\partial D} \zeta = \iint_D \mathrm{d}\zeta.$$

8.7.8 斯托克斯定理与旋度

考虑 3 维 1 度微分形式 $\omega = P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z$。因为

$$\mathrm{d}(P\,\mathrm{d}x) = \mathrm{d}P \wedge \mathrm{d}x, \quad \mathrm{d}(Q\,\mathrm{d}y) = \mathrm{d}Q \wedge \mathrm{d}y, \quad \mathrm{d}(R\,\mathrm{d}z) = \mathrm{d}R \wedge \mathrm{d}z,$$

所以

$$\begin{aligned}
\mathrm{d}\omega &= \mathrm{d}(P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z) \\
&= \mathrm{d}P \wedge \mathrm{d}x + \mathrm{d}Q \wedge \mathrm{d}y + \mathrm{d}R \wedge \mathrm{d}z \\
&= (\partial_x P\,\mathrm{d}x + \partial_y P\,\mathrm{d}y + \partial_z P\,\mathrm{d}z) \wedge \mathrm{d}x \\
&\quad + (\partial_x Q\,\mathrm{d}x + \partial_y Q\,\mathrm{d}y + \partial_z Q\,\mathrm{d}z) \wedge \mathrm{d}y \\
&\quad + (\partial_x R\,\mathrm{d}x + \partial_y R\,\mathrm{d}y + \partial_z R\,\mathrm{d}z) \wedge \mathrm{d}z \\
&= (\partial_y P\,\mathrm{d}y \wedge \mathrm{d}x + \partial_x Q\,\mathrm{d}x \wedge \mathrm{d}y) + (\partial_z Q\,\mathrm{d}z \wedge \mathrm{d}y + \partial_y R\,\mathrm{d}y \wedge \mathrm{d}z) \\
&\quad + (\partial_z P\,\mathrm{d}z \wedge \mathrm{d}x + \partial_x R\,\mathrm{d}x \wedge \mathrm{d}z) \\
&= (\partial_x Q - \partial_y P)\,\mathrm{d}x\,\mathrm{d}y + (\partial_y R - \partial_z Q)\,\mathrm{d}y\,\mathrm{d}z + (\partial_z P - \partial_x R)\,\mathrm{d}z\,\mathrm{d}x
\end{aligned}$$

此式是斯托克斯定理的微分表达方式。

给定空间矢量场 $\vec{F}(x, y, z) = P\vec{e}_x + Q\vec{e}_y + R\vec{e}_z$，记

$$\operatorname{curl}\vec{F} = (\partial_y R - \partial_z Q)\vec{e}_x + (\partial_z P - \partial_x R)\vec{e}_y + (\partial_x Q - \partial_y P)\vec{e}_z,$$

称为该矢量场的旋度。也有记为 $\operatorname{rot}\vec{F}$ 的。

设空间中一片可定向曲面 M，其边界为 ∂M，上述矢量场沿边界的线积分，等于该矢量场的旋度在该曲面上的面积分。即

$$\int_{\partial M} \vec{F} \cdot \mathrm{d}\vec{s} = \iint_M \operatorname{curl}\vec{F} \cdot \mathrm{d}\vec{S}.$$

这是斯托克斯定理的积分表达方式。斯托克斯定理又称旋度定理。

记 $\omega = \vec{F} \cdot \mathrm{d}\vec{s}$，则 $\mathrm{d}\omega = \operatorname{curl}\vec{F} \cdot \mathrm{d}\vec{S}$, 所以斯托克斯定理实际上是说

$$\int_{\partial M} \omega = \int_M \mathrm{d}\omega.$$

8.7.9 高斯定理与散度

考虑 3 维 2 度微分形式 $\omega = P\,\mathrm{d}x\,\mathrm{d}y + Q\,\mathrm{d}y\,\mathrm{d}z + R\,\mathrm{d}z\,\mathrm{d}x$。有

$$\begin{aligned}\mathrm{d}\omega &= \mathrm{d}P \wedge \mathrm{d}x \wedge \mathrm{d}y + \mathrm{d}Q \wedge \mathrm{d}y \wedge \mathrm{d}z + \mathrm{d}R \wedge \mathrm{d}z \wedge \mathrm{d}x \\ &= \partial_z P\,\mathrm{d}z \wedge \mathrm{d}x \wedge \mathrm{d}y + \partial_x Q\,\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z + \partial_y R\,\mathrm{d}y \wedge \mathrm{d}z \wedge \mathrm{d}x \\ &= (\partial_x Q + \partial_y R + \partial_z P)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.\end{aligned}$$

这是高斯定理的微分表达方式。

给定矢量场 $\vec{F} = Q\vec{e}_x + R\vec{e}_y + P\vec{e}_z$，记

$$\operatorname{div}\vec{F} = \partial_x Q + \partial_y R + \partial_z P,$$

称为该矢量场的散度。

给定空间区域 Ω，记其边界为 $\partial\Omega$，则对上述矢量场 $\vec{F}$，有

$$\iint_{\partial\Omega} \vec{F} \cdot \mathrm{d}\vec{S} = \iiint_{\Omega} \operatorname{div}\vec{F}\,\mathrm{d}V,$$

其中 $\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$。这是高斯定理的积分表达方式。高斯定理又称散度定理。

记 $\omega = \vec{F} \cdot \mathrm{d}\vec{S}$，则 $\mathrm{d}\omega = \operatorname{div}\vec{F}\,\mathrm{d}V$，因此，高斯定理实际上是说

$$\int_{\partial\Omega} \omega = \int_{\Omega} \mathrm{d}\omega.$$

8.7.10 广义斯托克斯定理

给定 n 维可微流形中的 $k+1$ 维可微子流形上的有界区域 M，记其边界为 ∂M。如果 ∂M 是 k 维可微子流形，那么，n 维 k 度微分形式 ω 在边界 ∂M 上的积分，等于 $\mathrm{d}\omega$ 在区域 M 上的积分，即

$$\int_{\partial M} \omega = \int_{M} \mathrm{d}\omega.$$

这个称为广义斯托克斯定理。其证明与格林定理类似，增加度数并采用曲线坐标系，并不构成实质的困难。故不赘述。

梯度定理、旋度定理、散度定理分别是它在 $n=3,\ k=0,1,2$ 时的特例。因为我们的空间是三维的，所以这三个特例用得最多。

8.7.11 哈密顿算子与拉普拉斯方程

称 $\vec{\nabla} = \vec{e}_x\partial_x + \vec{e}_y\partial_y + \vec{e}_z\partial_z$ 为哈密顿算子，或矢量微分算子。

标量场的梯度 $\operatorname{grad} f = \vec{\nabla} f$。

矢量场的旋度 $\operatorname{curl}\vec{F} = \vec{\nabla}\times\vec{F}$。

矢量场的散度 $\operatorname{div}\vec{F} = \vec{\nabla}\cdot\vec{F}$。

因为 $\mathrm{d}^2 = 0$，对标量场 f，有 $\mathrm{d}(\mathrm{d}f) = 0$，表明标量场的梯度是无旋的；对矢量场 $\vec{F}$，有 $\mathrm{d}(\mathrm{d}\vec{F}) = 0$，表明矢量场的旋度是无源的。

标量场的梯度称为位势场。任何无旋的矢量场必是位势场。

矢量场的旋度称为螺线场。任何无源的矢量场必是螺线场。

称 $\triangle = \partial_{xx} + \partial_{yy} + \partial_{zz} = \vec{\nabla}^2 = \vec{\nabla}\cdot\vec{\nabla}$ 为拉普拉斯算子。$\triangle f(x,y,z) = 0$ 称为拉普拉斯方程。该方程的解称为调和函数。

如果标量场的梯度在所考察的区域内是无源的，那么该标量场在该区域内是调和的。比如，电势在无电荷区域内是调和的。

8.8 超越函数与平面度规

8.8.1 自然指数函数

定义 $e^x = 1 + x + \dfrac{1}{2!}x^2 + \cdots + \dfrac{1}{n!}x^n + \cdots$，称为自然指数函数。取 $x = 1$，该函数等于 e，约为 2.718281828。它和 π 一样，是个无理数，也是个超越数，称为自然常数。

易见，$e^{a+b} = e^a e^b$，$e^{-x} = \dfrac{1}{e^x}$。

e^x 是求导运算的不动点，即 $(e^x)' = e^x$。

8.8.2 双曲三角函数

定义
$$\cosh x = \frac{e^x + e^{-x}}{2}, \quad \sinh x = \frac{e^x - e^{-x}}{2},$$
分别称为双曲余弦函数和双曲正弦函数。

有 $e^x = \cosh x + \sinh x, \quad e^{-x} = \cosh x - \sinh x$。

有 $\cosh^2 x - \sinh^2 x = (\cosh x + \sinh x)(\cosh x - \sinh x) = e^x e^{-x} = 1$。

有 $\cosh(-x) = \cosh x, \quad \sinh(-x) = -\sinh x$。

有
$$\begin{aligned}
\cosh(a+b) &= \frac{1}{2}(e^{a+b} + e^{-a-b}) \\
&= \frac{1}{2}\Big((\cosh a + \sinh a)(\cosh b + \sinh b) \\
&\quad + (\cosh a - \sinh a)(\cosh b - \sinh b)\Big) \\
&= \cosh a\cosh b + \sinh a\sinh b, \\
\sinh(a+b) &= \frac{1}{2}(e^{a+b} - e^{-a-b})
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2}\Big((\cosh a+\sinh a)(\cosh b+\sinh b)\\
&\quad -(\cosh a-\sinh a)(\cosh b-\sinh b)\Big)\\
&= \sinh a\cosh b+\cosh a\sinh b.
\end{aligned}$$

易见，$(\cosh x)' = \sinh x, \quad (\sinh x)' = \cosh x$。

定义 $\tanh x = \dfrac{\sinh x}{\cosh x}, \quad \coth x = \dfrac{\cosh x}{\sinh x}$，分别称为双曲正切函数和双曲余切函数。

有

$$\begin{aligned}
\tanh(a+b) &= \frac{\sinh a\cosh b+\cosh a\sinh b}{\cosh a\cosh b+\sinh a\sinh b} = \frac{\tanh a+\tanh b}{1+\tanh a\tanh b},\\
\coth(a+b) &= \frac{\cosh a\cosh b+\sinh a\sinh b}{\sinh a\cosh b+\cosh a\sinh b} = \frac{1+\coth a\coth b}{\coth a+\coth b}.
\end{aligned}$$

有

$$\begin{aligned}
(\tanh x)' &= \frac{(\sinh x)'\cosh x-\sinh x(\cosh x)'}{\cosh^2 x} = \frac{1}{\cosh^2 x},\\
(\coth x)' &= \frac{(\cosh x)'\sinh x-\cosh x(\sinh x)'}{\sinh^2 x} = -\frac{1}{\sinh^2 x}.
\end{aligned}$$

因为 $\cosh^2 x-\sinh^2 x = 1$，所以

$$(\tanh x)' = 1-\tanh^2 x, \quad (\coth x)' = 1-\coth^2 x.$$

8.8.3 三角函数

定义

$$\cos x = \frac{e^{ix}+e^{-ix}}{2}, \quad \sin x = \frac{e^{ix}-e^{-ix}}{2i},$$

分别称为余弦函数和正弦函数。式中的 $i=\sqrt{-1}$ 是虚数单位。

有 $e^{ix} = \cos x+i\sin x, \quad e^{-ix} = \cos x-i\sin x$。

有 $e^{i\pi}+1=0$。这个式子包含 $0,1,i,\pi,e$ 五个最典型的数学常数，和加法、乘法、指数等三个最典型的运算，和表示平衡的等号，因此，被称为数学中最美丽的公式之一。

有 $\cos^2 x+\sin^2 x = (\cos x)^2-(i\sin x)^2 = e^{ix}e^{-ix} = 1$。

有 $\cos(-x) = \cos x, \quad \sin(-x) = -\sin x$。

有

$$\begin{aligned}
\cos(a+b) &= \frac{1}{2}(e^{ai+bi}+e^{-ai-bi})\\
&= \frac{1}{2}\Big((\cos a+i\sin a)(\cos b+i\sin b)\\
&\quad +(\cos a-i\sin a)(\cos b-i\sin b)\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \cos a\cos b-\sin a\sin b,\\
\sin(a+b) &= \frac{1}{2i}(e^{ai+bi}-e^{-ai-bi})\\
&= \frac{1}{2i}\Big((\cos a+i\sin a)(\cos b+i\sin b)\\
&\quad -(\cos a-i\sin a)(\cos b-i\sin b)\Big)\\
&= \sin a\cos b+\cos a\sin b.
\end{aligned}
$$

易见，$(\cos x)'=-\sin x,\quad (\sin x)'=\cos x$。

定义 $\tan x=\dfrac{\sin x}{\cos x},\quad \cot x=\dfrac{\cos x}{\sin x}$，分别称为正切函数和余切函数。

有

$$
\begin{aligned}
\tan(a+b) &= \frac{\sin a\cos b+\cos a\sin b}{\cos a\cos b-\sin a\sin b}=\frac{\tan a+\tan b}{1-\tan a\tan b},\\
\cot(a+b) &= \frac{\cos a\cos b-\sin a\sin b}{\sin a\cos b+\cos a\sin b}=\frac{\cot a\cot b-1}{\cot a+\cot b}.
\end{aligned}
$$

有

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

因为 $\cos^2 x+\sin^2 x=1$，所以

$$
(\tan x)'=1+\tan^2 x,\quad (\cot x)'=-(1+\cot^2 x).
$$

8.8.4 反双曲正切函数与反正切函数

如果 $x=\tanh y$，则记 $y=\operatorname{arctanh} x$，称为反双曲正切函数。

由于 $x=\tanh y=\dfrac{e^y-e^{-y}}{e^y+e^{-y}}=\dfrac{e^{2y}-1}{e^{2y}+1}$，即 $(e^{2y}+1)x=e^{2y}-1$，所以 $e^{2y}=\dfrac{1+x}{1-x}$，即

$$
y=\operatorname{arctanh} x=\frac{1}{2}\ln\frac{1+x}{1-x}.
$$

因为 $\dfrac{\mathrm{d}x}{\mathrm{d}y}=1-\tanh^2 y$，所以 $\dfrac{\mathrm{d}y}{\mathrm{d}x}=\dfrac{1}{1-\tanh^2 y}=\dfrac{1}{1-x^2}$。

如果 $x=\tan y$，则记 $y=\arctan x$，称为反正切函数。

由于 $x=\tan y=\dfrac{e^{iy}-e^{-iy}}{i(e^{iy}+e^{-iy})}=\dfrac{e^{2iy}-1}{i(e^{2iy}+1)}$，即 $(e^{2iy}+1)ix=e^{2iy}-1$，所以 $e^{2iy}=\dfrac{1+ix}{1-ix}$，即

$$
y=\arctan x=\frac{1}{2i}\ln\frac{1+ix}{1-ix}.
$$

因为 $\dfrac{\mathrm{d}x}{\mathrm{d}y}=1+\tan^2 y$，所以 $\dfrac{\mathrm{d}y}{\mathrm{d}x}=\dfrac{1}{1+\tan^2 y}=\dfrac{1}{1+x^2}$。

8.8.5 欧氏度规平面与闵氏度规平面

由勾股定理，单位圆上任意一点 $P(x_P, y_P)$ 满足 ${x_P}^2 + {y_P}^2 = 1$，这一关系恰与 $\cos^2\phi + \sin^2\phi = 1$ 相一致，令 $x_P = \cos\phi, \quad y_P = \sin\phi$，则可以用 ϕ 来刻画单位圆上的点。这就是所谓的角度。

把这个思路应用于 $\cosh^2\zeta - \sinh^2\zeta = 1$，令 $x_P = \cosh\zeta, \quad y_P = \sinh\zeta$，则有单位双曲线 ${x_P}^2 - {y_P}^2 = 1$，并且可用 ζ 来刻画单位双曲线上的点。这就是所谓的双曲角度。

点 $P(x, y)$ 到原点的普通距离 r 有 $r^2 = x^2+y^2$，双曲距离 s 有 $s^2 = x^2-y^2$。一般地，距离 ρ 可以定义为 $\rho^2 = g_{ij}x^ix^j$，其中 $g_{ij} = g_{ji}$。这样的 g_{ij} 称为黎曼度规张量。

对于欧氏距离，有 $g_{11} = g_{22} = 1, \quad g_{12} = g_{21} = 0$，称为欧几里德度规。配上这种度规的平面就称为欧氏度规平面。

对于双曲距离，有 $g_{11} = 1, \quad g_{22} = -1, \quad g_{12} = g_{21} = 0$，称为闵科夫斯基度规。配上这种度规的平面就称为闵氏度规平面。

通常，在一种几何中定义距离，就是把距离当作几何性质来讨论，这意味着距离是该几何主变换群作用下的不变量。所谓“几何的主变换群”，通常表现为特定类型的坐标变换构成的群。因此，通常可以认为，距离不随坐标系的选择而改变，即，距离是标量。

8.8.6 角度与弧长、双曲角度与双曲弧长

按上节的方法定义距离之后，普通的角度就可以被定义为单位圆上该角度所截的圆弧的弧长，而双曲角度可以被定义为单位双曲线上该角度所截的双曲线弧段的双曲弧长。既然弧长是标量，那么角度也是标量。简言之，保长必保角。

图 8.2 显示了两种角度的三角函数的区别。 图中粗线段所示就是相应函数

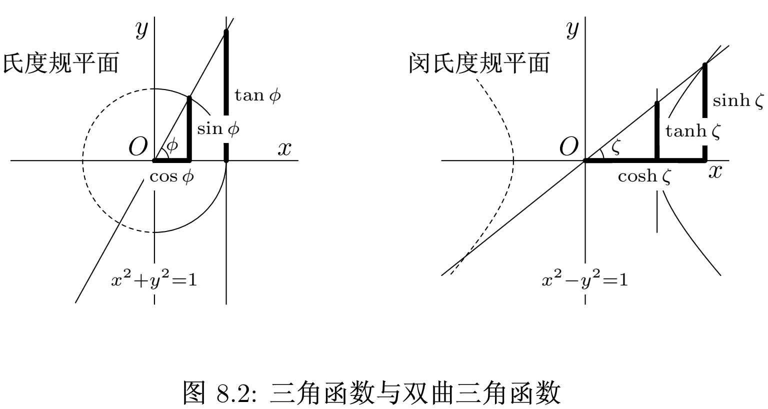

图 8.2: 三角函数与双曲三角函数

的值。从图上可以清楚地看出有如下变量代换：

$$\text{在欧氏度规平面上，}\quad x=\cos\phi,\quad y=\sin\phi,\quad \phi=\arctan\frac{y}{x}$$
$$\text{在闵氏度规平面上，}\quad x=\cosh\zeta,\quad y=\sinh\zeta,\quad \zeta=\operatorname{arctanh}\frac{y}{x}$$

8.8.7 平面曲线的高斯圆像曲率

前面讲过空间中曲面的高斯球面像曲率，是曲面上给定点的邻域内，闭合回路的单位球面像的有向面积与该回路的有向面积的比。

在平面上有类似的情况。如图 8.3 所示。

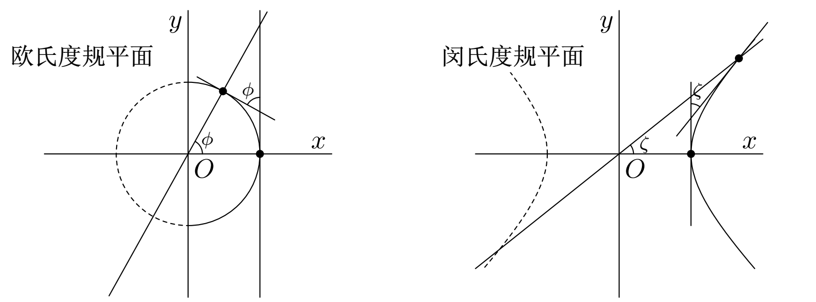

图 8.3: 高斯圆像曲率

设一动点从圆 $x^2+y^2=r^2$ 与 x 轴正半轴的交点出发，走一段有向弧长为 ϕr 的路程，过该动点的切线也偏转了角度 ϕ。后者比前者就是该圆的高斯圆像曲率，即 $\kappa=\dfrac{\phi}{\phi r}=\dfrac{1}{r}$。

设一动点从双曲线 $x^2-y^2=s^2$ 与 x 轴正半轴的交点出发，走一段有向双曲弧长为 ζs 的路程，过该动点的切线偏转了双曲角度 $-\zeta$。后者比前者就是该双曲线的高斯圆像曲率，即 $\kappa=\dfrac{-\zeta}{\zeta s}=-\dfrac{1}{s}$。

设一动点从直线 $x=a$ 与 x 轴的交点出发，走一段有向长度为 b 的路程，过该动点的切线偏转为 0，因此，该直线的高斯圆像曲率为 $\kappa=\dfrac{0}{b}=0$。

从这个意义上讲，圆是正曲率曲线，双曲线是负曲率曲线，直线是零曲率曲线。令 r 和 s 趋于无穷大，圆和双曲线都趋于直线。

因为闵氏度规应用于相对论，而后者又常谈到时空的负曲率弯曲，所以就容易产生一个错误的印象，认为闵氏度规刻画负曲率时空。其实不然。曲面和空间的曲率，与度规的导数有关，而欧氏度规和闵氏度规的所有分量都是常数，因此，曲率必定为零。即：欧氏平面和闵氏平面都是平直的。

8.8.8 把球面张到平面上

为了感受一下“不平直”的“平面”是个什么样子，我们通过南极投影，

把单位球面映射到赤道平面上。

设单位球面上除南极点之外的任意一点 $P(x,y,z)$，有球坐标 $(1,\theta,\phi)$，和坐标变换关系 $x=\sin\theta\cos\phi,\quad y=\sin\theta\sin\phi,\quad z=\cos\theta$。通过南极投影，映射到赤道平面上，因为圆周角等于圆心角的一半，所以有 $u=\tan\dfrac{\theta}{2}\cos\phi,\quad v=\tan\dfrac{\theta}{2}\sin\phi$。

单位球面上，一段曲线微元的长度满足 $\mathrm{d}s^2=\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2$，把球坐标代入，得 $\mathrm{d}s^2=\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\phi^2$。把映射关系代入，得 $\mathrm{d}s^2=\dfrac{4}{(1+u^2+v^2)^2}(\mathrm{d}u^2+\mathrm{d}v^2)$，即 $g_{ij}=\dfrac{4}{(1+u^2+v^2)^2}\delta_{ij}$。

在这个平面上，测地线不是直线，而是“过径圆”。所谓过径圆，是指该圆与单位圆有两个交点，而这两个交点的连线恰是单位圆的直径，就是说，该圆通过单位圆的一条直径的两个端点。

规定单位圆本身也算一个过径圆。这样，给定该平面上相异两点，可做且只可做一个过径圆与该两点接合。两点之间的最短距离是连接这两点的过径圆的劣弧的弧长。注意这个弧长不是欧氏几何中圆的弧长，而是按照上述度规 g_{ij} 积分出来的弧长。实际上，按照上述度规积分，所有过径圆的周长都是 2π。

该平面是“不平直”的，其标量曲率 $R=2$，高斯曲率为 $R/2=1$。以原点为圆心，过点 $P(u,v)$ 的圆的半径是 θ，满足 $\tan^2\dfrac{\theta}{2}=u^2+v^2$，即 $\theta=2\arctan\sqrt{u^2+v^2}$；周长是 $2\pi\sin\theta$。圆的周长半径比为 $2\pi\dfrac{\sin\theta}{\theta}<2\pi$。

这些结果并不奇怪。因为过径圆就是大圆的映像。这个平面实际上就是球面，是“张到了平面上”的球面。这个例子也说明了黎曼几何的威力。黎曼几何以 $\mathrm{d}s^2=g_{ij}x^iy^j=h_{\alpha\beta}u^\alpha v^\beta$ 为基本前提。当坐标系从 x^i 变换到 u^α 时，只要度规张量也从 g_{ij} 变换到 $h_{\alpha\beta}$，从而使得弧长微分 $\mathrm{d}s$ 是标量，在坐标系变换下保持不变，那么，黎曼几何所表达的流形的各种几何性质就与所选择的坐标系无关。

所谓坐标系的变换，就象上述球面上的 (θ,ϕ) 坐标，变换到平面上的 (u,v) 坐标一样，可以形象地看做是曲面的某种拉伸、扭曲。黎曼几何要求这种扭曲要保持点与点之间所定义的距离不变。

8.8.9 把闵氏双叶双曲面张到平面上

为了对比，我们再看一个负曲率曲面的例子。如图 8.4 所示。 把这个图竖起来绕 y 轴旋转一周，就做成单位双叶双曲面。出于习惯，把这个 y 轴改称 z 轴。则该双曲面的方程为 $z^2-x^2-y^2=1$。该空间的曲线微元的弧长为 $\mathrm{d}s^2=\mathrm{d}z^2-\mathrm{d}x^2-\mathrm{d}y^2$，因此是闵氏度规空间。图上点 Q 到北极点 $N(0,1)$ 的距离被定义为双曲线弧段 NQ 的闵氏弧长，也就是双曲角度 η。

如图做南极投影，由于周角是心角的一半，所以 $\lambda=\tanh\dfrac{\eta}{2}$。这样，双叶双曲面的上半片被映射到 xy-平面上单位圆 $x^2+y^2=1$ 的内部，而下半片被映

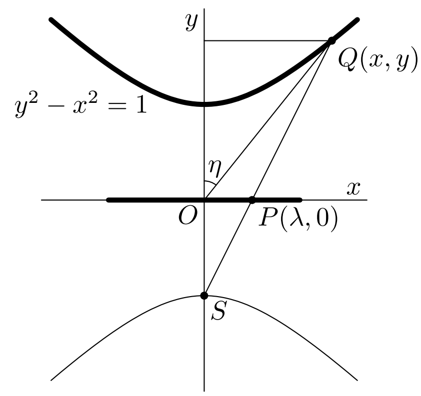

图 8.4: 把闵氏双叶双曲面张到平面上

射到该单位圆的外部。南极点 S 被映射到复平面的无穷远点。

双叶双曲面上的测地线是“大双曲线”，即过原点的平面与双叶双曲面的交线。大双曲线在南极投影下，被映射为与单位圆正交的圆。

取双曲面上的极坐标 (η,ϕ)，有 $x=\sinh\eta\cos\phi,\quad y=\sinh\eta\sin\phi,\quad z=\cosh\eta$。所以，$\mathrm{d}s^2=-\mathrm{d}\eta^2-\sinh^2\eta\,\mathrm{d}\phi^2$。设点 $Q(\eta,\phi)$ 被映射到平面上点 $P(u,v)$ 处，则 $u=\tanh\dfrac{\eta}{2}\cos\phi,\quad v=\tanh\dfrac{\eta}{2}\sin\phi$。这样，
$\mathrm{d}s^2=-\dfrac{4}{(1-u^2-v^2)^2}(\mathrm{d}u^2+\mathrm{d}v^2)$，即 $g_{ij}=-\dfrac{4}{(1-u^2-v^2)^2}\delta_{ij}$。

坐标系 (u,v) 配这样的度规 g_{ij}，可以解出该曲面的标量曲率为 $R=-2$，高斯曲率为 $R/2=-1$。以原点为圆心，过点 $P(u,v)$ 的圆的半径为 $\eta=2\,\mathrm{arctanh}\sqrt{u^2+v^2}$，周长为 $2\pi\sinh\eta$，圆的周长半径比为 $2\pi\dfrac{\sinh\eta}{\eta}>2\pi$。

8.9 圆几何

8.9.1 球极投影是保角的

如图 8.5 所示： 从球的北极点 N 发出的射线把球面上任意一点 P 投影到底面的点 P'。我们把 P 点的切面与顶面的交线称为上棱，与底面的交线称为下棱。考虑所有过直线 NP' 的平面。图中画了两个这样的平面。一个平面交上棱于 S_1，交下棱于 T_1。由于顶面平行于底面，所以 $NS_1\parallel T_1P'$，从而 $\triangle NS_1P\sim\triangle P'T_1P$。由球面的对称性质，可知 N 和 P 关于上棱与球心确定的平面成镜面反射对应，所以 $NS_1=PS_1$。从而 $PT_1=P'T_1$。另一个平面交上棱于 S_2，交下棱于 T_2。同理可知，$NS_2\parallel T_2P'$，$\triangle NS_2P\sim\triangle P'T_2P$，$NS_2=PS_2$，从而 $PT_2=P'T_2$。而 T_1T_2 是公共边，所以 $\triangle PT_1T_2\cong\triangle P'T_1T_2$，进而 $\angle T_1PT_2=\angle T_1P'T_2$。也就是说，在 P 点切面上，过 P 点的任意两直线之间的

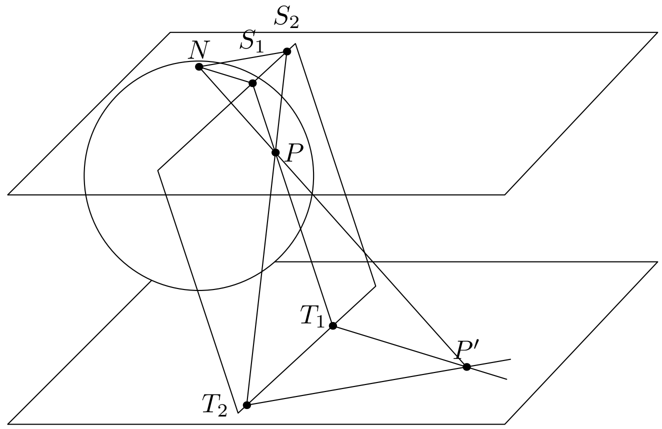

图 8.5: 球极投影是保角的

夹角如果是 θ，那么它们的影子之间的夹角也必定是 θ。所以，球极投影是保角的。 □

8.9.2 球极投影是保圆的

如图 8.6 所示： 球面上有一个圆，P 是其上任意一点，P' 是该点在底面

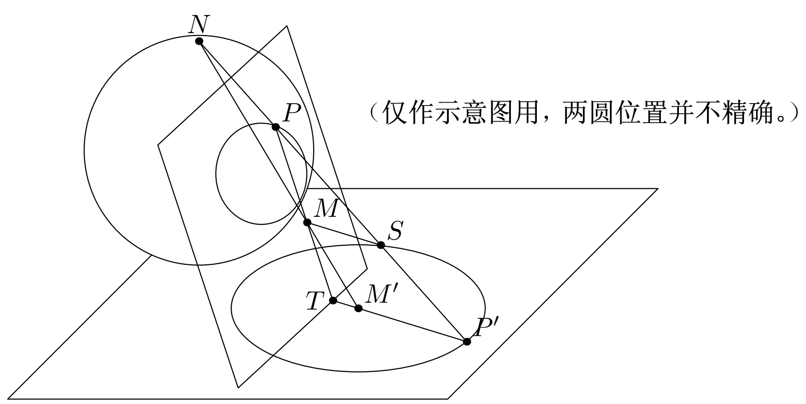

图 8.6: 球极投影是保圆的

上的投影。图中倾斜的平面是 P 点的切面。该切面与底面的交线仍旧称为下棱。当 P 点在圆上移动时，其切面也随之改变，并‘包络’出一个圆锥。该圆锥顶点为 M。该点不与 NP 共线，因而确定出一个过 NP' 的平面。该平面与下棱交于 T。前面我们已经证明，$PT = P'T$，因而 $\angle TPP' = \angle TP'P$。在此三角形中，过 M 做 $MS \parallel TP'$，则 $\angle MSP = \angle TP'P = \angle MPS$，因而 $MP = MS$。设 M' 是 M 在底面上的投影，则 $\dfrac{M'P'}{NM'} = \dfrac{MS}{NM}$，即 $M'P' =$

$\frac{NM'}{NM}MP$。当 P 点在圆上移动时，该式右边三个量都固定不变，因而 $M'P'$ 对所有 P' 都是相同的值。也即，所有 P' 点构成以 M' 为圆心的圆。 □

8.9.3 球面上的三种圆几何

三维欧氏空间中给定一个球面。若有平面与之相交，则交线必为圆。给定一个圆为基准圆，则

- 与基准圆正交的所有圆，在基准圆内部和外部的弧，各构成一个双曲几何。这些弧称为正交弧。该双曲几何中，过给定正交弧外一点可做两条正交弧与之相切，分别称为给定正交弧的过该点的左平行线和右平行线。过该点另外可做无数条正交弧与给定正交弧不相交。正交弧构成的三角形，内角和恒小于 π。
- 圆的直径的两个端点称为对径点。过基准圆对径点的所有圆，在基准圆内部和外部的弧，各构成一个椭圆几何。这些弧称为过径弧。该椭圆几何中，过给定过径弧外一点不可做过径弧与之相切，但是可做且只可做一条过径弧与之过相同的基准圆对径点，称其为给定过径弧的过该点的平行线。过径弧构成的三角形，内角和恒大于 π。
- 当基准圆缩为一点时，上述两种几何都过渡为抛物几何。基准圆外部的正交弧或过径弧，过渡为过同一个点的圆，称为共点圆。过给定共点圆外一点，可做且只可做一个共点圆与之相切，称其为给定共点圆的过该点的平行线。共点圆构成的三角形，内角和恒等于 π。

8.9.4 复平面上的测地圆几何

球极投影可以把整个欧氏平面映射到球面上除北极点之外的所有区域。在欧氏平面上添加一个无穷远点，与北极点对应，则构成所谓的复平面。复平面上的点与球面上的点在球极投影下做成一一映射。

把复平面上的直线看做过无穷远点的半径无限大的圆。则易见复平面上的圆与球面上的圆在球极投影下做成一一映射。

球面上的三种圆几何经球极投影映射到复平面上，分别称为复平面上的正交弧几何、过径弧几何和共点圆几何。三种圆，即正交圆、过径圆和共点圆，统称测地圆。它们都是相应几何中的测地线。所以，复平面上的三种几何也统称为测地圆几何。

当复平面上一个基准圆的圆心在有限范围内，而半径无限增大时，在球极投影映射下，球面上对应的基准圆反而是收缩的，最终收缩为北极点。此时复平面上基准圆内部的正交弧或过径弧都过渡为直线。从球面上看，这是过北极

点的共点圆构成的抛物几何。由于北极点对应复平面的无穷远点，所以，如果去掉这个点，该抛物几何就恢复成传统的欧氏几何。

8.9.5 正交弧几何

现在我们来看看这三种几何长什么样。如图 8.7 所示： C_b 是基准圆。C_0

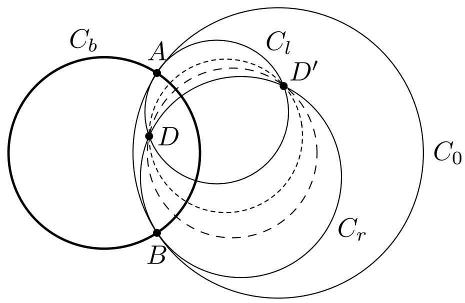

图 8.7: 正交弧几何中的平行线和不相交线（一）

是给定的任意一个正交圆，它与基准圆正交于 A 点和 B 点。D 为 C_0 外任意一点。过该点可做一个正交圆 C_l 与 C_0 相切于 A，称为左平行线。过 D 点还可做一个正交圆 C_r 与 C_0 相切于 B，称为右平行线。过 D 点还可做无限多个正交圆与 C_0 不相交，我们用虚线画了其中的两个作为代表。

注意，每个正交圆都被基准圆分割为两段，一段是基准圆内部的劣弧，另一段是基准圆外部的优弧。所有劣弧在基准圆内部构成双曲几何，所有优弧在基准圆外部也构成双曲几何。两者之间只相差一个反演变换。在这个意义上，可以认为两者完全等价。比如 D 点与 D' 点就是反演等价的。

图 8.8 显示了 D 在 C_0 ‘另一侧’时的情形。 所谓正交弧的‘两侧’，是

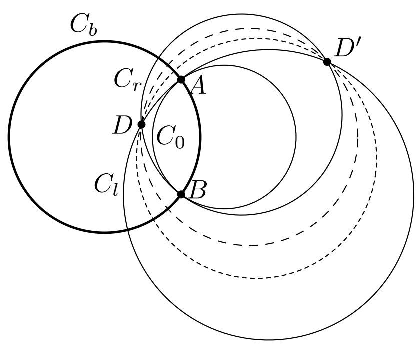

图 8.8: 正交弧几何中的平行线和不相交线（二）

指正交弧把基准圆内部分成两个部分，每个部分算一侧。

在所有过 D 点的正交圆中，除了与 C_0 相切的和不相交的，剩下的当然都是与 C_0 在有限远处相交的。两个这样的正交圆与 C_0 在基准圆内部和外部各构成一个双曲三角形。如图 8.9 所示。 基准圆内部的三角形为 $\triangle ABD$，基准圆

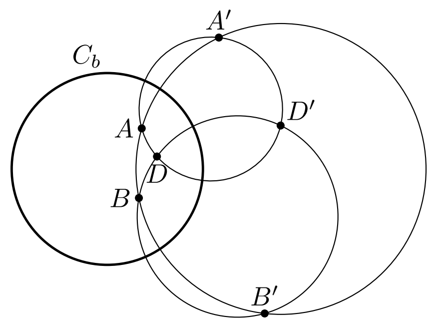

图 8.9: 正交弧几何中的三角形

外部的三角形为 $\triangle A'B'D'$。尽管后者看起来很怪异，但它的的确确是个三角形。实际上，它与前者互为对方的反演。

双曲三角形的内角和小于 π。举例来说，在图 8.7 中，C_0 与 C_l 和 C_r 也可以看做是一个双曲三角形，只是 $\angle A$ 和 $\angle B$ 都等于零而已。这样，该三角形的内角和就等于 $\angle D$，而 $\angle D$ 显然是小于 π 的。

8.9.6 反演变换

如图 8.10 所示。 如果基准圆的半径为 r，则 $OA * OA' = r^2$，$OB * OB' =$

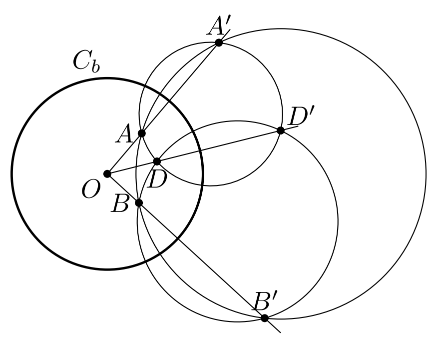

图 8.10: 正交弧几何中的反演变换

r^2，$OD * OD' = r^2$。实际上，基准圆内部除 O 点外的任何一个点 X，都可以按照 $OX * OX' = r^2$ 的关系在基准圆外部找到对应点 X'。当然，反过来也一样，给定 X' 可以找到对应点 X。对于 O 点，指定无穷远点与之对应。这样，

整个基准圆的内部就可以按此对应关系变换到外部，而外部则变换到内部，基准圆上的每一个点都变换到自身。这个变换，称为关于基准圆的反演变换。

反演变换是一种十分重要的保圆变换。它把任意一个正交圆在基准圆内部的劣弧与该圆在基准圆外部的优弧交换，从而保持该圆整体不变。也就是说，反演变换保持所有正交圆都不变。因此，整个图 8.9 在关于基准圆的反演变换下都是不变的。

如果我们取一个正交圆，设其圆心是 O，半径是 r，则同理可以定义关于该正交圆的反演变换。我们前面讲过，正交圆把基准圆内部的区域分为两个部分，称为该正交弧的两侧。显见，关于该正交圆的反演变换把这样的两侧相互交换。

复平面上一般的保圆变换可以用复变函数中的分式线性变换来表达。即：

$$w = \frac{az+b}{cz+d} \quad \text{或} \quad w = \frac{a\overline{z}+b}{c\overline{z}+d} \quad ；\text{其中} \quad ad-bc \neq 0$$

关于单位圆的反演则可以表示为 $w = 1/\overline{z}$。

可证复平面上任意一个保圆变换可以表为最多三个反演变换的合成。特别的，如果只考虑保持基准圆不变且把基准圆内部仍旧变换为基准圆内部的保圆变换，那么，这样的变换可以表为最多三个关于正交圆的反演变换的合成。

如图 8.11 所示。 当基准圆过无穷远点的时候，它就变成了一条直线。正

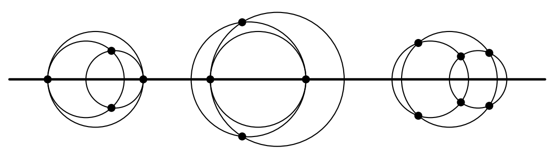

图 8.11: 基准圆过无穷远点时的正交弧几何

交圆就变成了圆心在该直线上的圆。关于基准圆的反演就变成了关于该直线的反射。尤其值得一提的是，复数的共轭，就是以实轴为基准圆的反演。

8.9.7 过径弧几何

图 8.12 展示了过径弧几何中的平行线。 其中 C_b 是基准圆，C_0 是给定的任意一个过径弧，P 和 P' 是它所过的对径点。D_1 是 C_0 外任意一点。从图上不难看出，过 D_1 可做且只可做一条过径弧与 C_0 过相同的对径点。这条过径弧就是 C_0 的过 D_1 的平行线。图上还画了 C_0 的过 D_2 的平行线。

图 8.13 展示了过径弧几何中的三角形。 与正交弧几何类似的是，每个过径圆也被基准圆分为两部分：一部分是劣弧，在基准圆内部；一部分是优弧，在基准圆外部。它们在基准圆内部和外部各自构成椭圆几何，并且两者之间只

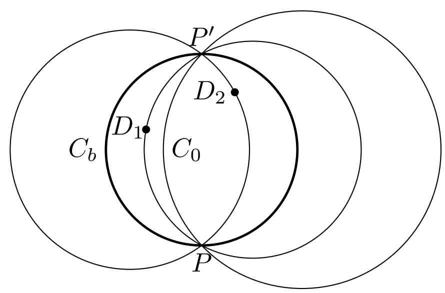

图 8.12: 过径弧几何中的平行线

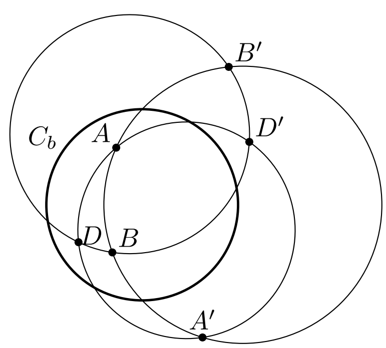

图 8.13: 过径弧几何中的三角形

相差一个对径变换。在这个意义上，也可以认为两者是等价的。比如，A 与 A' 是对径等价的，B 与 B' 是对径等价的。

关于基准圆的对径变换把基准圆上的对径两点互换，因此保持基准圆不变。把基准圆内部与外部互换。把过径圆在基准圆内部的劣弧与在基准圆外部的优弧互换，因而保持所有过径圆不变。实际上，整个图 8.13 在关于基准圆的对径变换下不变。

必须强调：对径中心与圆心是两个有微妙联系又不相同的概念，我们已经知道，在适当的球极投影下，每个过径圆都与球面上的一个大圆相对应。对径中心是球心的投影，而球心是所有大圆的圆心。所以，对径中心是过径圆所对应的球面大圆的圆心的投影。这与过径圆自己的圆心是有区别的。由于基准圆对应赤道，而赤道是唯一水平的大圆，所以基准圆的圆心恰好与对径中心重合。对于其他的过径圆，对径中心都是偏离圆心的。

过径弧几何中的‘对径点’这个概念，被定义为通过对径中心的直线与圆的两个交点，而并不是通过圆心的直线与圆的两个交点。这与球面上的‘对径点’的概念，表面上看似乎有很大区别，但实际上却是相合的。球面上，两个大圆必交于一直径，我们称这两个大圆是共径的。他们在球极投影下的像，是两个相交的圆，所交的两个交点的连线，必通过对径中心。因此，我们也称这样的两个相交的圆是‘共径’的。如果两个圆共径，那么他们的地位也是平等的。当过径圆过基准圆的一对对径点的时候，基准圆也过过径圆的一对对径点。

过径弧三角形内角和大于 π。举例来说，在图 8.12 中，我们可以加一条过基准圆圆心的水平线。这是个半径无限大的过径弧。它与 C_0 及其某条平行线可以构成一个三角形。但是 C_0 和该平行线都与水平线正交，仅这两个角的和就已经等于 π 了。而 C_0 和该平行线并不相切，而是成一个大于零的夹角。所以，该三角形的内角和显然大于 π。

8.9.8 对径变换

毫不奇怪，对径变换与反演变换存在内在的联系。如图 8.14 所示。 A'' 是 A 的关于基准圆圆心的中心对称点，同时又是 A' 的反演变换对应点。也就是说，对径变换是关于基准圆的反演变换和关于基准圆圆心的中心对称变换的合成。先做反演再做中心对称，和先做中心对称再做反演，得到的结果是一样的。

过基准圆圆心，做两个相互正交的直线，那么一个中心对称变换等于先做关于一个直线的反射，再做关于另一个直线的反射。这两个直线，实际上都是过无穷远点的正交圆。而关于正交圆的反演把基准圆内的点从正交弧的一侧变换到另一侧。尤其是当正交圆过无穷远点而变成直线的时候，关于该正交圆的反演就变成了关于该直线的反射。因此，对径变换实际上等于三个反演变换的

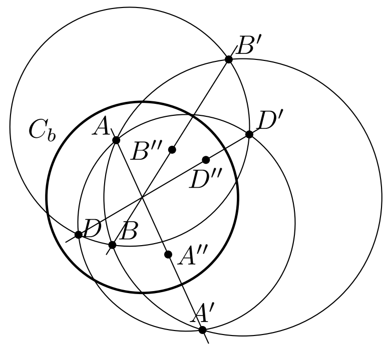

图 8.14: 过径弧几何中的对径变换

合成。我们在前面讲过，任何保圆变换都可以表为最多三个反演变换的合成。对径变换为我们提供了一个很好的例证。

8.9.9 双曲几何的布安加雷模型

反演本质上也是反射，是关于‘赤道面’的反射。设想把复平面水平放置。给定复平面上一个基准圆，设其半径为 r，那么，我们可以在该基准圆圆心位置，放置一个直径为 r 的球面。从该球面最高点做球极投影下来，正好把球面的赤道投影为底面上的基准圆。

我们来证在球面上关于赤道面反射对称的两个点，经球极投影后，它们在复平面上的对应点关于基准圆反演对称。如图 8.15 所示：P 和 P' 关于球面

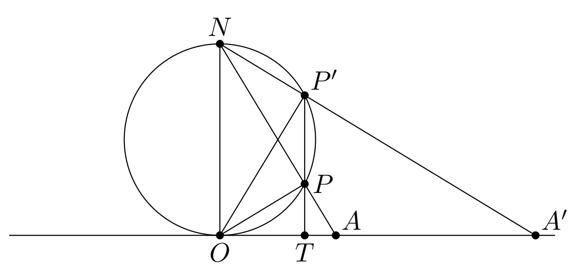

图 8.15: 关于基准圆的反演本质上是关于赤道面的反射

的赤道面反射对称。它们与球面北极点 N 确定一个竖直的平面。图上显示的就是该竖直平面上的情形。A 和 A' 分别是 P 和 P' 在底面上的投影。由于 P' 与 P 关于赤道面反射对称，所以 P' 到顶面（即过 N 点的水平面）的距离等于 P 到底面的距离，即 $P'T+PT=NO$。又由于 P 在球面上，所以 P 到赤道面的垂足，垂足到球心，球心回到 P 构成一个直角三角形。根据勾股定理，有 $\dfrac{(TP'-TP)^2}{4}+OT^2=\dfrac{NO^2}{4}=\dfrac{(TP'+TP)^2}{4}$，解得 $TP*TP'=OT^2$。由

于 P' 在球面上，所以 $OP' \perp NA'$，易证 $\triangle NOA' \sim \triangle OTP'$。设想把 $\triangle NOA'$ 绕 O 点旋转 $90°$，也就是把它扶起来，显见 $\frac{NO}{OT} = \frac{OA'}{TP'}$。同理，$OP \perp NA$，$\triangle NOA \sim \triangle OTP$，$\frac{NO}{OT} = \frac{OA}{TP}$。两式相乘，得 $\frac{NO^2}{OT^2} = \frac{OA' * OA}{TP' * TP}$。已证 $TP * TP' = OT^2$，所以 $OA' * OA = NO^2$。由于 NO 恰为底面上基准圆的半径，所以 A 与 A' 关于基准圆反演对应。

实际上，在水平面上放一球面，用一个竖直的平面来剖分该球面，允许剖成大小不等的两部分，则剖面边缘必定是个与赤道正交的圆。该圆一半在南半球，一半在北半球，两者关于赤道面反射对称。在球极投影下，这个与赤道正交的圆投影为与基准圆正交的圆，它在南半球的一半投影为正交圆在基准圆内部的劣弧，在北半球的一半投影为正交圆在基准圆外部的优弧。两者关于基准圆反演对称。反过来，任何一个正交圆都可以确定一个竖直的平面。这样，竖直平面与正交圆之间一一对应。这被称为双曲几何的布安加雷模型。该模型充分体现了反射与反演的关系。

在球心上做直角坐标系，把球面上任意一点关于三个坐标面连续做三次反射变换，就等于做了一次对径变换。即，对径变换等于关于三个坐标面的反射变换的合成。在球极投影下，每个坐标面与球面交成的圆被投影为复平面上的基准圆，因此，也可以说，对径变换等于关于三个坐标面所对应基准圆的反演变换的合成。

8.9.10 四圆坐标

如果把球面做一个就地转动，使得 N 点不在任何一个坐标轴上，那么三个坐标面就对应三个相互正交的真正的圆了，称其为坐标圆。从而立体的直角坐标系就变成了平面的正交坐标圆系。设球面上一点 P 的直角坐标是 (x, y, z)，三个坐标必须满足约束方程 $x^2 + y^2 + z^2 - r^2 = 0$。三个坐标面的方程分别是 $x = 0$、$y = 0$ 和 $z = 0$。球极投影到复平面，则复平面上任何一点，都可以配上三个数 (x, y, z)，它们满足基本关系 $x^2 + y^2 + z^2 - r^2 = 0$，其中 r 是给定的非零常数。因此三个数中只有两个是独立的。所有满足 $x = 0$ 的点构成一个圆，这就是坐标面 $x = 0$ 对应的坐标圆。同理，$y = 0$ 和 $z = 0$ 也各自确立一个坐标圆。

真正有趣的是，我们可以进一步扩展这个模型，使得 r 成为与 x, y, z 地位平等的变量，因而可以取零，同时允许四个变量都可以取复数。这样，$r = 0$ 也确立一个坐标圆，不过这是个‘虚圆’，无法在实的复平面上和其他三个坐标圆一起显示。同时我们约定，取任意非零实数 ρ，则 (x, y, z, r) 与 $(x', y', z', r') = (\rho x, \rho y, \rho z, \rho r)$ 表示同一个点。这个约定称为齐次等价条件。由此，四个复数变量，两个复数条件，正好可以用来刻画复数二维的复平面上的点。称这样的四个复变量为复平面上点的四圆坐标。

8.9.11 共点圆几何

图 8.16 展示了共点圆几何中的平行线。 其中 C_b 为缩成一点的基准圆。C_0

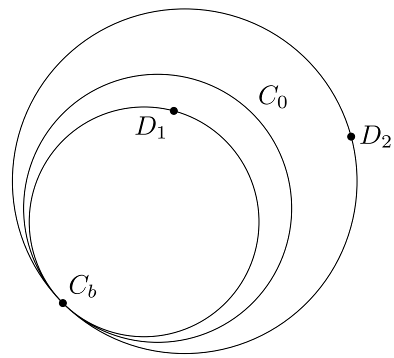

图 8.16: 共点圆几何中的平行线

为给定的任意一个共点圆。D_1 为 C_0 外任意一点。显见，过该点可做且只可做一个共点圆与 C_0 相切。该共点圆称为 C_0 的过 D_1 的平行线。图上还画了 C_0 的过 D_2 的平行线。

图 8.17 展示了共点圆几何中的三角形。

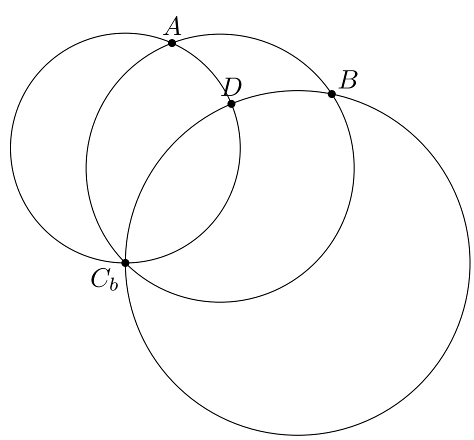

图 8.17: 共点圆几何中的三角形

共点圆几何中，三角形的内角和等于 π。首先，做一个球极投影，把复平面上共点圆三角形投影到球面上。然后，把球面做一个就地的转动，把共点圆所共之点，即 C_b，转到北极点的位置。再做球极投影，把球面上的三角形重新投影到复平面上。这样，共点圆三角形就变成了欧氏几何的直线三角形，其内角和当然等于 π。而我们已经知道，两次球极投影和一次球面就地转动，都不改变三个角的角度。

8.9.12 配极

三种几何之间的关系，还有另一种更为深刻的表述方式：给定空间任一球面，再给定空间任一点，则过该点的所有平面与球面的交线称为测地圆。如果平面与球面真的相交，测地圆就是实圆；如果相切，测地圆就是点圆；如果不相交，测地圆就是虚圆。而且，

- 如果给定点在给定球面外部，那么所有测地圆构成正交圆几何；
- 如果给定点在给定球面上，那么所有测地圆构成共点圆几何；
- 如果给定点在给定球面内部，那么所有测地圆构成过径圆几何。

图 8.6 显示的是第一种情况，M 是给定点。过 M 可做一圆锥与球面相切。所有切点构成球面的一个小圆。该小圆在底面上的投影是基准圆。过 M 的平面如果与球面相交，则交圆必与切点小圆正交，其在底面上的投影便是正交圆。所以，该情形刻画了一个正交圆几何。令 M 向 M' 移动，并越过 M' 移到无穷远，则切点小圆就成为球面的大圆，过 M 的平面都垂直于该大圆。这便是布安加雷模型。

令 M 向 N 移动，移动到与球面相交时，切点小圆就缩并成 M 本身。显见，该情形刻画了一个共点圆几何。

令 M 继续向 N 移动，进入球面内部，则所有过 M 的平面都与球面交成实圆。这些实圆在底面上的投影构成过径圆几何。注意，所有过径圆中，有且只有一个过径圆的圆心与过径中心重合，该过径圆便是基准圆。在图 8.6 中，切点小圆所在的平面与直线 NM' 有一交点。当 M 移动到该交点位置时，切点小圆就是过 M 的一个平面与球面的交圆，其投影恰满足圆心与过径中心重合之条件，因此，其投影就是该过径圆几何的基准圆。

球面是空间中的二次曲面。空间中的点与平面之间，关于给定的二次曲面，存在一种映射关系叫做配极。具体到球面，对于球面外部任意一点 M，可做且只可做一个切点圆。该切点圆所在之平面，称为 M 关于该球面的配极平面，简称极面。相应地，M 被称为该平面的配极点，简称极点。显然，当且仅当 M 在球面上的时候，极点和极面才会‘接合’，即‘极点在极面上’，或者说‘极面过极点’，此时称该极点或极面是自共轭的。注意，此时极点便是切点，极面便是切面，因此，配极映射是求导运算的一种扩展。球面本身可以被刻画成所有自共轭点的集合，或者所有自共轭平面的包络。

配极有一个美妙的性质：如果点 A 在点 B 的极面上，那么点 B 一定也在点 A 的极面上。因此称点 A 和 B 是共轭点对。易证，过给定点的所有平面，他们的配极点恰可构成一个平面。这个平面就是给定点的配极平面。当给定点在给定球面内部时，其配极平面一定在球面外部，与球面不相交。要做出这个配极平面，只需在过给定点的所有平面中任选三个不共线的平面，则相应的三个配极点也不共线，因而可以确定一个平面。该平面就是给定点的配极平面。

我们可以用配极来寻找过径圆几何的基准圆。给定球极投影框架下，给定球面内部一点 M，北极点 N 与点 M 的连线交底面于点 M'，该点就是过径中心。点 M 的配极平面与直线 NM 必有一个交点，该交点的配极平面过点 M 且与球面有实的交圆。该交圆投影到底面上，就是该过径圆几何的基准圆。

正交弧几何的基准圆过无穷远点时，基准圆变成直线，正交圆变成圆心在该直线上的圆。过径弧几何则不然。从图 8.6 中可以看出，当切点小圆过北极点时，直线 NM 必然处于水平位置，原本在球面内部的‘给定点’此时与北极点重合，从‘球面内’变到了‘球面上’。所以，当过径弧几何的基准圆过无穷远点时，过径弧几何就变成了欧氏几何。

8.10 测地复数与万能时空维度公式

8.10.1 对数测地变换

在狭义相对论中，考虑如下变量代换：

$$\begin{cases} ct = \exp\xi\cosh\eta \\ x = \exp\xi\sinh\eta \end{cases}$$

不难验证：

$$\begin{aligned}
\begin{bmatrix} \mathrm{d}(ct) & \mathrm{d}x \\ \mathrm{d}x & \mathrm{d}(ct) \end{bmatrix} &= \begin{bmatrix} \exp\xi & 0 \\ 0 & \exp\xi \end{bmatrix} \begin{bmatrix} \cosh\eta & \sinh\eta \\ \sinh\eta & \cosh\eta \end{bmatrix} \begin{bmatrix} \mathrm{d}\xi & \mathrm{d}\eta \\ \mathrm{d}\eta & \mathrm{d}\xi \end{bmatrix} \\
&= \exp\left(\xi \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \exp\left(\eta \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) \begin{bmatrix} \mathrm{d}\xi & \mathrm{d}\eta \\ \mathrm{d}\eta & \mathrm{d}\xi \end{bmatrix}
\end{aligned}$$

其中方阵 $\mathbf{A}$ 的指数函数是如下定义的一个方阵：

$$\exp\mathbf{A} = 1 + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \cdots + \frac{\mathbf{A}^n}{n!} + \cdots$$

我们的变量代换实际上是一类算符的指数函数：

$$\begin{aligned}
\begin{bmatrix} ct & x \\ x & ct \end{bmatrix} &= \exp\left(\xi \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \eta \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) \\
&= \exp\left(\begin{bmatrix} \xi & \eta \\ \eta & \xi \end{bmatrix}\right)
\end{aligned}$$

再看变量代换：

$$\begin{cases} ct = \exp\xi\cos\eta \\ x = \exp\xi\sin\eta \end{cases}$$

不难验证：

$$\begin{aligned}
\begin{bmatrix} \mathrm{d}(ct) & -\mathrm{d}x \\ \mathrm{d}x & \mathrm{d}(ct) \end{bmatrix} &= \begin{bmatrix} \exp\xi & 0 \\ 0 & \exp\xi \end{bmatrix} \begin{bmatrix} \cos\eta & -\sin\eta \\ \sin\eta & \cos\eta \end{bmatrix} \begin{bmatrix} \mathrm{d}\xi & -\mathrm{d}\eta \\ \mathrm{d}\eta & \mathrm{d}\xi \end{bmatrix} \\
&= \exp\left(\xi \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \exp\left(\eta \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right) \begin{bmatrix} \mathrm{d}\xi & -\mathrm{d}\eta \\ \mathrm{d}\eta & \mathrm{d}\xi \end{bmatrix}
\end{aligned}$$

该变量代换是另一类算符的指数函数：

$$\begin{aligned}\begin{bmatrix} ct & -x \\ x & ct \end{bmatrix} &= \exp\left(\xi\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \eta\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right) \\ &= \exp\left(\begin{bmatrix} \xi & -\eta \\ \eta & \xi \end{bmatrix}\right)\end{aligned}$$

注意到：

$$\begin{aligned}\exp\left(\begin{bmatrix} \xi & \kappa^2\eta \\ \eta & \xi \end{bmatrix}\right) &= \exp\left(\xi\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \eta\begin{bmatrix} 0 & \kappa^2 \\ 1 & 0 \end{bmatrix}\right) \\ &= \exp\left(\xi\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\exp\left(\eta\begin{bmatrix} 0 & \kappa^2 \\ 1 & 0 \end{bmatrix}\right) \\ &= \begin{bmatrix} \exp\xi & 0 \\ 0 & \exp\xi \end{bmatrix}\begin{bmatrix} \cosh\kappa\eta & \kappa^2(\frac{1}{\kappa}\sinh\kappa\eta) \\ \frac{1}{\kappa}\sinh\kappa\eta & \cosh\kappa\eta \end{bmatrix} \\ &= \begin{bmatrix} ct & \kappa^2 x \\ x & ct \end{bmatrix}\end{aligned}$$

确定了变量代换：

$$\begin{cases} ct = \exp\xi \ \ \cosh\kappa\eta \\ x = \exp\xi \ \ (\frac{1}{\kappa}\sinh\kappa\eta) \end{cases}$$

此式称为对数测地极坐标到直角坐标的变量代换，简称对数测地变换。相应的算符称为对数测地算符。式中的 ξ 称为对数测地线长，η 称为对数测地幅角，κ 称为方根曲率，$-\kappa^2$ 称为高斯曲率，又称主曲率或面积曲率。在以后的讨论中，除非特别说明，否则我们就认为 κ^2 是一个选定的固定不变的实数。我们使用‘对数’一词，是因为形式上存在一个对数函数的关系：

$$\begin{bmatrix} \xi & \kappa^2\eta \\ \eta & \xi \end{bmatrix} = \ln\left(\begin{bmatrix} ct & \kappa^2 x \\ x & ct \end{bmatrix}\right)$$

从对数测地变换，取 $\kappa = 1$，即得刚才的第一类变换。取 $\kappa = i$，i 为虚数单位，即得刚才的第二类变换。令 $\kappa \to 0$，可以得到第三类变换：

$$\begin{cases} ct = \exp\xi \\ x = \exp\xi \ \ \eta \end{cases}$$

这三类变量代换和相应的算符分别被称为是双曲型的、椭圆型的和抛物型的。

8.10.2 常性修正因子与线性修正因子

注意到，一般对数测地变换可以写成

$$\begin{cases} ct = \exp\xi \ \ \cosh\kappa\eta \\ x = \exp\xi \ \ \eta \ \dfrac{\sinh\kappa\eta}{\kappa\eta} \end{cases}$$

其中因子 $\dfrac{\sinh\theta}{\theta} = \operatorname{shc}\theta$ 常被视为一个专门的关于自变量 θ 的函数，当 $\theta \to 0$ 时，该函数和 $\cosh\theta$ 一样，也趋于 1。因此，一般形式的对数测地变换又可以看做是对抛物型对数测地变换的修正。$\cosh\kappa\eta$ 是对 1 的修正，称为常性修正因子，$\dfrac{\sinh\kappa\eta}{\kappa\eta} = \operatorname{shc}\kappa\eta$ 是对 η 的修正，称为线性修正因子。

8.10.3 测地复数

从代数的角度来讲，所有对数测地算符按照矩阵加法和乘法可以构成一个含幺交换环。它不能构成数域，是因为，当 $\kappa^2 > 0$ 时，如果 $(ct)^2 = \kappa^2 x^2 \neq 0$，则

$$\begin{bmatrix} ct & \kappa^2 x \\ x & ct \end{bmatrix} \begin{bmatrix} ct & -\kappa^2 x \\ -x & ct \end{bmatrix} = \begin{bmatrix} (ct)^2 - \kappa^2 x^2 & 0 \\ 0 & (ct)^2 - \kappa^2 x^2 \end{bmatrix} = \mathbf{0}$$

即，两个非零元素的乘积为零。这样的非零元素称为零因子。当 $\kappa^2 = 0$ 时，那些 $ct = 0$ 而 $x \neq 0$ 的元素也是零因子。当 $\kappa^2 < 0$ 时，零因子问题不存在，测地算符就可以构成数域。普通的复数就是测地算符在 $\kappa^2 = -1$ 时的特例。

零因子的存在使得乘法消去律不成立。幸运的是，就我们的这个具体问题而言，加一个适当的约束，就可以化解大部分的零因子带来的麻烦。我们称满足 $(ct)^2 - \kappa^2 x^2 > 0$ 的对数测地算符为类时的，满足 $(ct)^2 - \kappa^2 x^2 = 0$ 的为类光的，满足 $(ct)^2 - \kappa^2 x^2 < 0$ 的为类空的，注意到，两个类时测地算符的乘积仍旧是类时的，而且所有类时测地算符中不含零因子，因此乘法消去律成立。从而使得 0 与所有类时测地算符一起，几乎就要构成数域了。可惜，减法封闭性不满足，即两个类时测地算符的差不一定仍旧是类时的。怎么办呢？数学上无法保证，就从物理上保证。如果两个类时测地算符的差是类时的或者为 0，则称该减法运算对这两个算符是可行的，否则就是不可行的。这一规定是有物理基础的。可以说，在我们的讨论中，如果需要使用减法，那一定是可行的减法。不可行的减法没有物理意义。

这样， 0 与所有类时测地算符构成的集合，在矩阵加法，可行减法，矩阵乘法和马上就要定义的除法运算下，可以视同数域处理。不妨称之为可行类时对数测地复数，或简称测地数。在不致混淆的情况下，我们将使用 a, b, w, z 等小写字母表示测地数。

令测地数 $a = \left[\begin{smallmatrix} \xi & \kappa^2\eta \\ \eta & \xi \end{smallmatrix}\right]$，则其共轭测地数定义为 $\overline{a} = \left[\begin{smallmatrix} \xi & -\kappa^2\eta \\ -\eta & \xi \end{smallmatrix}\right]$。物理上，这相当于对 a 做了一个宇称变换。

令测地数 $b = \left[\begin{smallmatrix} \phi & \kappa^2\theta \\ \theta & \phi \end{smallmatrix}\right]$，则两个测地数相乘的运算定义为：

$$\begin{aligned} ab &= \begin{bmatrix} \xi & \kappa^2\eta \\ \eta & \xi \end{bmatrix} \begin{bmatrix} \phi & \kappa^2\theta \\ \theta & \phi \end{bmatrix} \\ &= \begin{bmatrix} \xi\phi + \kappa^2\eta\theta & \kappa^2(\eta\phi + \xi\theta) \\ \eta\phi + \xi\theta & \xi\phi + \kappa^2\eta\theta \end{bmatrix} \end{aligned}$$

注意到：$a=\xi\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]+\eta\left[\begin{smallmatrix}0&\kappa^2\\1&0\end{smallmatrix}\right]$。我们引入测地虚数单位 $\iota=\left[\begin{smallmatrix}0&\kappa^2\\1&0\end{smallmatrix}\right]$，则测地数可以有更简洁的写法：$a=\xi+\eta\iota$。前面的乘法运算就可以写成：

$$\begin{aligned}ab &= (\xi+\eta\iota)(\phi+\theta\iota)\\ &= (\xi\phi+\kappa^2\eta\theta)+(\eta\phi+\xi\theta)\iota\end{aligned}$$

显然，$\iota^2=\kappa^2$。注意 ι 和 κ 的联系与区别。κ 被我们当作一个实数或者纯虚数，从而 κ^2 一定是个实数。但是 ι 却始终是一个算符。当 $\kappa^2=-1$ 时，$\kappa=\pm i$，ι 也可以看做是 i。但是当 $\kappa^2=1$ 时，κ 是实数 ±1，而 ι 却是算符 $\left[\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right]$，而不是算符 $\pm\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$。后者可以看做是实数 ±1，而前者却完全不能。当 $\kappa^2\to0$ 时，ι 就是对偶数的虚数单位 ε。对偶数被 W. K. Clifford 和 E. Study 用于直线几何学。这一点都不奇怪，因为 $\kappa^2=0$ 意味着空间是平直的，其上的所有测地线都是直线。

绝大多数普通复数的相关概念都可以直接推广到测地数中来。

比如，我们称 ξ 为 a 的实部，η 为 a 的虚部。

乘积 $a\overline{a}=\xi^2-\kappa^2\eta^2$ 称为 a 的模长平方。模长平方为 1 的测地数称为幺模的。注意到：

$$\begin{aligned}&\exp\eta\iota\;\overline{\exp\eta\iota}\\ =&\;(\cosh\kappa\eta+\frac{1}{\kappa}\sinh\kappa\eta\,\iota)\,(\cosh\kappa\eta-\frac{1}{\kappa}\sinh\kappa\eta\,\iota)\\ =&\;\cosh^2\kappa\eta-\sinh^2\kappa\eta\\ \equiv&\;1\end{aligned}$$

因此 $\exp\eta\iota$ 是幺模的，对应的算符称为对数测地旋转算符。

除法运算定义为，当 $b\neq0$ 时，

$$\begin{aligned}\frac{a}{b} &= \frac{a\overline{b}}{b\overline{b}}\\ &= \frac{(\xi+\eta\iota)(\phi-\theta\iota)}{(\phi+\theta\iota)(\phi-\theta\iota)}\\ &= \frac{1}{\phi^2-\kappa^2\theta^2}\big((\xi\phi-\kappa^2\eta\theta)+(\eta\phi-\xi\theta)\iota\big)\end{aligned}$$

测地数 a 也可以写成对数测地极坐标的形式：

$$\begin{aligned}a &= \exp\,(\mu+\nu\iota)\\ &= \exp\mu\,\big(\cosh\kappa\nu+\frac{1}{\kappa}\sinh\kappa\nu\;\iota\big)\end{aligned}$$

其中对数测地线长 μ 和对数测地幅角 ν 与实部 ξ 和虚部 η 的关系是：

$$\begin{cases}\mu = \frac{1}{2}\ln\,(\xi^2-\kappa^2\eta^2)\\ \nu = \frac{1}{2\kappa}\ln\dfrac{\xi+\kappa\eta}{\xi-\kappa\eta}\end{cases}$$

8.10.4 测地复数平面

复平面可以推广为测地数平面，只是当 $\kappa^2 \geq 0$ 时，为了保留相对论的习惯，我们取横轴表示虚部，纵轴表示实部。考虑对数测地旋转时，取顺时针为正方向。当 $\kappa^2 < 0$ 时，恢复采用复数的习惯。诸如开集、极限、可导、解析等概念在测地数平面上都可定义，只是要注意把减法限制为可行减法，相应地，把连续曲线限制为可行的连续曲线。当 $\kappa^2 < 0$ 时，连续曲线总是可行的。但是，对于 $\kappa^2 \geq 0$ 的情况，可行的连续曲线在每一点都只能向绝对将来前进，无法回到起点形成封闭曲线而围出一个区域。解决的办法是修改区域的定义。有相同起点和终点的两个中途不相交的可行连续曲线形成封闭曲线，围出的部分叫做区域。这样，任何区域都有一个起点和一个终点。起点的绝对将来与终点的绝对过去的交集是一个区域，称为可行区域，在测地数平面上是一个平行四边形。可行连续曲线被限制在该平行四边形内部。

正弦曲线恰可作为可行连续曲线的典型。不知道这和物质波有没有关系。

8.10.5 时空量子猜想

洛伦兹变换说：$ct' + x'\iota = \exp(\Delta\xi + \Delta\eta\iota)(ct + x\iota)$。如果坐标系变换是以量子的方式逐步进行的，那么可设 $\Delta\xi = m\xi_0, \quad \Delta\eta = n\eta_0$，其中 ξ_0 称为元时子，η_0 称为元空子。这种观点认为元时子和元空子是各自独立地吸收和释放的。在洛伦兹变换中，只涉及到元空子的吸收或释放。

也可设 $\Delta\xi + \Delta\eta\iota = k(a\xi_0 + \eta_0\iota)$。这种观点认为元时子和元空子按固定的比例组成时空子 $\zeta_0 = a\xi_0 + \eta_0\iota$，洛伦兹变换是通过吸收或释放时空子 ζ_0 而实现的。之所以观测不到 $\exp\xi_0$ 的效应，是因为 a 太过微小的缘故。

以上两种猜想都称为时空量子猜想。

8.10.6 万能时空维度公式

来看下面这样一组公式：

$$
\begin{aligned}
a &= 1\varepsilon\frac{\sinh\alpha\varepsilon}{\alpha\varepsilon}\\
ct &= 1\cosh\alpha\varepsilon\,\xi\frac{\sinh\tau\xi}{\tau\xi}\\
x &= 1\cosh\alpha\varepsilon\ \cosh\tau\xi\,\eta\frac{\sinh\kappa_x\eta}{\kappa_x\eta}\\
y &= 1\cosh\alpha\varepsilon\ \cosh\tau\xi\ \cosh\kappa_x\eta\,\phi\frac{\sinh\kappa_y\phi}{\kappa_y\phi}\\
b &= 1\cosh\alpha\varepsilon\ \cosh\tau\xi\ \cosh\kappa_x\eta\ \cosh\kappa_y\phi\,\theta_0\frac{\sinh\kappa_z\theta_0}{\kappa_z\theta_0}\\
w &= 1\cosh\alpha\varepsilon\ \cosh\tau\xi\ \cosh\kappa_x\eta\ \cosh\kappa_y\phi\ \cosh\kappa_z\theta_0\,\psi_0\frac{\sinh\kappa_a\psi_0}{\kappa_a\psi_0}.
\end{aligned}
$$

这组公式称为万能时空维度公式。

其中的 ε 是取值极其微小的物理量，从而 a 几乎恒等于零。把 ε 所表示的维度称为零值维度。

其中的 θ_0 和 ψ_0 是取值几乎恒定的物理量，从而 b 和 w 总保持固定的比例，两者当中只有一个是独立的。把 θ_0 和 ψ_0 所表示的维度称为常值维度。

其他的 ξ、η 和 ϕ 等维度在取值微小时，该组公式就是 $ct = \xi$, $x = \eta$, $y = \phi$。这个称为弱场近似。

当 ξ、η 和 ϕ 取值并不微小时，做适当单位调整，则：

$$\begin{aligned}
ct &= \xi\frac{\sinh\tau\xi}{\tau\xi} \\
x &= \cosh\tau\xi\,\eta\frac{\sinh\kappa_x\eta}{\kappa_x\eta} \\
y &= \cosh\tau\xi\ \cosh\kappa_x\eta\,\phi\frac{\sinh\kappa_y\phi}{\kappa_y\phi} \\
b &= \cosh\tau\xi\ \cosh\kappa_x\eta\ \cosh\kappa_y\phi
\end{aligned}$$

显有 $b^2 - {\kappa_y}^2y^2 - {\kappa_x}^2x^2 - \tau^2(ct)^2 = 1$。当 $\tau^2 = -1$, ${\kappa_x}^2 = {\kappa_y}^2 = 1$ 时，有 $(ct)^2 - x^2 - y^2 = 1 - b^2$，如果 $1 - b^2$ 极其微小，那么，这就是光速不变原理。如果光速不变原理严格成立，那么 $b = \pm 1$。可以认为，$b = +1$ 表示正物质，$b = -1$ 表示反物质。这称为物质正反性的几何化。这也许是第一个把光速不变原理与物质的正反性联系起来的公式。

在万能时空维度公式中，a 之前还可以有任意多个零值维度，b 之后还可以有任意多个常值维度。也许时空维度取决于所讨论的问题。有些物理量，在一个问题中是常数，在另一个问题中可能就是变量。可能并不存在唯一的、“客观的”、绝对的标准。

8.10.7 第五维度假说与 CPT 守恒

第五维度假说认为，时空除了时间 t 和空间 x, y, z，还有一个与时间、空间地位平等的维度，称为伍间 w。该维度决定了正物质与反物质的区别。时间、空间和伍间一起构成的‘超空间’称为增广时空。

物理定律总是用物理量满足的约束方程组来表示的。设这样的方程组的一般形式为 $f_i(t, x, y, z, w) = 0$。在 $f_i(t, x, y, z, w) = 0$ 成立的前提下，如果

- $f_i(-t, x, y, z, w) = 0$ 成立，则称该定律关于时间反转变换守恒，简称 T 变换守恒。
- $f_i(t, -x, -y, -z, w) = 0$ 成立，则称该定律关于宇称变换守恒，简称 P 变换守恒。

- $f_i(t,x,y,z,-w)=0$ 成立，则称该定律关于正反变换守恒，简称 C 变换守恒。
- $f_i(-t,-x,-y,-z,w)=0$ 成立，则称该定律关于 PT 联合变换守恒。
- $f_i(-t,-x,-y,-z,-w)=0$ 成立，则称该定律关于 CPT 联合变换守恒。

在已经发现的物理定律中，有 T、P、C 单项不守恒的，有两项联合也不守恒的，但是没有三项联合都不守恒的。也就是说，到目前为止，所有物理定律，关于 CPT 联合变换都是守恒的。

注意到，5 维 2 度外积空间和 5 维 3 度外积空间都是 $\dfrac{5!}{3!2!}=10$ 维的。设 ω 是一个 5 维 2 度微分形式，那么 $\mathrm{d}\omega$ 是 5 维 3 度的。这样，恰存在一个广义斯托克斯定理来表述两者之间的关系。这种表述与 CPT 守恒有关系吗？与现有的十维空间的理论有关系吗？笔者才疏学浅，尚未找出答案。留给列位看官去发现吧。

8.11 度规与短程线

度规说到底是与配极有关。在本节，我们先认识一些度规的实例。

8.11.1 内积与度规

向量空间中给定基底 $\{\vec{e}_i\}$，则任何非零向量均可表为基矢的线性组合，即 $\vec{a}=a^i\vec{e}_i$。两向量 $\vec{a}$ 和 $\vec{b}$ 的内积定义为 $\vec{a}\cdot\vec{b}=g_{ij}a^ib^j$，其中 g_{ij} 称为度规。

任意两向量的内积可以表成所有基矢两两求内积乘系数再求和，即

$$\vec{a}\cdot\vec{b}=(a^i\vec{e}_i)\cdot(b^j\vec{e}_j)=a^ib^j\vec{e}_i\cdot\vec{e}_j.$$

可见，$g_{ij}=\vec{e}_i\cdot\vec{e}_j$。

如果两向量的内积为零，则称这两向量正交。

一个向量 $\vec{a}$ 与自己的内积，称为该向量的模长平方。再开方，就得到该向量的模长，也称长度，记为 $|\vec{a}|$。有 $|\vec{a}|^2=\vec{a}\cdot\vec{a}$。

8.11.2 坐标变换与度规的张量属性

设有另一个基底 $\vec{e}_{i'}$，使得 $\vec{e}_i=A_i^{i'}\vec{e}_{i'}$，从而任意非零向量 $\vec{a}=a^i\vec{e}_i=a^iA_i^{i'}\vec{e}_{i'}=a^{i'}\vec{e}_{i'}$，即 $a^{i'}=a^iA_i^{i'}$。这称为换基操作下的坐标变换。相应的逆变换记为 $a^i=A_{i'}^ia^{i'}=A_{i'}^iA_j^{i'}a^j$，显有 $A_{i'}^iA_j^{i'}=\delta_j^i$。

如果换基操作保持任意两个非零向量 $\vec{a}$ 和 $\vec{b}$ 的内积不变，那么，有 $\vec{a}\cdot\vec{b}=g_{i'j'}a^{i'}b^{j'}=g_{i'j'}(a^iA_i^{i'})(b^jA_j^{j'})=g_{i'j'}A_i^{i'}A_j^{j'}a^ib^j=g_{ij}a^ib^j$，因为 $\vec{a}$ 和 $\vec{b}$ 可以是

任意非零向量，所以该式最后一个等号成立，就意味着两边 $a^i b^j$ 的系数分别相等，即 $g_{ij} = g_{i'j'} A_i^{i'} A_j^{j'}$。可见 g_{ij} 就是前面讲过的二协张量。也称 g_{ij} 为协变度规。

8.11.3 逆变度规与共轭基底

给定协变度规 g_{ij}，有相应的逆变度规 g^{jk}，满足 $g_{ij}g^{jk} = \delta_i^k$。注意该式与 $A_{i'}^i A_j^{i'} = \delta_j^i$ 意义完全不同。

用逆变度规可以定义一组新的基矢 $\vec{e}^{\,i} = g^{ij}\vec{e}_j$，它们构成一个新的基底，称为原基底的共轭基底。

两个相互共轭的基底 $\vec{e}_i$ 和 $\vec{e}^{\,i}$ 有一个重要的关系：$\vec{e}^{\,i} \cdot \vec{e}_j = g^{ik}\vec{e}_k \cdot \vec{e}_j = g^{ik}g_{kj} = \delta_j^i$。利用该关系，易证 $\vec{e}^{\,i} \cdot \vec{e}^{\,j} = g^{ik}\vec{e}_k \cdot \vec{e}^{\,j} = g^{ik}\delta_k^j = g^{ij}$。可见逆变度规与共轭基底的关系，和度规与基底的关系完全一致。特别是：$g_{ij}\vec{e}^{\,j} = g_{ij}g^{jk}\vec{e}_k = \delta_i^k\vec{e}_k = \vec{e}_i$。

任何一个向量 $\vec{a}$ 既可以在基底上展开成 $\vec{a} = a^i\vec{e}_i$，又可以在共轭基底上展开成 $\vec{a} = a_i\vec{e}^{\,i}$。一方面，我们有 $\vec{a} \cdot \vec{e}^{\,i} = a^k\vec{e}_k \cdot \vec{e}^{\,i} = a^k\delta_k^i = a^i$，另一方面，我们又有 $\vec{a} \cdot \vec{e}^{\,i} = a_k\vec{e}^{\,k} \cdot \vec{e}^{\,i} = a_k g^{ki}$，因此，$a^i = g^{ij}a_j$。同理可得 $a_i = g_{ij}a^j$。故而 g^{ij} 又称角标提升算符，g_{ij} 又称角标下降算符。

8.11.4 平面极坐标系

前面讲过，任何流形都可以看做是曲线坐标系变换中取一些变量为守恒变量而得来。最简单的曲线坐标系是欧氏度规平面上的极坐标系。

$$
\begin{aligned}
x &= r\cos\theta = x(r,\theta) = u^1(v^1, v^2) \\
y &= r\sin\theta = y(r,\theta) = u^2(v^1, v^2)
\end{aligned}
$$

我们特意把有关变量重新命名为 u^i 和 v^i，为的是容易与张量简记法对应。

设平面上一矢径 $\vec{R} = x\,\vec{\imath} + y\,\vec{\jmath}$ 的终点的邻域内有一微小向量 $\mathrm{d}\vec{R} = \mathrm{d}x\,\vec{\imath} + \mathrm{d}y\,\vec{\jmath} = \dfrac{\partial \vec{R}}{\partial u^i}\,\mathrm{d}u^i$。该微小向量的所有可能取值构成局部切空间。由上述变量代换可知：

$$
\mathrm{d}x = \frac{\partial x}{\partial r}\,\mathrm{d}r + \frac{\partial x}{\partial \theta}\,\mathrm{d}\theta = \cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta = \frac{\partial u^1}{\partial v^i}\,\mathrm{d}v^i
$$

$$
\mathrm{d}y = \frac{\partial y}{\partial r}\,\mathrm{d}r + \frac{\partial y}{\partial \theta}\,\mathrm{d}\theta = \sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta = \frac{\partial u^2}{\partial v^i}\,\mathrm{d}v^i
$$

这种关系就是逆变矢量的坐标变换：$\mathrm{d}u^i = \dfrac{\partial u^i}{\partial v^{i'}}\,\mathrm{d}v^{i'} = \partial_{i'}^i\,\mathrm{d}v^{i'}$。

该曲线坐标变换的 Jacobi 矩阵为

$$\frac{\partial(u^i)}{\partial(v^{i'})}=\frac{\partial(x,y)}{\partial(r,\theta)}=\begin{bmatrix}\dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta}\\ \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta}\end{bmatrix}=\begin{bmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{bmatrix}$$

设同一微小向量在曲线坐标系活动标架上展开为 $\mathrm{d}\vec{R}=\mathrm{d}r\,\vec{e}_r+\mathrm{d}\theta\,\vec{e}_\theta=\dfrac{\partial\vec{R}}{\partial v^i}\,\mathrm{d}v^i$。易见

$$\begin{aligned}\vec{e}_r &= \cos\theta\,\vec{\imath}+\sin\theta\,\vec{\jmath}\\ \vec{e}_\theta &= -r\sin\theta\,\vec{\imath}+r\cos\theta\,\vec{\jmath}\end{aligned}$$

进而 $g_{rr}=\vec{e}_r\cdot\vec{e}_r=1,\quad g_{r\theta}=g_{\theta r}=\vec{e}_r\cdot\vec{e}_\theta=0,\quad g_{\theta\theta}=\vec{e}_\theta\cdot\vec{e}_\theta=r^2$。这个结果应该也能从直角坐标系的度规 $h_{xx}=1,\quad h_{xy}=h_{yx}=0,\quad h_{yy}=1$，即 $h_{ij}=\delta_{ij}$，按照二阶协变张量的坐标变换法则做变换而得到，即 $g_{ij}=\partial_i^{i'}\partial_j^{j'}h_{i'j'}$。

该微小向量的模长平方 $\mathrm{d}l^2$ 在直角坐标系中表示为 $\mathrm{d}l^2=h_{ij}\,\mathrm{d}u^i\,\mathrm{d}u^j=\mathrm{d}x^2+\mathrm{d}y^2$，在极坐标系中表示为 $\mathrm{d}l^2=g_{ij}\,\mathrm{d}v^i\,\mathrm{d}v^j=\mathrm{d}r^2+r^2\,\mathrm{d}\theta^2$。两者表达的是同样的实质内容，相互之间只是相差一个坐标变换而已。

逆变度规张量 g^{ij} 的分量显见有：$g^{rr}=1,\quad g^{r\theta}=g^{\theta r}=0,\quad g^{\theta\theta}=\dfrac{1}{r^2}$。因此，共轭基矢为

$$\begin{aligned}\vec{e}^{\,r} &= g^{rr}\vec{e}_r+g^{r\theta}\vec{e}_\theta = \vec{e}_r &&= \cos\theta\,\vec{\imath}+\sin\theta\,\vec{\jmath}\\ \vec{e}^{\,\theta} &= g^{\theta r}\vec{e}_r+g^{\theta\theta}\vec{e}_\theta = \frac{1}{r^2}\vec{e}_\theta &&= -\frac{1}{r}\sin\theta\,\vec{\imath}+\frac{1}{r}\cos\theta\,\vec{\jmath}\end{aligned}$$

不难验证：$\vec{e}^{\,r}\cdot\vec{e}_r=1,\quad \vec{e}^{\,r}\cdot\vec{e}_\theta=\vec{e}^{\,\theta}\cdot\vec{e}_r=0,\quad \vec{e}^{\,\theta}\cdot\vec{e}_\theta=1$，即 $\vec{e}^{\,i}\cdot\vec{e}_j=\delta^i_j$。

8.11.5 闵氏度规平面上的双曲极坐标系

现在来看闵氏度规平面上的双曲极坐标系。

$$\begin{aligned}x &= s\cosh\zeta = x(s,\zeta) = u^1(v^1,v^2)\\ y &= s\sinh\zeta = y(s,\zeta) = u^2(v^1,v^2)\end{aligned}$$

我们还是把有关变量重新命名为 u^i 和 v^i，以便与张量简记法对应。

前述微小向量 $\mathrm{d}\vec{R}=\mathrm{d}x\,\vec{\imath}+\mathrm{d}y\,\vec{\jmath}=\dfrac{\partial\vec{R}}{\partial u^i}\,\mathrm{d}u^i$ 的分量有

$$\mathrm{d}x=\frac{\partial x}{\partial s}\,\mathrm{d}s+\frac{\partial x}{\partial\zeta}\,\mathrm{d}\zeta=\cosh\zeta\,\mathrm{d}s+s\sinh\zeta\,\mathrm{d}\zeta=\frac{\partial u^1}{\partial v^i}\,\mathrm{d}v^i$$

$$\mathrm{d}y=\frac{\partial y}{\partial s}\,\mathrm{d}s+\frac{\partial y}{\partial\zeta}\,\mathrm{d}\zeta=\sinh\zeta\,\mathrm{d}s+s\cosh\zeta\,\mathrm{d}\zeta=\frac{\partial u^2}{\partial v^i}\,\mathrm{d}v^i$$

这种关系就是逆变矢量的坐标变换： $\mathrm{d}u^i=\dfrac{\partial u^i}{\partial v^{i'}}\,\mathrm{d}v^{i'}=\partial^i_{i'}\,\mathrm{d}v^{i'}$。

该曲线坐标变换的 Jacobi 矩阵为

$$\frac{\partial(u^i)}{\partial(v^{i'})} = \frac{\partial(x,y)}{\partial(s,\zeta)} = \begin{bmatrix} \dfrac{\partial x}{\partial s} & \dfrac{\partial x}{\partial \zeta} \\ \\ \dfrac{\partial y}{\partial s} & \dfrac{\partial y}{\partial \zeta} \end{bmatrix} = \begin{bmatrix} \cosh\zeta & s\sinh\zeta \\ \sinh\zeta & s\cosh\zeta \end{bmatrix}$$

设同一微小向量在曲线坐标系活动标架上展开为 $\mathrm{d}\vec{R} = \mathrm{d}s\,\vec{e}_s + \mathrm{d}\zeta\,\vec{e}_\zeta = \dfrac{\partial \vec{R}}{\partial v^i}\,\mathrm{d}v^i$。易见

$$\begin{aligned} \vec{e}_s &= \cosh\zeta\,\vec{\imath} + \sinh\zeta\,\vec{\jmath} \\ \vec{e}_\zeta &= s\sinh\zeta\,\vec{\imath} + s\cosh\zeta\,\vec{\jmath} \end{aligned}$$

进而 $g_{ss} = \vec{e}_s \cdot_u \vec{e}_s = 1,\quad g_{s\zeta} = g_{\zeta s} = \vec{e}_s \cdot_u \vec{e}_\zeta = 0,\quad g_{\zeta\zeta} = \vec{e}_\zeta \cdot_u \vec{e}_\zeta = -s^2$。这里我们特意为内积运算符加了下标，明确表示这是双曲内积，而不是前面的普通内积。两向量 a^i 和 b^j 的双曲内积定义为 $\vec{a}\cdot_u\vec{b} = \eta_{ij}a^ib^j$。其中的 η_{ij} 就是前面讲过的闵氏度规，即闵氏平面上直角坐标系的度规。为了和上一例对比，我们记 $h_{xx} = 1,\quad h_{xy} = h_{yx} = 0,\quad h_{yy} = -1$，即 $h_{ij} = \eta_{ij}$。那么，$g_{ij} = \partial_i^{i'}\partial_j^{j'}h_{i'j'}$。

该微小向量的模长平方 $\mathrm{d}l^2$ 在直角坐标系中表示为 $\mathrm{d}l^2 = h_{ij}\,\mathrm{d}u^i\,\mathrm{d}u^j = \mathrm{d}x^2 - \mathrm{d}y^2$，在双曲极坐标系中表示为 $\mathrm{d}l^2 = g_{ij}\,\mathrm{d}v^i\,\mathrm{d}v^j = \mathrm{d}s^2 - s^2\,\mathrm{d}\zeta^2$。两者表达的是同样的实质内容，相互之间只是相差一个坐标变换而已。

对闵氏平面双曲极坐标系，逆变度规张量 g^{ij} 的分量显见有：$g^{ss} = 1,\quad g^{s\zeta} = g^{\zeta s} = 0,\quad g^{\zeta\zeta} = -\dfrac{1}{s^2}$。因此，共轭基矢为

$$\begin{aligned} \vec{e}^{\,s} &= g^{ss}\vec{e}_s + g^{s\zeta}\vec{e}_\zeta = \vec{e}_s &&= \cosh\zeta\,\vec{\imath} + \sinh\zeta\,\vec{\jmath} \\ \vec{e}^{\,\zeta} &= g^{\zeta s}\vec{e}_s + g^{\zeta\zeta}\vec{e}_\zeta = -\frac{1}{s^2}\vec{e}_\zeta &&= -\frac{1}{s}\sinh\zeta\,\vec{\imath} - \frac{1}{s}\cosh\zeta\,\vec{\jmath} \end{aligned}$$

不难验证：$\vec{e}^{\,s}\cdot_u\vec{e}_s = 1,\quad \vec{e}^{\,s}\cdot_u\vec{e}_\zeta = \vec{e}^{\,\zeta}\cdot_u\vec{e}_s = 0,\quad \vec{e}^{\,\zeta}\cdot_u\vec{e}_\zeta = 1$，即 $\vec{e}^{\,i}\cdot_u\vec{e}_j = \delta^i_j$。可见，基底与共轭基底的这种关系与度规无关，与曲线坐标系的选择无关。

8.11.6　空间中的球坐标系与球面上的曲线坐标系

正如前面讲过的，球面可以通过三维欧氏空间极坐标系取极径为守恒变量而得到。所以，我们先看三维欧氏空间的极坐标系。

$$\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned}$$

各坐标的微元有

$$\begin{aligned} \mathrm{d}x &= \sin\theta\cos\phi\,\mathrm{d}r + r\cos\theta\cos\phi\,\mathrm{d}\theta - r\sin\theta\sin\phi\,\mathrm{d}\phi \\ \mathrm{d}y &= \sin\theta\sin\phi\,\mathrm{d}r + r\cos\theta\sin\phi\,\mathrm{d}\theta + r\sin\theta\cos\phi\,\mathrm{d}\phi \\ \mathrm{d}z &= \cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta \end{aligned}$$

因此，曲线坐标系的基矢有

$$\begin{aligned}\vec{e}_r &= \sin\theta\cos\phi\,\vec{\imath}+\sin\theta\sin\phi\,\vec{\jmath}+\cos\theta\,\vec{k}\\ \vec{e}_\theta &= r\cos\theta\cos\phi\,\vec{\imath}+r\cos\theta\sin\phi\,\vec{\jmath}-r\sin\theta\,\vec{k}\\ \vec{e}_\phi &= -r\sin\theta\sin\phi\,\vec{\imath}+r\sin\theta\cos\phi\,\vec{\jmath}\end{aligned}$$

度规的各分量为 $g_{rr}=\vec{e}_r\cdot\vec{e}_r=1$, $g_{r\theta}=g_{\theta r}=\vec{e}_r\cdot\vec{e}_\theta=0$, $g_{r\phi}=g_{\phi r}=\vec{e}_r\cdot\vec{e}_\phi=0$, $g_{\theta\phi}=g_{\phi\theta}=\vec{e}_\theta\cdot\vec{e}_\phi=0$, $g_{\theta\theta}=\vec{e}_\theta\cdot\vec{e}_\theta=r^2$, $g_{\phi\phi}=\vec{e}_\phi\cdot\vec{e}_\phi=r^2\sin^2\theta$。逆变度规的分量为 $g^{rr}=1$, $g^{\theta\theta}=\dfrac{1}{r^2}$, $g_{\phi\phi}=\dfrac{1}{r^2\sin^2\theta}$，其他分量都为零。共轭基矢为

$$\begin{aligned}\vec{e}^{\,r} &= \sin\theta\cos\phi\,\vec{\imath}+\sin\theta\sin\phi\,\vec{\jmath}+\cos\theta\,\vec{k}\\ \vec{e}^{\,\theta} &= \frac{1}{r}\cos\theta\cos\phi\,\vec{\imath}+\frac{1}{r}\cos\theta\sin\phi\,\vec{\jmath}-\frac{1}{r}\sin\theta\,\vec{k}\\ \vec{e}^{\,\phi} &= -\frac{1}{r\sin\theta}\sin\phi\,\vec{\imath}+\frac{1}{r\sin\theta}\cos\phi\,\vec{\jmath}\end{aligned}$$

微小向量的模长平方为

$$\mathrm{d}l^2=\mathrm{d}r^2+r^2\,\mathrm{d}\theta^2+r^2\sin^2\theta\,\mathrm{d}\phi^2$$

取 $r\equiv 1$，则 $\mathrm{d}r\equiv 0$。这样，我们就得到一个单位球面。

该球面上的点只由两个坐标变量 θ 和 ϕ 来刻画。该球面上微小向量 $(\mathrm{d}\theta,\mathrm{d}\phi)$ 的模长平方为

$$\mathrm{d}l^2=\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\phi^2$$

即该球面的这个坐标系的度规为

$$g_{\theta\theta}=1,\quad g_{\phi\phi}=\sin^2\theta,\quad g_{\theta\phi}=g_{\phi\theta}=0$$

8.11.7 过径弧几何中的度规

前面讲过，可把单位球面通过南极投影映射到赤道面上。设球面上一点 $P(\theta,\phi)$ 映射到赤道面上的点 $Q(u,v)$，其中 u,v 为平面直角坐标系的坐标，有

$$\begin{aligned}u &= \tan\frac{\theta}{2}\cos\phi=\frac{\sin\theta}{1+\cos\theta}\cos\phi\\ v &= \tan\frac{\theta}{2}\sin\phi=\frac{\sin\theta}{1+\cos\theta}\sin\phi\end{aligned}$$

微元的变换关系是

$$\begin{aligned}\mathrm{d}u &= \frac{1}{1+\cos\theta}(\cos\phi\,\mathrm{d}\theta-\sin\theta\sin\phi\,\mathrm{d}\phi)\\ \mathrm{d}v &= \frac{1}{1+\cos\theta}(\sin\phi\,\mathrm{d}\theta+\sin\theta\cos\phi\,\mathrm{d}\phi)\end{aligned}$$

该变换的行列式为

$$|\Phi| = \frac{\sin\theta}{(1+\cos\theta)^2}$$

易见其逆变换为

$$\begin{aligned}
\mathrm{d}\theta &= \frac{1+\cos\theta}{\sin\theta}(\sin\theta\cos\phi\,\mathrm{d}u + \sin\theta\sin\phi\,\mathrm{d}v) \\
\mathrm{d}\phi &= \frac{1+\cos\theta}{\sin\theta}(-\sin\phi\,\mathrm{d}u + \cos\phi\,\mathrm{d}v)
\end{aligned}$$

因而，与上述球面坐标系度规等距的度规是 $g_{uu} = (1+\cos\theta)^2$, $g_{uv} = g_{vu} = 0$, $g_{vv} = (1+\cos\theta)^2$。其中的非零分量用 u,v 来表示，即得 $g_{ij} = \dfrac{4}{(1+u^2+v^2)^2}\delta_{ij}$。

8.11.8 变分法与欧拉方程

现在我们用这些度规来求短程线。设二维流形上两点之间有光滑曲线 $\Gamma : x_i(t)$, $t_1 \le t \le t_2$。其长度为

$$L = \int_\Gamma \mathrm{d}l = \int_{t_1}^{t_2} \sqrt{g_{ij}\dot{x}^i\dot{x}^j}\,\mathrm{d}t$$

所谓短程线，就是两点之间所有可能的光滑曲线中，长度最短者。这意味着，任何曲线只要与它有偏离，长度都会增加。因此，短程线是个‘极值点’。当然这不是普通的极值，而是‘变分’的极值，表示为 $\delta L = 0$。

先来看一般化的目标函数 $F(t,\theta,\phi,\dot{\theta},\dot{\phi})$ 的变分问题。设 $\theta = \theta(t)$, $\phi = \phi(t)$ 是问题的解，而 $\tilde{\theta} = \tilde{\theta}(t)$, $\tilde{\phi} = \tilde{\phi}(t)$ 是任意的另外一条曲线，两者的偏离为 $\delta\theta = \tilde{\theta} - \theta$, $\delta\phi = \tilde{\phi} - \phi$。两者的一阶导数的偏离为 $\delta\dot{\theta} = \dot{\tilde{\theta}} - \dot{\theta} = \dfrac{\mathrm{d}}{\mathrm{d}t}(\tilde{\theta} - \theta) = \dfrac{\mathrm{d}}{\mathrm{d}t}\delta\theta$, $\delta\dot{\phi} = \dfrac{\mathrm{d}}{\mathrm{d}t}\delta\phi$。那么，目标函数的偏离可以展开成这些偏离的线性组合：

$$\begin{aligned}
\delta F &= F(t,\tilde{\theta},\tilde{\phi},\dot{\tilde{\theta}},\dot{\tilde{\phi}}) - F(t,\theta,\phi,\dot{\theta},\dot{\phi}) \\
&= \frac{\partial F}{\partial\theta}\delta\theta + \frac{\partial F}{\partial\phi}\delta\phi + \frac{\partial F}{\partial\dot{\theta}}\delta\dot{\theta} + \frac{\partial F}{\partial\dot{\phi}}\delta\dot{\phi} \\
&= \frac{\partial F}{\partial\theta}\delta\theta + \frac{\partial F}{\partial\phi}\delta\phi + \frac{\partial F}{\partial\dot{\theta}}\frac{\mathrm{d}}{\mathrm{d}t}\delta\theta + \frac{\partial F}{\partial\dot{\phi}}\frac{\mathrm{d}}{\mathrm{d}t}\delta\phi \\
&= \frac{\partial F}{\partial\theta}\delta\theta + \frac{\partial F}{\partial\phi}\delta\phi + \frac{\mathrm{d}}{\mathrm{d}t}(\frac{\partial F}{\partial\dot{\theta}}\delta\theta) - (\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial\dot{\theta}})\delta\theta + \frac{\mathrm{d}}{\mathrm{d}t}(\frac{\partial F}{\partial\dot{\phi}}\delta\phi) - (\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial\dot{\phi}})\delta\phi
\end{aligned}$$

因此，总变分等于各时刻偏离的积分：

$$\begin{aligned}
\delta\int_{t_1}^{t_2} F\,\mathrm{d}t &= \int_{t_1}^{t_2} \delta F\,\mathrm{d}t \\
&= \left(\frac{\partial F}{\partial\dot{\theta}}\delta\theta\right)\Bigg|_{t_1}^{t_2} + \left(\frac{\partial F}{\partial\dot{\phi}}\delta\phi\right)\Bigg|_{t_1}^{t_2} + \int_{t_1}^{t_2}\left(\frac{\partial F}{\partial\theta} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial\dot{\theta}}\right)\delta\theta\,\mathrm{d}t \\
&\quad + \int_{t_1}^{t_2}\left(\frac{\partial F}{\partial\phi} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial\dot{\phi}}\right)\delta\phi\,\mathrm{d}t
\end{aligned}$$

由于偏离函数 $\delta\theta$ 和 $\delta\phi$ 在两个端点上的值都为零，因此，上式前两项全函数在端点上的值也是零，两端点上值的差还是零。后两项中的偏离函数 $\delta\theta$ 和 $\delta\phi$ 是可以任意取的，在这种情况下，要想让积分恒为零，只有一个办法，那就是让括号里的表达式恒为零，即

$$\frac{\partial F}{\partial \theta}-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial \dot{\theta}}=0, \qquad \frac{\partial F}{\partial \phi}-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial \dot{\phi}}=0$$

这组方程称为 Euler 方程。

顺便提一下，在直线上运动的质点的动能为 $K=\frac{1}{2}m\dot{x}^2$，质点在保守力场 V 中的受力为 $F=-\frac{\partial V}{\partial x}$。如果取目标函数为该质点的拉格朗日函数 $L=K-V$，那么

$$\frac{\partial L}{\partial x}=-\frac{\partial V}{\partial x}=F, \qquad \frac{\partial L}{\partial \dot{x}}=\frac{\partial K}{\partial \dot{x}}=m\dot{x}$$

代入 Euler 方程，就是牛顿第二定律 $F=m\ddot{x}$。对于更为一般的情况，力学系统的拉格朗日函数可以表述为广义位置 $q^i(t)$ 和广义速度 $\dot{q}^i(t)$ 的函数 $L=L(q^i(t),\dot{q}^i(t))$。$\frac{\partial L}{\partial q^i}$ 称为广义力 f_i。$\frac{\partial L}{\partial \dot{q}^i}$ 称为广义动量 p_i。那么 Euler 方程表述的就是：广义力等于广义动量对时间的导数，即 $f_i=\dot{p}_i$。拉格朗日函数关于时间的积分称为作用量 S。给定起点和终点，保守力场中的质点总是选择能够使作用量最小的运动方式来运动。这就是最小作用量原理，常表述为 $\delta S=0$。在量子力学中，$\exp(iS/\hbar)$ 用来表示相位的变化。

8.11.9 短程线方程

现在，取目标函数为 $F^2(t,\theta,\phi,\dot{\theta},\dot{\phi})=g_{\theta\theta}\dot{\theta}\dot{\theta}+g_{\theta\phi}\dot{\theta}\dot{\phi}+g_{\phi\theta}\dot{\phi}\dot{\theta}+g_{\phi\phi}\dot{\phi}\dot{\phi}$，采用前面讲过的简记法，该式又可写成 $F^2(t,x^i,\dot{x}^i)=g_{ij}\dot{x}^i\dot{x}^j$。

先请注意一个事实，g_{ij} 只含自变量 x^i 而与 $\dot{x}^i$ 无关，但是，由于 x^i 随时间而变，所以 $\frac{\mathrm{d}g_{ij}}{\mathrm{d}t}=\frac{\partial g_{ij}}{\partial x^m}\dot{x}^m$。

把 F^2 对 x^i 求偏导，

$$\frac{\partial F^2}{\partial x^i}=2F\frac{\partial F}{\partial x^i}=(\frac{\partial g_{mn}}{\partial x^i})\dot{x}^m\dot{x}^n$$

把 F^2 对 $\dot{x}^i$ 求偏导，

$$\frac{\partial F^2}{\partial \dot{x}^i}=2F\frac{\partial F}{\partial \dot{x}^i}=2g_{ij}\dot{x}^j$$

那么，Euler 方程的对时间求导项就是

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial \dot{x}^i} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{F}g_{ij}\dot{x}^j\right)\\ &= \frac{1}{F}g_{ij}\ddot{x}^j+\frac{1}{F}\left(\frac{\partial g_{ij}}{\partial x^m}\dot{x}^m\right)\dot{x}^i-\frac{1}{F^2}\frac{\mathrm{d}F}{\mathrm{d}t}g_{ij}\dot{x}^j\end{aligned}$$

在求短程线这个具体问题中，我们可以取弧长作为‘时间’，也就是曲线的参数。这样 $F\equiv 1$，进而 $\dfrac{\mathrm{d}F}{\mathrm{d}t}\equiv 0$。代入 Euler 方程，得

$$g_{ij}\ddot{x}^j+\frac{\partial g_{ij}}{\partial x^m}\dot{x}^m\dot{x}^j-\frac{1}{2}\frac{\partial g_{mn}}{\partial x^i}\dot{x}^m\dot{x}^n=0$$

这就是短程线应当满足的二阶微分方程组。注意，这个方程组表明，短程线完全由度规所决定。另外，这个方程组把 $\ddot{x}^i$ 用 $\dot{x}^i$ 的二次型表示出来。

8.11.10 球面几何的短程线

对前述球面的例子，考虑到 $g_{\theta\phi}=g_{\phi\theta}=0$，上述微分方程组展开成：

$$\begin{aligned}
g_{\theta\theta}\ddot{\theta}+\frac{\partial g_{\theta\theta}}{\partial\theta}\dot{\theta}\dot{\theta}+\frac{\partial g_{\theta\theta}}{\partial\phi}\dot{\phi}\dot{\theta}-\frac{1}{2}\frac{\partial g_{\theta\theta}}{\partial\theta}\dot{\theta}\dot{\theta}-\frac{1}{2}\frac{\partial g_{\phi\phi}}{\partial\theta}\dot{\phi}\dot{\phi}\;&=\;0\\
g_{\phi\phi}\ddot{\phi}+\frac{\partial g_{\phi\phi}}{\partial\theta}\dot{\theta}\dot{\phi}+\frac{\partial g_{\phi\phi}}{\partial\phi}\dot{\phi}\dot{\phi}-\frac{1}{2}\frac{\partial g_{\theta\theta}}{\partial\phi}\dot{\theta}\dot{\theta}-\frac{1}{2}\frac{\partial g_{\phi\phi}}{\partial\phi}\dot{\phi}\dot{\phi}\;&=\;0
\end{aligned}$$

把 $g_{\theta\theta}=1,\quad g_{\phi\phi}=\sin^2\theta$ 代入，得

$$\begin{aligned}
\ddot{\theta}-\sin\theta\cos\theta\,\dot{\phi}\dot{\phi}\;&=\;0\\
\sin\theta\,\ddot{\phi}+2\cos\theta\,\dot{\theta}\dot{\phi}\;&=\;0
\end{aligned}$$

这就是球面上的上述度规所决定的短程线微分方程组。这个方程组的解是众所周知的：球面上的短程线是大圆。在这里，我们验证一下大圆确实满足这个方程组，这样心里就踏实了。

单位球面上任何一个大圆可以用方程组

$$\begin{aligned}
ax+by+cz\;&=\;0\\
x^2+y^2+z^2\;&=\;1
\end{aligned}$$

来刻画，其中 a,b,c 不全为零。按前面讲过的变量代换，把 x,y,z 用 θ,ϕ 表示，则得到球面上大圆的方程：

$$a\sin\theta\cos\phi+b\sin\theta\sin\phi+c\cos\theta=0$$

把 θ,ϕ 都看做是时间的函数，方程两边对时间求导，得：

$$a\cos\theta\cos\phi\,\dot{\theta}-a\sin\theta\sin\phi\,\dot{\phi}+b\cos\theta\sin\phi\,\dot{\theta}+b\sin\theta\cos\phi\,\dot{\phi}-c\sin\theta\,\dot{\theta}=0$$

这是 $\dot{\theta}$ 和 $\dot{\phi}$ 之间的关系。此式两边再对时间求导，得：

$$\begin{aligned}
&-a\sin\theta\cos\phi\,\dot{\theta}\dot{\theta}-a\cos\theta\sin\phi\,\dot{\theta}\dot{\phi}+a\cos\theta\cos\phi\,\ddot{\theta}\\
&-a\cos\theta\sin\phi\,\dot{\theta}\dot{\phi}-a\sin\theta\cos\phi\,\dot{\phi}\dot{\phi}-a\sin\theta\sin\phi\,\ddot{\phi}\\
&-b\sin\theta\sin\phi\,\dot{\theta}\dot{\theta}+b\cos\theta\cos\phi\,\dot{\theta}\dot{\phi}+b\cos\theta\sin\phi\,\ddot{\theta}\\
&+b\cos\theta\cos\phi\,\dot{\theta}\dot{\phi}-b\sin\theta\sin\phi\,\dot{\phi}\dot{\phi}+b\sin\theta\cos\phi\,\ddot{\phi}\\
&\qquad\qquad-c\cos\theta\,\dot{\theta}\dot{\theta}-c\sin\theta\,\ddot{\theta}\;=\;0
\end{aligned}$$

其中的三个 $\dot\theta\dot\theta$ 项显见系数和为零。再把 $\ddot\theta$ 和 $\ddot\phi$ 的表达式代入，不难验证，$\dot\phi\dot\phi$ 和 $\dot\theta\dot\phi$ 的项的系数和分别也都为零。因此，大圆确实满足球面短程线方程。

8.11.11 过径弧几何的短程线

现在来看过径弧几何。由 $g_{uv} = g_{vu} = 0$，短程线方程组可以展开成

$$
\begin{aligned}
g_{uu}\ddot u + \frac{\partial g_{uu}}{\partial u}\dot u\dot u + \frac{\partial g_{uu}}{\partial v}\dot v\dot u - \frac{1}{2}\frac{\partial g_{uu}}{\partial u}\dot u\dot u - \frac{1}{2}\frac{\partial g_{vv}}{\partial u}\dot v\dot v &= 0\\
g_{vv}\ddot v + \frac{\partial g_{vv}}{\partial u}\dot u\dot v + \frac{\partial g_{vv}}{\partial v}\dot v\dot v - \frac{1}{2}\frac{\partial g_{uu}}{\partial v}\dot u\dot u - \frac{1}{2}\frac{\partial g_{vv}}{\partial v}\dot v\dot v &= 0
\end{aligned}
$$

把 $g_{uu} = g_{vv} = \dfrac{4}{(1+u^2+v^2)^2}$ 代入，得

$$
\begin{aligned}
(1+u^2+v^2)\ddot u - 2u\dot u\dot u - 4v\dot v\dot u + 2u\dot v\dot v &= 0\\
(1+u^2+v^2)\ddot v - 4u\dot u\dot v - 2v\dot v\dot v + 2v\dot u\dot u &= 0
\end{aligned}
$$

这就是复平面上的上述度规所决定的短程线微分方程组。这个方程组的解我们也知道：复平面上过径弧几何的短程线是过径圆。下面，我们验证一下过径圆确实满足这个方程组。

复平面上过径弧几何中的任何一个过径圆可以用方程

$$(u - a\cos\beta)^2 + (v - a\sin\beta)^2 = a^2 + 1$$

来刻画，其中 a 表示过径圆圆心到基准圆圆心的距离，β 表示两个圆心的连线与选定的参考线，比如 x 轴，的夹角。该方程还可以变换成另一种形式：

$$(1+u^2+v^2) = 2(1 + a\cos\beta\, u + a\sin\beta\, v)$$

方程两边对时间求导，得 $\dot u, \dot v$ 之间的关系

$$(u - a\cos\beta)\dot u + (v - a\sin\beta)\dot v = 0$$

再对时间求导，得：

$$\dot u\dot u + (u - a\cos\beta)\ddot u + \dot v\dot v + (v - a\sin\beta)\ddot v = 0$$

两边同乘 $(1+u^2+v^2)$，消去 $\ddot u$ 和 $\ddot v$。再代入 $\dot u, \dot v$ 之间的关系，可证方程成立，即过径圆是该度规所决定的短程线。

按照上述给定的度规，球极投影把球面上的大圆变换为复平面上的过径圆，即把短程线变换为短程线。这种特点的变换称为保短变换，又称短程变换。

8.11.12 空间双曲坐标系与闵氏双叶双曲面几何

现在来看三维闵氏度规空间中的单位双叶双曲面上的度规。三维闵氏空间中的点的直角坐标 (x, y, z) 和双曲极坐标 (s, ζ, ϕ) 之间有变换关系：

$$
\begin{aligned}
x &= s \sinh\zeta \cos\phi \\
y &= s \sinh\zeta \sin\phi \\
z &= s \cosh\zeta
\end{aligned}
$$

与直角坐标系配套的闵氏度规为 $\eta_{xx} = \eta_{yy} = -1$, $\eta_{zz} = 1$, $\eta_{xy} = \eta_{yx} = \eta_{xz} = \eta_{zx} = \eta_{yz} = \eta_{zy} = 0$。微小向量的模长平方为 $\mathrm{d}l^2 = -\,\mathrm{d}x^2 - \mathrm{d}y^2 + \mathrm{d}z^2$，

各坐标的微元有

$$
\begin{aligned}
\mathrm{d}x &= \sinh\zeta \cos\phi\, \mathrm{d}s + s\cosh\zeta\cos\phi\, \mathrm{d}\zeta - s\sinh\zeta\sin\phi\, \mathrm{d}\phi \\
\mathrm{d}y &= \sinh\zeta \sin\phi\, \mathrm{d}s + s\cosh\zeta\sin\phi\, \mathrm{d}\zeta + s\sinh\zeta\cos\phi\, \mathrm{d}\phi \\
\mathrm{d}z &= \cosh\zeta\, \mathrm{d}s + s\sinh\zeta\, \mathrm{d}\zeta
\end{aligned}
$$

因此，曲线坐标系的基矢有

$$
\begin{aligned}
\vec{e}_s &= \sinh\zeta\cos\phi\,\vec{\imath} + \sinh\zeta\sin\phi\,\vec{\jmath} + \cosh\zeta\,\vec{k} \\
\vec{e}_\zeta &= s\cosh\zeta\cos\phi\,\vec{\imath} + s\cosh\zeta\sin\phi\,\vec{\jmath} + s\sinh\zeta\,\vec{k} \\
\vec{e}_\phi &= -s\sinh\zeta\sin\phi\,\vec{\imath} + s\sinh\zeta\cos\phi\,\vec{\jmath}
\end{aligned}
$$

度规的各分量为 $g_{ss} = \vec{e}_s \cdot_\eta \vec{e}_s = 1$, $g_{s\zeta} = g_{\zeta s} = \vec{e}_s \cdot_\eta \vec{e}_\zeta = 0$, $g_{s\phi} = g_{\phi s} = \vec{e}_s \cdot_\eta \vec{e}_\phi = 0$, $g_{\zeta\phi} = g_{\phi\zeta} = \vec{e}_\zeta \cdot_\eta \vec{e}_\phi = 0$, $g_{\zeta\zeta} = \vec{e}_\zeta \cdot_\eta \vec{e}_\zeta = -s^2$, $g_{\phi\phi} = \vec{e}_\phi \cdot_\eta \vec{e}_\phi = -s^2\sinh^2\zeta$。微小向量的模长平方为 $dl^2 = \mathrm{d}s^2 - s^2\,\mathrm{d}\zeta^2 - s^2\sinh^2\zeta\,\mathrm{d}\phi^2$。

现在，令 $s \equiv 1$，则曲线坐标系的流动点 $(1, \zeta, \phi)$ 刻画单位双叶双曲面。该曲面上微小向量的模长平方为 $dl^2 = -\,\mathrm{d}\zeta^2 - \sinh^2\zeta\,\mathrm{d}\phi^2$。即相应的度规为：

$$
g_{\zeta\zeta} = -1, \quad g_{\phi\phi} = -\sinh^2\zeta, \quad g_{\zeta\phi} = g_{\phi\zeta} = 0
$$

考虑到 $g_{\zeta\phi} = g_{\phi\zeta} = 0$，该度规所决定的短程线的微分方程组可以展开成：

$$
\begin{aligned}
g_{\zeta\zeta}\ddot{\zeta} + \frac{\partial g_{\zeta\zeta}}{\partial\zeta}\dot{\zeta}\dot{\zeta} + \frac{\partial g_{\zeta\zeta}}{\partial\phi}\dot{\phi}\dot{\zeta} - \frac{1}{2}\frac{\partial g_{\zeta\zeta}}{\partial\zeta}\dot{\zeta}\dot{\zeta} - \frac{1}{2}\frac{\partial g_{\phi\phi}}{\partial\zeta}\dot{\phi}\dot{\phi} &= 0 \\
g_{\phi\phi}\ddot{\phi} + \frac{\partial g_{\phi\phi}}{\partial\zeta}\dot{\zeta}\dot{\phi} + \frac{\partial g_{\phi\phi}}{\partial\phi}\dot{\phi}\dot{\phi} - \frac{1}{2}\frac{\partial g_{\zeta\zeta}}{\partial\phi}\dot{\zeta}\dot{\zeta} - \frac{1}{2}\frac{\partial g_{\phi\phi}}{\partial\phi}\dot{\phi}\dot{\phi} &= 0
\end{aligned}
$$

把 $g_{\zeta\zeta} = -1$, $g_{\phi\phi} = -\sinh^2\zeta$ 代入，得

$$
\begin{aligned}
\ddot{\zeta} - \sinh\zeta\cosh\zeta\,\dot{\phi}\dot{\phi} &= 0 \\
\sinh\zeta\,\ddot{\phi} + 2\cosh\zeta\,\dot{\zeta}\dot{\phi} &= 0
\end{aligned}
$$

这就是双叶双曲面上的上述度规所决定的短程线微分方程组。很容易想到，这个方程组的解是‘大双曲线’，即过原点的平面与该曲面的交线。我们来验证一下这个猜测。

单位双叶双曲面上的大双曲线可以用方程组

$$
\begin{aligned}
ax + by + cz &= 0 \\
-x^2 - y^2 + z^2 &= 1
\end{aligned}
$$

来刻画，其中 a, b, c 不全为零。按前面讲过的变量代换，把 x, y, z 用 ζ, ϕ 表示，则得到双叶双曲面上大双曲线的方程：

$$a \sinh\zeta \cos\phi + b \sinh\zeta \sin\phi + c \cosh\zeta = 0$$

把 ζ, ϕ 都看做是时间的函数，方程两边对时间求导，得：

$$a \cosh\zeta \cos\phi\, \dot{\zeta} - a \sinh\zeta \sin\phi\, \dot{\phi} + b \cosh\zeta \sin\phi\, \dot{\zeta} + b \sinh\zeta \cos\phi\, \dot{\phi} + c \sinh\zeta\, \dot{\zeta} = 0$$

这是 $\dot{\zeta}$ 和 $\dot{\phi}$ 之间的关系。此式两边再对时间求导，得：

$$
\begin{aligned}
a \sinh\zeta \cos\phi\, \dot{\zeta}\dot{\zeta} - a \cosh\zeta \sin\phi\, \dot{\zeta}\dot{\phi} + a \cosh\zeta \cos\phi\, \ddot{\zeta} \\
-a \cosh\zeta \sin\phi\, \dot{\zeta}\dot{\phi} - a \sinh\zeta \cos\phi\, \dot{\phi}\dot{\phi} - a \sinh\zeta \sin\phi\, \ddot{\phi} \\
+b \sinh\zeta \sin\phi\, \dot{\zeta}\dot{\zeta} + b \cosh\zeta \cos\phi\, \dot{\zeta}\dot{\phi} + b \cosh\zeta \sin\phi\, \ddot{\zeta} \\
+b \cosh\zeta \cos\phi\, \dot{\zeta}\dot{\phi} - b \sinh\zeta \sin\phi\, \dot{\phi}\dot{\phi} + b \sinh\zeta \cos\phi\, \ddot{\phi} \\
+c \cosh\zeta\, \dot{\zeta}\dot{\zeta} + c \sinh\zeta\, \ddot{\zeta} \;&=\; 0
\end{aligned}
$$

其中的三个 $\dot{\zeta}\dot{\zeta}$ 项显见系数和为零。再把 $\ddot{\zeta}$ 和 $\ddot{\phi}$ 的表达式代入，不难验证，$\dot{\phi}\dot{\phi}$ 和 $\dot{\zeta}\dot{\phi}$ 的项的系数和分别也都为零。因此，大双曲线确实是双叶双曲面上前述度规所决定的短程线。

8.11.13 正交弧几何

如前所述，把闵氏双叶双曲面张到平面上，我们就可以在复平面上建立坐标系：

$$
\begin{aligned}
u &= \tanh\frac{\zeta}{2} \cos\phi = \frac{\sinh\zeta}{1 + \cosh\zeta} \cos\phi \\
v &= \tanh\frac{\zeta}{2} \sin\phi = \frac{\sinh\zeta}{1 + \cosh\zeta} \sin\phi
\end{aligned}
$$

微元的变换关系是

$$
\begin{aligned}
\mathrm{d}u &= \frac{1}{1 + \cosh\zeta} (\cos\phi\, \mathrm{d}\zeta - \sinh\zeta \sin\phi\, \mathrm{d}\phi) \\
\mathrm{d}v &= \frac{1}{1 + \cosh\zeta} (\sin\phi\, \mathrm{d}\zeta + \sinh\zeta \cos\phi\, \mathrm{d}\phi)
\end{aligned}
$$

该变换的行列式为

$$|\Phi| = \frac{\sinh\zeta}{(1+\cosh\zeta)^2}$$

易见其逆变换为

$$\begin{aligned}
\mathrm{d}\zeta &= \frac{1+\cosh\zeta}{\sinh\zeta}(\sinh\zeta\cos\phi\,\mathrm{d}u + \sinh\zeta\sin\phi\,\mathrm{d}v) \\
\mathrm{d}\phi &= \frac{1+\cosh\zeta}{\sinh\zeta}(-\sin\phi\,\mathrm{d}u + \cos\phi\,\mathrm{d}v)
\end{aligned}$$

因而，与上述单位双叶双曲面上的度规等距的度规是

$$g_{uu} = -(1+\cosh\zeta)^2, \quad g_{uv} = g_{vu} = 0, \quad g_{vv} = -(1+\cosh\zeta)^2,$$

其中的非零分量用 u, v 来表示，即得 $g_{ij} = \dfrac{-4}{(1-u^2-v^2)^2}\delta_{ij}$。

把该度规代入短程线的微分方程组，得

$$\begin{aligned}
(1-u^2-v^2)\ddot{u} + 2u\dot{u}\dot{u} + 4v\dot{v}\dot{u} - 2u\dot{v}\dot{v} &= 0 \\
(1-u^2-v^2)\ddot{v} + 4u\dot{u}\dot{v} + 2v\dot{v}\dot{v} - 2v\dot{u}\dot{u} &= 0
\end{aligned}$$

这个方程组的解我们也猜得到：复平面上正交弧几何的短程线是正交圆。下面，我们来验证正交圆确实满足这个方程组。

复平面上正交弧几何中的任何一个正交圆可以用方程

$$(u - a\cos\beta)^2 + (v - a\sin\beta)^2 = a^2 - 1$$

来刻画，其中 a 表示正交圆圆心到基准圆圆心的距离，β 表示两个圆心的连线与选定的参考线，比如 x 轴，的夹角。该方程还可以变换成另一种形式：

$$(1-u^2-v^2) = 2(1 - a\cos\beta\, u - a\sin\beta\, v)$$

方程两边对时间求导，得 $\dot{u}, \dot{v}$ 之间的关系

$$(u - a\cos\beta)\dot{u} + (v - a\sin\beta)\dot{v} = 0$$

再对时间求导，得：

$$\dot{u}\dot{u} + (u - a\cos\beta)\ddot{u} + \dot{v}\dot{v} + (v - a\sin\beta)\ddot{v} = 0$$

两边同乘 $(1-u^2-v^2)$，消去 $\ddot{u}$ 和 $\ddot{v}$。再代入 $\dot{u}, \dot{v}$ 之间的关系，可证方程成立，即正交圆是该度规所决定的短程线。

8.11.14 平面几何的短程线

了解这些曲面之后，我们对平面的理解也更深刻了。比如，平面直角坐标系 (x, y) 上的欧氏度规 δ_{ij}。它各个分量都是常数，对 x 和对 y 的偏导都为零，

因此，它决定的短程线的微分方程是：$\ddot{x}=0, \quad \ddot{y}=0$。这当然就是平面上的直线。其实，这就是物理上的牛顿第一定律。

即使换成曲线坐标系，如果度规保长的话，也不会改变短程线的性质。比如，象我们前面讲过的那样，把欧氏平面的直角坐标系 (x,y) 换成极坐标系 (r,θ)，度规换成 $g_{rr}=1, \quad g_{\theta\theta}=r^2, \quad g_{r\theta}=g_{\theta r}=0$。代入短程线的微分方程组，得

$$\begin{aligned}\ddot{r}-r\dot{\theta}\dot{\theta} &= 0\\ r\ddot{\theta}+2\dot{r}\dot{\theta} &= 0\end{aligned}$$

用极坐标表示的直线的方程为

$$r\sin(\theta+\theta_0)=d$$

其中 θ_0 刻画直线的方向，而 d 则是原点到直线的距离。不难验证，直线满足上述短程线的微分方程组。

把直线方程的两边对时间求导，得

$$\sin(\theta+\theta_0)\,\dot{r}+r\cos(\theta+\theta_0)\,\dot{\theta}=0$$

再求导，把微分方程组决定的 $\ddot{r}$ 和 $\ddot{\theta}$ 的表达式代入，即可得证。

闵氏平面上的情况与欧氏平面类似。用直角坐标系 (x,y) 配上常数度规 η_{ij} 就可以刻画平直的二维空间。此时的短程线微分方程组也是 $\ddot{x}=0, \quad \ddot{y}=0$。它的解当然也是直线。

换成双曲极坐标 (s,ζ) 和配套的度规 $g_{ss}=-1, \quad g_{\zeta\zeta}=-s^2, \quad g_{s\zeta}=g_{\zeta s}=0$。代入短程线的微分方程组，得

$$\begin{aligned}\ddot{s}+s\dot{\zeta}\dot{\zeta} &= 0\\ s\ddot{\zeta}+2\dot{s}\dot{\zeta} &= 0\end{aligned}$$

用双曲极坐标表示的直线的方程为

$$s\sinh(\zeta+\zeta_0)=d$$

其中 ζ_0 刻画直线的方向，而 d 则是原点到直线的双曲距离。不难验证，直线满足上述短程线的微分方程组。

把直线方程的两边对时间求导，得

$$\sinh(\zeta+\zeta_0)\,\dot{s}+s\cosh(\zeta+\zeta_0)\,\dot{\zeta}=0$$

再求导，把微分方程组决定的 $\ddot{s}$ 和 $\ddot{\zeta}$ 的表达式代入，即可得证。

8.11.15 共点圆几何

平直空间并不要求测地线是直线。只要度规定义得当，测地线为曲线的空间也可以是平直的。比如前面讲过的共点圆几何。

把欧氏度规平面上的所有点，做关于单位圆的反演，那么所有的直线，就都被变换成过坐标原点的圆。原先的以直线为测地线的欧氏平面几何，就变成了以坐标原点为所共之点的共点圆几何，也就是抛物几何。

平面反演变换是一个曲线坐标系变换，把欧氏度规平面上直角坐标系中除坐标原点之外的任意一点 (x, y)，变换到同一坐标系中另外一点 (u, v)，变换关系为：

$$u = \frac{x}{x^2+y^2}, \qquad v = \frac{y}{x^2+y^2}.$$

这样，就满足 $(x^2+y^2)(u^2+v^2) = 1$，因而这是关于单位圆的反演变换。该变换的逆为：

$$x = \frac{u}{u^2+v^2}, \qquad y = \frac{v}{u^2+v^2}.$$

显有：

$$\frac{\partial x}{\partial u} = \frac{v^2-u^2}{(u^2+v^2)^2}, \quad \frac{\partial x}{\partial v} = \frac{-2uv}{(u^2+v^2)^2}, \quad \frac{\partial y}{\partial u} = \frac{-2uv}{(u^2+v^2)^2}, \quad \frac{\partial y}{\partial v} = \frac{u^2-v^2}{(u^2+v^2)^2}.$$

代入 $g_{ij} = \partial_i^m \partial_j^n h_{mn}$，得 $g_{ij} = \dfrac{1}{(u^2+v^2)^2}\delta_{ij}$。代入短程线微分方程组，得：

$$\begin{aligned}
(u^2+v^2)\ddot{u} - 2u\dot{u}\dot{u} - 4v\dot{u}\dot{v} + 2u\dot{v}\dot{v} \;&=\; 0 \\
(u^2+v^2)\ddot{v} - 4u\dot{u}\dot{v} - 2v\dot{v}\dot{v} + 2v\dot{u}\dot{u} \;&=\; 0
\end{aligned}$$

现在我们来验证，共点圆满足该方程组。原来的直线 $ax+by+c=0$，被变换成共点圆 $au+bv+c(u^2+v^2)=0$。两边对时间求导，得：

$$a\dot{u} + b\dot{v} + 2c(u\dot{u} + v\dot{v}) = 0$$

也就是

$$(a+2cu)\dot{u} + (b+2cv)\dot{v} = 0$$

这个式子反映了 $\dot{u}$ 和 $\dot{v}$ 的关系，在下面的推导中用得着。把前一个式子两边对时间再求导，得：

$$a\ddot{u} + b\ddot{v} + 2c(\dot{u}\dot{u} + u\ddot{u} + \dot{v}\dot{v} + v\ddot{v}) = 0$$

两边同乘 (u^2+v^2)，把短程线微分方程组的 $\ddot{u}$ 和 $\ddot{v}$ 的表达式代入，就得到 $\dot{u}$ 和 $\dot{v}$ 的二次型，利用共点圆方程以及 $\dot{u}$ 和 $\dot{v}$ 的关系，不难验证该二次型恒等于零。

想想看，平面上两点之间有且只有一条直线，沿这条直线走路程最短。‘平面’上两点之间有且只有一个圆能够过坐标原点，沿这个圆走‘路程’最

短。这两种说法居然描述的是同一个客观事实，两者完全等价！真是可笑而又神奇。

实际上，我们可以把欧氏度规平面上的直线用球极投影映射到球面上，成为球面上过投影中心的圆。然后，把这样的圆关于赤道面做一个镜像变换，使南北半球互换，再用同样的球极投影把球面上的圆映射回平面上，就是共点圆。这个共点圆与原来的直线互为反演。

8.11.16 球面上的平面几何

不仅平面上的曲线可以用来刻画平直空间，连球面上的曲线，都可以刻画平直空间。过径圆几何是把球面做到平面上，现在我们来把平面做到球面上。

以球面南极点为投影中心，将上述共点圆几何投影到球面上，则成为球面上以北极点为所共之点的共点圆几何。该几何同样刻画平直空间。这个球极投影相当于做了一个变量代换：

$$u = \frac{\sin\theta}{1+\cos\theta}\cos\phi, \qquad v = \frac{\sin\theta}{1+\cos\theta}\sin\phi$$

平面共点圆几何中微小向量的模长平方 $\mathrm{d}l^2 = \dfrac{1}{(u^2+v^2)^2}(\,\mathrm{d}u^2 + \,\mathrm{d}v^2)$，变换成 $\mathrm{d}l^2 = \dfrac{1}{(1-\cos\theta)^2}(\,\mathrm{d}\theta^2 + \sin^2\theta\,\,\mathrm{d}\phi^2)$。也就是说，在这个球面共点圆几何中，我们定义度规

$$g_{\theta\theta} = \frac{1}{(1-\cos\theta)^2}, \quad g_{\phi\phi} = \frac{\sin^2\theta}{(1-\cos\theta)^2}, \quad g_{\theta\phi} = g_{\phi\theta} = 0$$

代入短程线微分方程组，得到短程线方程组：

$$\begin{aligned}\sin\theta\,\ddot\theta - (1+\cos\theta)\dot\theta\dot\theta + \sin^2\theta\,\dot\phi\dot\phi &= 0\\ \sin\phi\,\ddot\phi - 2\dot\theta\dot\phi &= 0\end{aligned}$$

把平面共点圆几何中共点圆方程的 u, v 代换成球面几何中的 θ, ϕ 的表达式，就得到球面共点圆几何的测地线：

$$a\sin\theta\cos\phi + b\sin\theta\sin\phi + c(1-\cos\theta) = 0$$

用三维欧氏空间直角坐标系的 x, y, z 表示就是：

$$\begin{aligned}ax + by + c(1-z) &= 0\\ x^2 + y^2 + z^2 &= 1\end{aligned}$$

这是过单位球面北极点的平面与单位球面的交线。

我们都知道球面是正曲率的曲面，其上测地线三角形的内角和大于 π。刚才我们在球面上做了一个零曲率的平面，其上测地线三角形的内角和等于 π。

8.11.17 球面上的双曲几何

我们还可以在球面上做一个负曲率的曲面，其上测地线三角形的内角和小于 π。

实际上，我们已经在平面上做了一个负曲率的曲面，即正交圆几何。那是用把闵氏度规空间中单位双叶双曲面上的大双曲线南极投影到平面上而得到的。现在，用和前面共点圆几何和过径圆几何同样的球极投影，把平面上的正交圆几何投影到球面上，即把点投影过去，把度规也变换过去，就得到了球面上的正交圆几何。

把球极投影的变量代换关系代入平面正交圆几何的微小向量模长平方 $\mathrm{d}l^2 = \dfrac{-4}{(1-u^2-v^2)^2}(\,\mathrm{d}u^2+\,\mathrm{d}v^2)$，得 $\mathrm{d}l^2 = \dfrac{-1}{\cos^2\theta}(\,\mathrm{d}\theta^2+\sin^2\theta\,\,\mathrm{d}\phi^2)$。即，该球面正交圆几何的度规是

$$g_{\theta\theta} = \frac{-1}{\cos^2\theta}, \quad g_{\phi\phi} = \frac{-\sin^2\theta}{\cos^2\theta}, \quad g_{\theta\phi} = g_{\phi\theta} = 0$$

该几何的测地线是：

$$a\cos\beta\sin\theta\cos\phi + a\sin\beta\sin\theta\sin\phi = 1$$

这个是通过把 u, v 的表达式代入平面正交圆方程而得到的。用三维欧氏空间直角坐标系的 x, y, z 表示就是：

$$\begin{aligned} a\cos\beta\, x + a\sin\beta\, y \,&=\, 1 \\ x^2+y^2+z^2 \,&=\, 1 \end{aligned}$$

这是垂直于赤道面的平面与单位球面的交线。这正是双曲几何的布安加雷模型。

8.11.18 平面上用直线表述的双曲几何

用平行光把球面正交圆垂直投影到赤道面上，我们就得到一个测地线为直线的负曲率曲面模型。没错哦，这些测地线看起来是直的，但是度规定义出来的长度却是不均匀的，因而，它们实际上刻画的是曲面。怎么个不均匀法呢？我们把这个度规求出来看看就知道了。

把球面上一点 (θ, ϕ) 垂直投影到赤道面上，在赤道面上建立适当的直角坐标系 (x, y) ，使得：

$$x = \sin\theta\cos\phi, \qquad y = \sin\theta\sin\phi$$

则微小向量各分量之间的变换关系为

$$\begin{aligned} \mathrm{d}x \,&=\, \cos\theta\cos\phi\,\,\mathrm{d}\theta - \sin\theta\sin\phi\,\,\mathrm{d}\phi \\ \mathrm{d}y \,&=\, \cos\theta\sin\phi\,\,\mathrm{d}\theta + \sin\theta\cos\phi\,\,\mathrm{d}\phi \end{aligned}$$

逆变换为

$$\begin{aligned}
\mathrm{d}\theta &= \frac{1}{\cos\theta}(\cos\phi\ \mathrm{d}x + \sin\phi\ \mathrm{d}y) \\
\sin\theta\ \mathrm{d}\phi &= -\sin\phi\ \mathrm{d}x + \cos\phi\ \mathrm{d}y
\end{aligned}$$

代入球面正交圆几何的度规，得

$$\begin{aligned}
-\mathrm{d}l^2 &= \frac{1}{\cos^2\theta}(\ \mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\phi^2) \\
&= \frac{1}{\cos^4\theta}[\cos^2\phi\ \mathrm{d}x^2 + \sin 2\phi\ \mathrm{d}x\,\mathrm{d}y + \sin^2\phi\ \mathrm{d}y^2 \\
&\quad + \cos^2\theta(\sin^2\phi\ \mathrm{d}x^2 - \sin 2\phi\ \mathrm{d}x\,\mathrm{d}y + \cos^2\phi\ \mathrm{d}y^2)] \\
&= \frac{1}{(1-x^2-y^2)^2}[(1-y^2)\,\mathrm{d}x^2 + 2xy\,\mathrm{d}x\,\mathrm{d}y + (1-x^2)\,\mathrm{d}y^2]
\end{aligned}$$

可见，这种平面上的直线刻画的负曲率曲面所使用的度规为

$$g_{xx} = \frac{-(1-y^2)}{(1-x^2-y^2)^2},\quad g_{xy} = g_{yx} = \frac{-xy}{(1-x^2-y^2)^2},\quad g_{yy} = \frac{-(1-x^2)}{(1-x^2-y^2)^2}$$

如图 8.18 所示。 上述方法是从大双曲线上一点 Q 出发，通过南极投影映

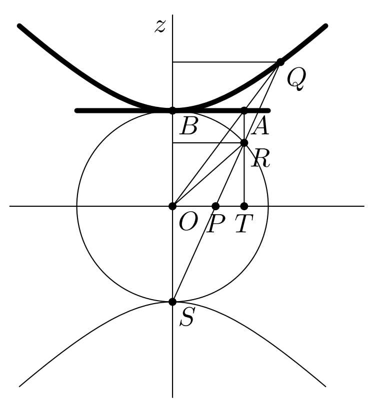

图 8.18: 平面上的直线也可以是双曲几何的测地线

射到 P，大双曲线变成平面正交圆，再通过球极投影映射到 R，平面正交圆变成球面正交圆，再通过垂直投影映射到 T，球面正交圆变成平面直线。实际上，还有另一种更加直接的映射方法：通过中心投影从 Q 直接映射到 A，大双曲线直接变成直线。

记 $\angle BOQ$ 的双曲角度为 ζ，$\angle BOR$ 的普通角度为 θ。首先我们来证 $\cos\theta\cosh\zeta = 1$。从图上不难看出

$$\frac{\sin\theta}{1+\cos\theta} = \frac{\sinh\zeta}{1+\cosh\zeta} = OP$$

将前两式平方，用三角及双曲三角恒等式，易证

$$\frac{1-\cos\theta}{1+\cos\theta}=\frac{\cosh\zeta-1}{\cosh\zeta+1}$$

进而

$$\frac{\cos\theta}{1+\cos\theta}=\frac{1}{\cosh\zeta+1}$$

两边求倒数，即得 $\cos\theta\cosh\zeta=1$。

这样，代入

$$\sin\theta=\frac{1+\cos\theta}{1+\cosh\zeta}\sinh\zeta=\tanh\zeta$$

而 $\tanh\zeta$ 正是 AB 的长度。因此，我们可以在平面上建立适当的直角坐标系使得

$$x=\tanh\zeta\cos\phi,\qquad y=\tanh\zeta\sin\phi$$

这个变换把大双曲线映射为大双曲线所在平面与过 AB 的水平面的交线。由于 $\sin\theta=\tanh\zeta$，所以这个变换与球面正交圆垂直投影法是一致的。

度规的情况也是殊途同归。从闵氏度规空间中双叶双曲面上的度规 $\mathrm{d}l^2=-\mathrm{d}\zeta^2-\sinh^2\zeta\,\mathrm{d}\phi^2$ 出发，按上述变量代换，计算出来的度规与球面正交圆垂直投影法的结果完全一致。

8.11.19 平面上用直线表述的球面几何

只要度规定义得合适，平面上的直线也能刻画正曲率曲面。如图 8.19 所示。我们把球面大圆上任意一点 R 中心投影到点 A，则球面大圆变成过点 B

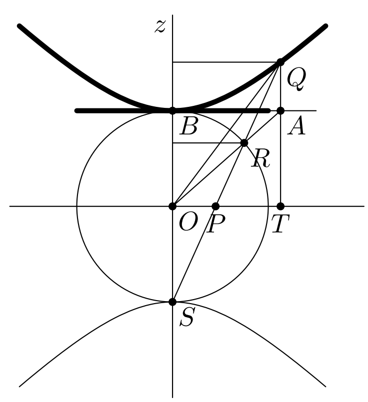

图 8.19: 平面上的直线也可以是椭圆几何的测地线

的水平面上的直线。在此平面上建立直角坐标系使得

$$x=\tan\theta\cos\phi,\qquad y=\tan\theta\sin\phi$$

微小向量各分量的变换关系为

$$\begin{aligned}
\mathrm{d}x &= \frac{1}{\cos^2\theta}\cos\phi\ \mathrm{d}\theta - \tan\theta\sin\phi\ \mathrm{d}\phi \\
\mathrm{d}y &= \frac{1}{\cos^2\theta}\sin\phi\ \mathrm{d}\theta + \tan\theta\cos\phi\ \mathrm{d}\phi
\end{aligned}$$

逆变换为

$$\begin{aligned}
\mathrm{d}\theta &= \cos^2\theta(\cos\phi\ \mathrm{d}x + \sin\phi\ \mathrm{d}y) \\
\mathrm{d}\phi &= \frac{\cos^2\theta}{\sin\theta}\left(\frac{-1}{\cos\theta}\sin\phi\ \mathrm{d}x + \frac{1}{\cos\theta}\cos\phi\ \mathrm{d}y\right)
\end{aligned}$$

代入球面大圆的度规，得

$$\begin{aligned}
\mathrm{d}l^2 &= \mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\phi^2 \\
&= \frac{1}{(1+x^2+y^2)^2}[(1+y^2)\,\mathrm{d}x^2 - 2xy\,\mathrm{d}x\,\mathrm{d}y + (1+x^2)\,\mathrm{d}y^2]
\end{aligned}$$

即，这个以平面直线为测地线的‘单叶球面几何’的度规为

$$g_{xx} = \frac{1+y^2}{(1+x^2+y^2)^2},\quad g_{xy} = g_{yx} = \frac{-xy}{(1+x^2+y^2)^2},\quad g_{yy} = \frac{1+x^2}{(1+x^2+y^2)^2}$$

8.11.20 双叶双曲面上的球面几何

现在我们有两种方法得到双叶双曲面上的‘单叶球面几何’。一种是用垂直投影把点 A 映射到点 Q，另一种是用球极投影把赤道面上的过径圆上的点 P 映射到点 Q。两种方法的结果是一样的。原因是，前面讲过，该图中 $\cos\theta\cosh\zeta = 1$，代入 $\sinh\zeta = \frac{1+\cosh\zeta}{1+\cos\theta}\sin\theta = \tan\theta$。即点 Q 恰在点 A 的正上方。

我们来求一下这种几何的度规。先用对平面过径圆做双曲南极投影的方法。双曲南极投影所决定的曲线坐标变换关系是

$$u = \frac{\sinh\zeta}{1+\cosh\zeta}\cos\phi,\quad v = \frac{\sinh\zeta}{1+\cosh\zeta}\sin\phi$$

平面过径圆几何的度规是 $\mathrm{d}l^2 = \frac{4}{(1+u^2+v^2)^2}(\,\mathrm{d}u^2 + \,\mathrm{d}v^2)$。注意到

$$1+u^2+v^2 = \frac{2\cosh\zeta}{1+\cosh\zeta},\quad \left(\frac{\sinh\zeta}{1+\cosh\zeta}\right)' = \frac{1}{1+\cosh\zeta}$$

所以，微小向量各分量的变换关系是

$$\begin{aligned}
\mathrm{d}u &= \frac{1}{1+\cosh\zeta}(\cos\phi\ \mathrm{d}\zeta - \sinh\zeta\sin\phi\ \mathrm{d}\phi) \\
\mathrm{d}v &= \frac{1}{1+\cosh\zeta}(\sin\phi\ \mathrm{d}\zeta + \sinh\zeta\cos\phi\ \mathrm{d}\phi)
\end{aligned}$$

代入微小向量模长平方的表达式，得 $\mathrm{d}l^2 = \dfrac{1}{\cosh^2\zeta}(\,\mathrm{d}\zeta^2 + \sinh^2\zeta\,\mathrm{d}\phi^2)$。即，这个做在双叶双曲面上的‘单叶球面’的度规是：

$$g_{\zeta\zeta} = \frac{1}{\cosh^2\zeta},\quad g_{\phi\phi} = \frac{\sinh^2\zeta}{\cosh^2\zeta},\quad g_{\zeta\phi} = g_{\phi\zeta} = 0$$

换另一种方法，从平面上直线的‘单叶球面’几何垂直投影到双叶双曲面上。这时的坐标变换关系是

$$x = \sinh\zeta\cos\phi,\qquad y = \sinh\zeta\sin\phi$$

微小向量模长平方为

$$\mathrm{d}l^2 = \frac{1}{(1+x^2+y^2)^2}[(1+y^2)\,\mathrm{d}x^2 - 2xy\,\mathrm{d}x\,\mathrm{d}y + (1+x^2)\,\mathrm{d}y^2]$$

微小向量各分量的变换关系是

$$\begin{aligned}
\mathrm{d}x &= \cosh\zeta\cos\phi\ \mathrm{d}\zeta - \sinh\zeta\sin\phi\ \mathrm{d}\phi\\
\mathrm{d}y &= \cosh\zeta\sin\phi\ \mathrm{d}\zeta + \sinh\zeta\cos\phi\ \mathrm{d}\phi
\end{aligned}$$

代入，得

$$\begin{aligned}
\mathrm{d}l^2 &= \frac{1}{\cosh^4\zeta}[\,\mathrm{d}x^2 + \,\mathrm{d}y^2 + (y\,\mathrm{d}x - x\,\mathrm{d}y)^2]\\
&= \frac{1}{\cosh^4\zeta}[\cosh^2\zeta\ \mathrm{d}\zeta^2 + \sinh^2\zeta\ \mathrm{d}\phi^2 + \sinh^4\zeta\ \mathrm{d}\phi^2]\\
&= \frac{1}{\cosh^2\zeta}(\,\mathrm{d}\zeta^2 + \sinh^2\zeta\ \mathrm{d}\phi^2)
\end{aligned}$$

与前一种方法得到的结果完全一致。

8.11.21 反演变换与球面共点圆几何

如图 8.20 所示。 前面讲过，把欧氏度规平面上的直线关于单位圆做反演就是共点圆。该图显示了直线上任意一点 Q，经反演变成共点圆上的一点 P。从南极 S 做球极投影，平面上的共点圆就被映射为以北极 N 为所共之点的球面共点圆，其中任意点 P 被映射为球面共点圆上的点 A。现在我们要说的是，点 A 可以直接从点 Q 经北极球极投影而得到。

前面我们利用的是 $\triangle SPO \sim \triangle SAB$，得 $OP = \dfrac{\sin\theta}{1+\cos\theta}$。现在我们利用 $\triangle NAB \sim \triangle NQO$，得 $OQ = \dfrac{\sin\theta}{1-\cos\theta}$。即，从欧氏度规平面几何到球面共点圆几何的坐标变换关系是：

$$x = \frac{\sin\theta}{1-\cos\theta}\cos\phi,\qquad y = \frac{\sin\theta}{1-\cos\theta}\sin\phi$$

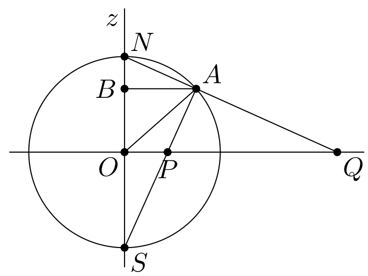

图 8.20: 反演变换与球面共点圆几何

微小向量各分量的变换关系是

$$
\begin{aligned}
\mathrm{d}x &= \frac{-1}{1-\cos\theta}(\cos\phi\ \mathrm{d}\theta + \sin\theta\sin\phi\ \mathrm{d}\phi) \\
\mathrm{d}y &= \frac{-1}{1-\cos\theta}(\sin\phi\ \mathrm{d}\theta - \sin\theta\cos\phi\ \mathrm{d}\phi)
\end{aligned}
$$

代入微小向量模长平方的表达式

$$
\begin{aligned}
\mathrm{d}l^2 &= \mathrm{d}x^2 + \mathrm{d}y^2 \\
&= \frac{1}{(1-\cos\theta)^2}(\ \mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\phi^2)
\end{aligned}
$$

该式所表达的度规与前一种方法得到的结果完全一致。

8.11.22 双叶双曲面上的平面几何

现在我们来看双叶双曲面上的情况。如图 8.21 所示。 欧氏度规平面上任

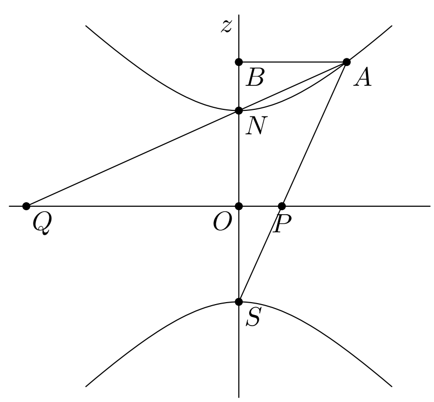

图 8.21: 对径变换与双叶双曲面上的共点二次曲线几何

意直线上一点 Q，经北极投影映射为双叶双曲面上的点 A。整个直线被映射为

双叶双曲面上的一条二次曲线。由 $\triangle NBA \sim \triangle NOQ$ 得坐标变换关系

$$x = \frac{\sinh\zeta}{1-\cosh\zeta}\cos\phi, \qquad y = \frac{\sinh\zeta}{1-\cosh\zeta}\sin\phi$$

微小向量各分量的变换关系为

$$\begin{aligned} \mathrm{d}x &= \frac{-1}{1-\cosh\zeta}(\cos\phi\ \mathrm{d}\zeta + \sinh\zeta\sin\phi\ \mathrm{d}\phi) \\ \mathrm{d}y &= \frac{-1}{1-\cosh\zeta}(\sin\phi\ \mathrm{d}\zeta - \sinh\zeta\cos\phi\ \mathrm{d}\phi) \end{aligned}$$

代入微小向量模长平方的表达式，得

$$\begin{aligned} \mathrm{d}l^2 &= \mathrm{d}x^2 + \mathrm{d}y^2 \\ &= \frac{1}{(1-\cosh\zeta)^2}(\mathrm{d}\zeta^2 + \sinh^2\zeta\,\mathrm{d}\phi^2) \end{aligned}$$

这就是这种建立在单位双叶双曲面上的平直空间的度规。该平直空间的测地线是过点 N 的平面与单位双叶双曲面的交线，可能是椭圆，抛物线，或双曲线。

很自然地，我们会想到一个问题：把这种几何通过南极投影再投回到平面上来，会是一种什么几何呢？还是看图 8.21。在南极投影下，点 A 被投影为平面上的点 P。显见，从点 Q 到点 P 是个对径变换。也就是说，把普通平面几何做一个关于单位圆的对径变换，就得到了我们想要的结果。

前面说过，对径变换等于反演变换与中心对称变换的复合。因此，有坐标变换关系

$$x = \frac{-u}{u^2+v^2}, \qquad y = \frac{-v}{u^2+v^2}$$

微小向量各分量的变换关系为

$$\begin{aligned} \mathrm{d}x &= \frac{1}{(u^2+v^2)^2}[(u^2-v^2)\,\mathrm{d}u + 2uv\,\mathrm{d}v] \\ \mathrm{d}y &= \frac{1}{(u^2+v^2)^2}[2uv\,\mathrm{d}u - (u^2-v^2)\,\mathrm{d}v] \end{aligned}$$

代入微小向量模长平方的表达式，得

$$\begin{aligned} \mathrm{d}l^2 &= \mathrm{d}x^2 + \mathrm{d}y^2 \\ &= \frac{1}{(u^2+v^2)^2}(\mathrm{d}x^2 + \mathrm{d}y^2) \end{aligned}$$

这个度规与反演变换得到的共点圆几何的度规是一样的。实际上，两者从同一条直线变换来的测地线做成中心对称。

但是这种映射也确有自己的独到之处。如图 8.22 所示。 过北极的平面与双叶双曲面相交，粗实线所示的平面交成椭圆，点线所示的平面交成抛物线，虚线所示的平面交成双曲线。这些看上去不同类型的二次曲线，经过南极投影映射到平面上，就是图中右半边所示的情况。椭圆被映射为过原点而与单位圆

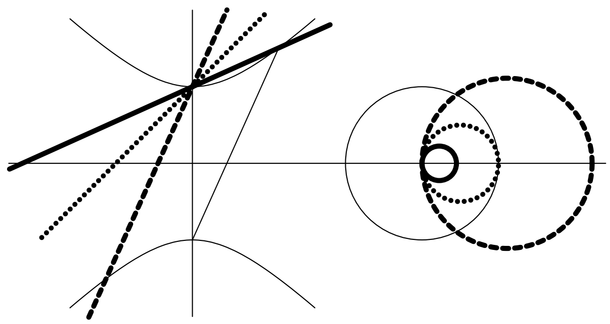

图 8.22: 双叶双曲面上三种二次曲线都被映射为共点圆

不相交的圆；抛物线被映射为过原点而与单位圆相切的圆；双曲线被映射为过原点而与单位圆有两个交点的圆。原本属于不同类型的二次曲线，在这种几何中得到了统一。这些共点二次曲线都是测地线，刻画一个平面几何。

注意，共点圆（或共点二次曲线）并非总与平面几何相联系。前面已经介绍过，同样是直线，配上不同的度规，就可以刻画不同的几何。把这些几何做关于单位圆的反演，就可以得到：同样的共点圆（或共点二次曲线），配上不同的度规，就可以刻画不同的几何。列位看官如果有兴趣，可以自己计算一下这些度规，权作练习。

8.12 射影变换与宇宙模型

8.12.1 点列与线束的底与保底射影变换

射影几何中把直线称为一阶点列的底，把二次曲线称为二阶点列的底。一系列的点经过射影变换变成一系列新的点，称为把一个点列射影变换成另一个点列。如果变换后的点列仍旧在原来的底上，则称为同底两点列的变换，称该射影变换是保底的。

对偶地，射影几何中把点称为一阶线束的底，把二次曲线称为二阶线束的底。原先‘点在曲线上’的概念对偶地变成‘直线与曲线相切’，因此，二阶线束的底是该线束的包络。如果经射影变换后的线束仍旧在原来的底上，则称为同底两线束的变换，称该射影变换是保底的。

8.12.2 透视对应与透视变换

给定一个点，再给定一个点列，给定点与点列中各点的连线构成一个线束。称该线束是从给定点到给定点列的透视线束。透视线束的底称为透视中

心。对偶地，给定点列称为透视点列。透视点列的底称为透视轴。称该透视点列与该透视线束之间成透视对应，也称配景对应。

称与同一个线束成透视对应的两个点列之间也成透视对应。从其中一个点列到另一个点列的射影变换称为透视变换。对偶地，称与同一个点列成透视对应的两个线束之间也成透视对应。

透视对应的独特之处在于：两透视对应点列的底的交点是该对应的不动点，以及对偶地，两透视对应线束的底的连线是该对应的不动直线。反之也成立。任何射影变换都把直线变到直线，如果源点列的底与目标点列的底的交点是不动点，那么该射影变换一定是透视变换。对偶地，任何射影变换都把点变到点。如果源线束的底与目标线束的底的连线是不动直线，那么该射影变换一定是透视变换。

8.12.3 角度、距离、弧长的一致性

如图 8.23、图 8.24、图 8.25 所示。 在原点的微小邻域内，基准圆直径与

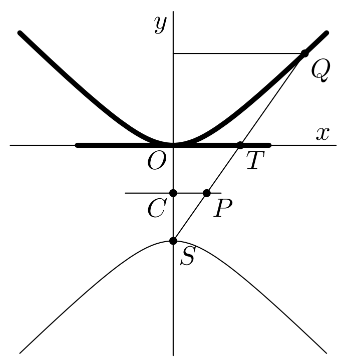

图 8.23: 正交弧双曲距离与辅助双曲线的直接比对

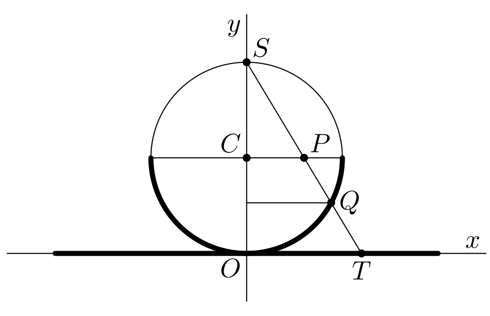

图 8.24: 过径弧椭圆距离与辅助圆的直接比对

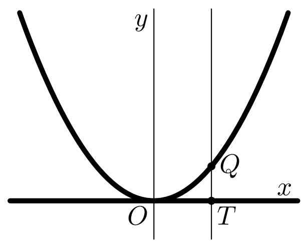

图 8.25: 共点圆抛物距离与辅助抛物线的直接比对

这些辅助二次曲线是局部重合的，两者关于距离的定义都过渡为欧氏几何的距离。

双曲线、椭圆、和抛物线在射影几何中是可以通过射影变换而相互转化的。它们的区别仅仅在于：双曲线与无穷远直线相交，有两个交点；椭圆与无穷远直线不相交，没有交点；抛物线与无穷远直线相切，算是有一个交点。既然无穷远直线可以被变换成普通直线，那么二次曲线的三种类型都可以被变换成椭圆与一个普通直线的关系来考察。

如图 8.26 所示。 图中的圆表示一个一般的二次曲线。无穷远直线用 l_∞ 标

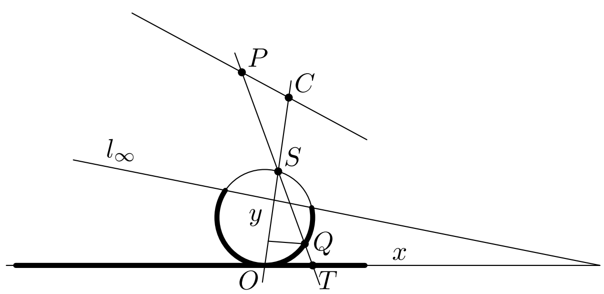

图 8.26: 辅助双曲线比对图的射影本质

出。该二次曲线与无穷远直线有两个交点，因此它是个双曲线。无穷远直线关于该二次曲线的配极点称为该二次曲线的中心，即图上的点 C。过中心的直线称为该二次曲线的直径。与椭圆不同的是，双曲线的直径可以与该双曲线或相交于两点，或不相交，或相切。如果相切，切点必在无穷远直线上。此时的直径称为渐近线。双曲线有两条渐近线。

所有直径中有且只有一条是无穷远直线与 y 轴交点的配极直线，称为对分直径，也称对分轴。图中 P 点就在该二次曲线的对分轴上。对分轴与无穷远直线的交点的配极直线，即图上的 y 轴，称为中分直径，也称中轴。中轴与该二次曲线恒有两个交点。对分轴和中轴统称主轴，也称主方向。

如果给定二次曲线与无穷远直线不相切，那么，实际上我们可以任意选取

二次曲线上的一点，与中心连成直线作为中轴，再做出相应的对分轴。两者是‘仿射正交’的。举例来说，图中直线 CQ 与 Q 点的切线就是仿射正交的。欧氏几何中的两直线正交是仿射正交的一种特例。

上述三条直线做成该二次曲线的‘自配极三线形’。无穷远直线也称绝对轴。因此自配极三线形也称由绝对轴、对分轴和中轴构成的三轴系统。这三条直线交成三点，称三个交点做成该二次曲线的‘自配极三点形’。以这三点建立射影坐标系，取中心 C 为坐标原点，可以使该二次曲线的方程具有很简洁、整齐的形式。这个图与图 8.23 表达的内容是完全一致的，两者只是相差一个射影变换而已。

图 8.27 显示了辅助圆的情形。 这与辅助双曲线几乎一样，只是无穷远直

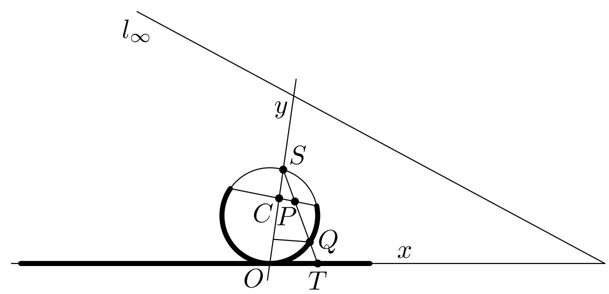

图 8.27: 辅助圆比对图的射影本质

线与二次曲线的对分轴交换了一下角色。显见，椭圆的任何直径都与椭圆有两个交点。

图 8.28 显示的是辅助抛物线的情形。 从图中可以看出，二次曲线与无穷

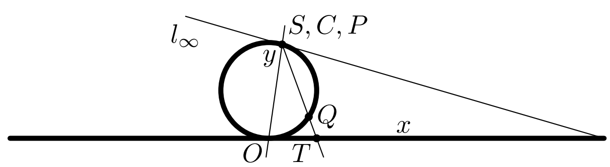

图 8.28: 辅助抛物线比对图的射影本质

远直线相切，切点就是抛物线的中心，对分轴与无穷远直线合二为一。由于欧氏平面上没有无穷远直线，所以，我们有时也说，抛物线没有中心，也没有对分轴。其他的直径全都平行于中轴。图中被比对的，不再是一个线段，而是整个直线。整个二次曲线与整个比对直线完全逐点一一对应。

从图 8.26、图 8.27和图 8.28 中我们看到，同一个射影变换，可以表现为同底两个一阶点列之间的变换，也可以表现为同底两个二阶点列之间的变换，还可以表现为同底两个一阶线束之间的变换。实际上，还可以表现为同底两个二

阶线束之间的变换，只是图上没有显示出来。所以，我们无论是讨论直线上两点之间的‘距离’，还是讨论二次曲线上两点之间的‘弧长’，还是讨论线束上两线之间的‘角度’，还是讨论二次曲线上两切线之间的‘偏转角度’，其实都是在讨论同底射影变换群的同一种性质。

8.12.4 两个同底一阶点列之间的射影变换

我们来看一个命题：在射影平面上，给定射影变换把一直线上的一阶点列变换成另一直线上的点列，则，要么该变换是透视变换，要么该变换是两个透视变换的积。

假设该命题成立，那么对于两点列同底的情形，可以选任意一个不同的直线为辅助线，选三直线外任意一点为中心，把其中一个点列透视变换到辅助线上。从而转化成两点列不同底的情形。这样，我们就证明了：同底两个一阶点列之间的射影变换，在射影平面上，总可表成至多三个透视变换的积。

现在我们来证，该命题的确成立。如图 8.29 所示。 三对对应点 $A \to$

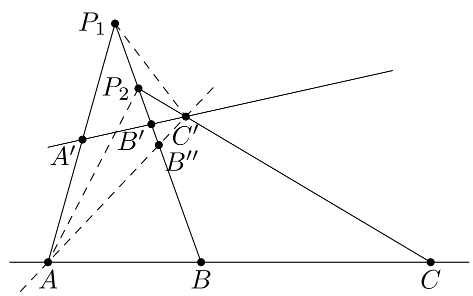

图 8.29: 不同底两个一阶点列间的射影变换可表为两透视之积

$A', B \to B', C \to C'$ 确定了两个一阶点列之间的射影变换。AA' 与 BB' 交于点 P_1。如果该点在 CC' 上，那么该射影变换为透视变换。如果该点不在 CC' 上，那么，设 CC' 与 BB' 的交点为 P_2。如图做三条辅助线。先以 P_2 为透视中心，把直线 AC 上的点列变换到直线 AC' 上。注意，A 为该对应的不动点。再以 P_1 为透视中心，把直线 AC' 上的点列变换到直线 $A'C'$ 上。注意，C' 为该对应的不动点。这样，我们就把给定射影变换表成了两个透视变换的积。

注意到，图中过渡用的辅助线 AC' 是两线束 P_1 和 P_2 的透视轴。直线 P_1P_2 当然就是该透视对应的不变直线。选用不同的过渡直线，可以做出不同的分解。如图 8.30 所示。 AB' 与 $A'B$ 交于点 B''，AC' 与 $A'C$ 交于点 C''。以这两点的连线为过渡直线，先以 A' 为透视中心，把点列 (ABC) 对应到 $(A''B''C'')$，F_1 为不动点。再以 A 为透视中心，把点列 $(A''B''C'')$ 对应到 $(A'B'C')$，F_2 为不动点。这样也能把从 (ABC) 到 $(A'B'C')$ 的射影对应分解为

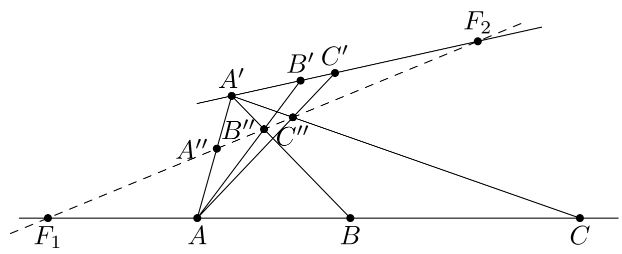

图 8.30: 射影变换分解为两透视之积的另一个方案

两个透视对应的积。显见，过渡直线是两个透视线束的透视轴。AA' 是其不动直线。

即使选定了不动直线，仍然有几种不同的方案来确定两个透视中心。图 8.29 中，我们用源点列和目标点列中两对对应点连线与不动直线的交点做透视中心。图 8.30 中，我们用源点列和过渡点列中两对对应点连线与不动直线的交点做透视中心。即使又选定了透视中心，也仍然有几种不同的方案来确定过渡直线。在图 8.29 中，我们也可以用 $A'C$ 做过渡直线。这些区别并不重要。重要的是：任何一对对应点的连线都可以扮演不动直线的角色。

8.12.5 非透视对应的射影对应

那么，这些相互对应的点的连线总体上有什么样的性质呢？图 8.31 展示了前例的情形。 很显然，成射影对应但非透视对应的不同底的两点列各对应点连

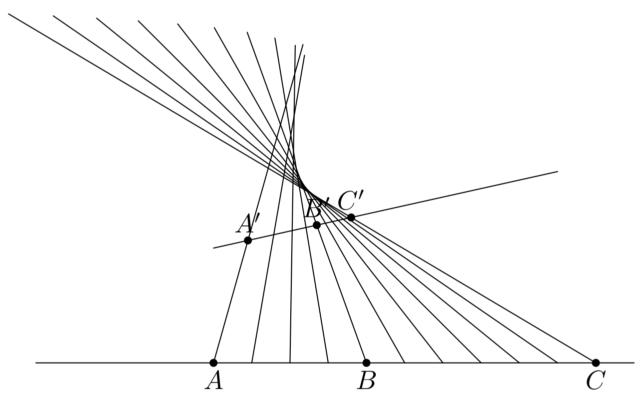

图 8.31: 成射影对应的两点列各对应点连线包络成二次曲线

线包络成二次曲线。

对偶地，成射影对应但非透视对应的不同底的两线束各对应直线的交点构成二次曲线。该命题的逆命题就是 Steiner 定理：以二次曲线上两相异点为两

个线束的中心，该二次曲线上流动点与两个中心分别连线做成两个线束，则两个线束按流动点做成的对应为射影对应，但非透视对应。

由此易知，在射影平面上，任意给定五个点，其中任何三点不共线，则有且只有一条二次曲线过给定五点。我们可以任选其中两点做线束中心，其他三点做流动点。由于三对对应点足以完全确定两个点列间的射影变换，对偶地，三对对应直线足以完全确定两个线束间的射影变换，所以，三个流动点与两个中心的连线足以完全确定两个线束间的射影变换。这样，两个线束的相互对应的直线的交点就构成了所求的二次曲线。

对偶地，在射影平面上，任意给定五条直线，其中任何三条直线不共点，则有且只有一条二次曲线与给定五直线相切。图 8.29、图 8.30、图 8.31 就是一个例子。给定两点列的底是两条直线，给定三对对应点的连线就是另外的三条直线，这五条直线足以确定两点列之间的射影变换，因而，流动对应点的连线包络出所求的二次曲线。

在二次曲线上任选五点，令其中一点 A 无限趋近于另一点 B，则以点 B 为线束中心，第三点 C 为另一个线束中心，线束 (B) 中的直线 AB 与线束 (C) 中的直线 AC 是三对对应直线之一，取极限时，AB 变成了二次曲线在 B 点的切线，AC 变成了 BC。但是，该切线和连线仍然是一对对应直线。因此，射影平面上，任意给定四点，其中任何三点不共线，再给定过其中一点的直线，其他三点都不在该直线上，那么，有且只有一条二次曲线过给定四点，且与给定直线在给定点上相切。

对偶地，射影平面上，任意给定四条直线，其中任何三线不共点，再给定其中一条直线上的一个点，其他三条直线都不过该点，那么，有且只有一条二次曲线与给定四直线相切，且通过给定的点。

更进一步，在上例中，令另外一点 D 无限趋近于点 C，则 CD 变成了二次曲线在 C 点的切线，DB 变成了 CB。两者仍然是一对对应直线。因此，射影平面上，任意给定不共线的三点，其中有两点，过每点给定一直线。这两直线都不过另外两点。那么，有且只有一条二次曲线过给定三点，且与给定的两直线在给定点上相切。

对偶地，射影平面上，任意给定不共点的三条直线，其中有两条直线，每条之上给定一点。这两点都不在另外两直线上。那么，有且只有一条二次曲线与给定三直线相切，且与那两条直线的切点为给定的点。

8.12.6 帕斯卡古典定理

讲到这些，我们就不能不提 Pascal 的古典定理：二次曲线内接六边形对边交点共线。这个只有十五个字的定理和 Pascal 本人一样，是个数学史上的传奇。如图 8.32 所示。 二次曲线内接六边形 $ABCDEF$ 的三对对边分别交于点 R, T, S。我们要证明这三点必定共线。取点 A 为透视中心，把二阶点列

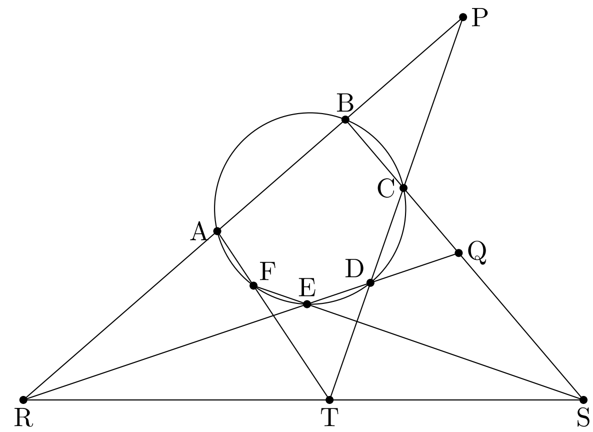

图 8.32: 二次曲线内接六边形对边交点共线

$(BCDF)$ 透视对应到一阶点列 $(PCDT)$。取点 E 为透视中心，把同样的二阶点列透视对应到一阶点列 $(BCQS)$。既然是从同一个二阶点列透视而来的，这两个一阶点列必定成射影对应。而点 C 又是不动点，因此该射影对应是透视对应。那么，两对对应点的连线 PB 和 DQ 的交点 R 是透视中心。另一对对应点 T 和 S 的连线必定过该透视中心。证毕。

二次曲线内接六边形对边交点所共之直线称为 Pascal 线。一个多边形，如果对任何一条边来说，其他顶点都处于该边所在直线的同侧，那么，就称这个多边形是凸的。Pascal 定理并不要求二次曲线的内接六边形是凸的。实际上，从一条二次曲线上任取 6 个不同点，可做 60 种不同的内接六边形，进而确定 60 条 Pascal 线。这些线每三条共一点。这样就有 20 个点。称为 Steiner 点。

Pascal 定理的对偶是 Brianchon 定理：二次曲线外切六边形对顶点连线共点。

如果二次曲线退化为两条相交直线，那么 Pascal 定理就转化为 Pappus 定理：两直线上各任选三点 A, B, C 和 A', B', C'，则 AB' 和 $A'B$ 的交点， AC' 和 $A'C$ 的交点， $B'C$ 和 BC' 的交点，三点共线。

8.12.7 “有限无界”与“无限有界”

如图 8.33 所示。 对单叶椭圆几何的距离有 $\phi = \arctan\lambda$。λ 可以取从 $-\infty$ 到 ∞，但是相应的 ϕ 却只能从 $-\dfrac{\pi}{2}$ 到 $\dfrac{\pi}{2}$。因此称该空间是有限无界的。‘有限’指椭圆几何的直线的总长度是有限的，恒为 π； ‘无界’指椭圆几何的直线是闭合的，没有边界点。

对双曲几何的距离，有 $\zeta = \operatorname{arctanh}\lambda$。$\lambda$ 取值范围是 $(-1, 1)$，而相应的 ζ 的取值范围是 $(-\infty, \infty)$。所以，称该空间是无限有界的。‘无限’指双曲几何

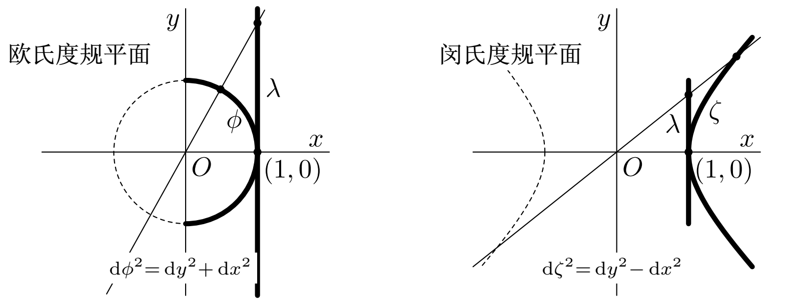

图 8.33: 双曲距离、椭圆距离与抛物距离的关系

的直线的长度是无限的；‘有界’指双曲几何的直线是开的，有两个不可逾越的边界点，对应图上的 $\lambda = \pm 1$。

但是，要注意，说椭圆几何中的直线‘长度有限等于 π，两端闭合无有界’，而双曲几何中的直线‘长度无限，两端有界’，其中前者的‘长度’指椭圆几何长度，后者的‘长度’指双曲几何长度。两者同词不同义，不可混淆。

8.12.8 欧氏正交与闵氏正交

不同的度规还导致不同的正交。图 8.34 显示了欧氏正交与闵氏正交的区

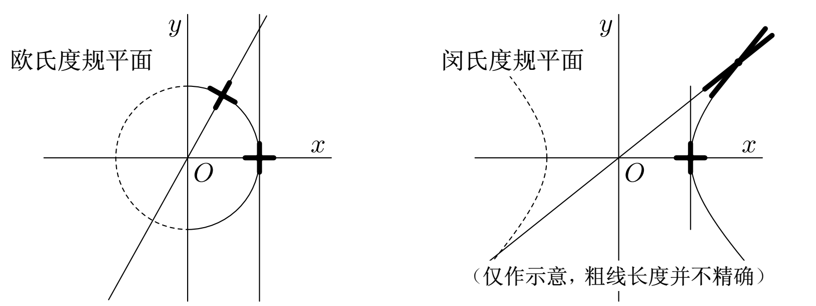

图 8.34: 欧氏正交与闵氏正交

别。在欧氏度规平面上，绕原点的旋转变换是保长的，保‘欧氏长度’的，因此也是保角的，保正交的。在闵氏度规平面上，绕原点的双曲旋转变换，物理上即洛伦兹变换，也是保长的，但这个保长，是保‘闵氏长度’的。因此，也是保角的，保‘双曲角度’的，保正交的，保‘闵氏正交’的。

为了从解析的角度分析这个问题，我们让圆和双曲线的半径分别取为变量 ρ 和 τ，即：在欧氏度规平面上，有 $\rho^2 = x^2 + y^2$，在闵氏度规平面上，有 $\tau^2 = x^2 - y^2$。当一点在旋转变换群的作用下沿着圆或者双曲线移动时，这两个

二次型的值应该不变。所以，在欧氏度规平面上，

$$\mathrm{d}(\rho^2) = \frac{\partial(\rho^2)}{\partial x}\,\mathrm{d}x + \frac{\partial(\rho^2)}{\partial y}\,\mathrm{d}y = 2x\,\mathrm{d}x + 2y\,\mathrm{d}y = 0$$

即：

$$x\,\mathrm{d}x + y\,\mathrm{d}y = 0$$

这就是矢径 (x, y) 与切线 (dx, dy) 正交的解析表达式。在闵氏度规平面上，

$$\mathrm{d}(\tau^2) = \frac{\partial(\tau^2)}{\partial x}\,\mathrm{d}x + \frac{\partial(\tau^2)}{\partial y}\,\mathrm{d}y = 2x\,\mathrm{d}x - 2y\,\mathrm{d}y = 0$$

即：

$$x\,\mathrm{d}x - y\,\mathrm{d}y = 0$$

这是矢径 (x, y) 与切线 (dx, dy) 闵氏正交的解析表达式。此式也可以写成 $x\,\mathrm{d}x + y(-\,\mathrm{d}y) = 0$ 或者 $x(-\,\mathrm{d}x) + y\,\mathrm{d}y = 0$。也就是说，如果我们把双曲线的切线 (dx, dy) 以过切点的水平线为轴做反射，变成 $(dx, -dy)$，或者以竖直线为轴做反射，变成 $(-dx, dy) = -(dx, -dy)$，那么，反射后的切线与矢径就是欧氏正交的了。

当 ζ 趋于正、负无穷大的时候，矢径与切线都会合二为一，从而形成两条渐近线。每条渐近线都是两条相互闵氏正交的直线合并而来的，因此，是自己与自己闵氏正交的。我们称这两条渐近线是自正交的，或者迷向的。

8.12.9 仿射变换、绝对直线与圆点

在射影几何中，保持同一直线整体不变的所有射影变换构成一个群，称为仿射变换群。以该群为主变换群的几何称为仿射几何。该直线称为仿射几何的绝对直线。在双曲旋转变换下，绝对直线上其他点都会被变换到新的位置，只有两个点是不动的。一般把无穷远直线选为仿射几何的绝对直线，这样单位双曲线与无穷远直线的两个交点就是双曲旋转变换的不动点。单位双曲线的中心到这两个点的连线就是渐近线，即迷向直线。

在欧氏度规平面上，过单位圆的圆心也有两条迷向直线。不过，这两条迷向直线是‘虚直线’。单位圆与无穷远直线有两个虚的交点。实际上，所有圆与无穷远直线都交于这两个虚点。因此，它们被称为圆点。这两个圆点是普通旋转变换下的不动点。圆心到这两个圆点所连的两条虚直线都是迷向的。

8.12.10 双曲运动群

实际上，保持射影直线上给定两点为不动点的所有射影变换构成一个群，称为双曲运动群。直线上的双曲距离是双曲运动群作用下的不变量。

如图 8.35 所示。 我们考察保持图中水平射影直线上点 P 和点 Q 不动的双

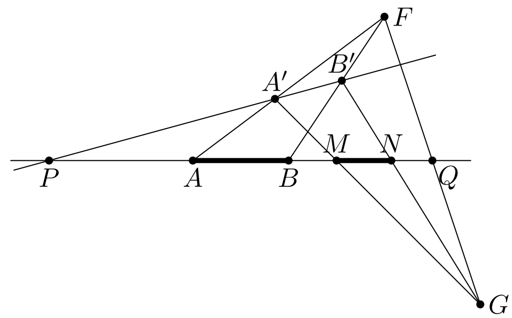

图 8.35: 线段 AB 双曲运动到 MN

曲运动群。我们已经知道，三对对应点完全确定射影直线上的一个射影变换。现在，设一个双曲运动变换把点 A 变换到点 M，或者简单的说，点 A 运动到点 M，这提供了一对对应点。点 P 和点 Q 都对应到自身，这提供了两对对应点。由此，该射影变换完全确定。那么，直线上任何一点 B 运动到何处就是完全确定的。

具体作图方法如下。过点 P 或点 Q 任做一直线。图中选了过点 Q 做直线。在其上任选两点 F 和 G。记 FA 和 GM 的交点为 A'。记 PA' 与 FB 的交点为 B'。则 GB' 与 PQ 的交点 N 就是点 B 在该双曲运动变换下的对应点。换句话说，线段 AB 做了一个双曲运动，变成了线段 MN，而其‘长度’，并不因‘运动’而改变。就象同一把尺子，不会因为摆放在不同的地方就改变‘长度’一样。

射影直线上的双曲运动与具体作图方法无关。图 8.36 显示了另外一种作图

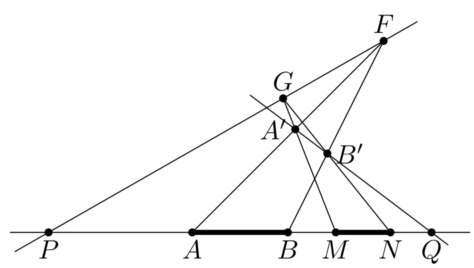

图 8.36: 双曲运动与具体作图方法无关

方法。可以看出，最终确定的点 N 的位置与上一种方法得到的结果完全一致。

如果取 AB 的长度为标准长度，那么，做一个双曲运动使点 A 移动到点 B，并改记为 A_1，而点 B 则移动到点 B_1，则有 AB_1 的长度为两个标准长度。再做一次同样的双曲运动，点 A_1 移动到点 B_1，而点 B_1 则移动到点 B_2。那么，AB_2 的长度为三个标准长度。如此，我们便可以在 PQ 之间建立数轴。取

标准长度足够小，就可以建立从负无穷大到正无穷大的整个实数数轴。

双曲距离与双曲角度的对应关系还有另一种理解。如图 8.37 所示。 当所

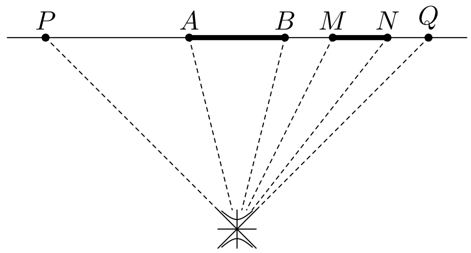

图 8.37: 双曲距离与双曲角度的关系

考察的射影直线被理解为无穷远直线的时候，相当于图中的‘单位双曲线’的‘半径’变得无限小，小到整个双曲线退化成两条渐近线。但是，按图中的透视关系，透视线束却不受这一过程的影响。因此，无穷远直线上两点的双曲距离，实际上是用以坐标原点为中心的线束中相应两直线之间的双曲角度来刻画的。

8.12.11 椭圆运动群

同理，如图 8.38 所示， 对于椭圆距离，当所考察的射影直线被理解为无

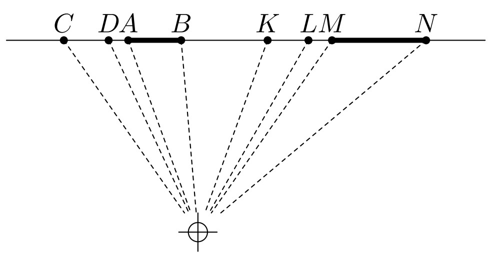

图 8.38: 椭圆距离与欧氏角度的关系

穷远直线的时候，相当于图中的‘单位圆’的‘半径’变得无限小，小到整个圆退化成一个点。（实际上也是退化成两条渐近线，只不过这两条渐近线是虚直线。）然而，按图中的透视关系，透视线束不受这个取极限过程的影响。因此，无穷远直线上两点的椭圆距离，实际上是用以坐标原点为中心的线束中相应两直线之间的欧氏角度来刻画的。显然，从左边无限远处到右边无限远处，相应两直线之间的欧氏角度是 π，因此我们说，这个无穷远直线的长度是 π。实际上，在椭圆几何中，所有直线的长度都是 π。

椭圆运动就是绕给定点的普通旋转。以该给定点为中心的线束中，任何两直线的欧氏角度是椭圆运动群的不变量。

椭圆运动没有不动点。为了完全确定一个椭圆运动，我们必须给出三对对应点，比如图 8.38 中的 $(ACD) \to (MKL)$。

具体方法是这样的：平面上所有满足 $\angle COD = \angle KOL$ 的点 O，构成一条四次曲线。所有满足 $\angle DO'A = \angle LO'M$ 的点 O'，构成另一条四次曲线。两者若有交点，则它必可作为图 8.38 中的透视中心。从而，对其他任何点，比如点 B，做 $\angle NOM = \angle BOA$，那么点 N 就是该椭圆运动变换下点 B 的对应点。

8.12.12 平面透视变换与直移

这两例中让单位双曲线或单位圆半径无限小的取极限的过程，其实是射影平面上的一类射影变换，称为透视变换，也称同调变换。同调变换保持给定直线上所有点不动，保持该直线外一给定点不动。该给定直线称为同调变换的轴，也称透视轴，该给定点称为同调变换的心，也称透视中心。

注意，我们在讲射影直线时，也提到过两个点列之间的透视对应。这种透视对应可以用‘选定射影平面上的两点列外一点为透视中心，对应两点的连线过透视中心’这一条件来确立。对偶的情况是两线束之间的透视对应。即‘选定射影平面上的两线束外一直线为透视轴，对应两直线的交点在透视轴上’所确立的对应。这种意义上的‘透视中心’和‘透视轴’与同调的心和轴，名称一样，含义上也有关联，但毕竟是用在不同的上下文当中，刻画不同事物的，还是要注意它们的区别。

由于射影变换保持点与直线的接合关系不变，因此，轴上任何一点与心的连线都是整体不变的。从而该连线上的点在同调变换下实际上是做了一个同底的一维射影变换。因为轴上的点和心是该一维射影变换的两个不动点，所以，该一维射影变换是双曲运动变换。

同轴且同心的所有同调变换构成群，称为同调变换群。它只有一个参数。实际上，它与轴上任何一点与心的连线上，以轴上该点和心为不动点的双曲运动群同构。

当一个‘同调’的心在其轴上的时候，按定义就不再是同调，所以‘改’称为扩张，也称直移。扩张的心称为直移中心，扩张的轴称为直移轴。

8.12.13 哈勃定律

有个宇宙学原理说：宇宙是时间上大尺度稳定，空间上大尺度均匀的。可是，天文观测上又有一个哈勃定律说：宇宙中的星系都在以特定的速度远离我们，其速度与该星系离我们的距离大体成正比。这两者有没有办法统一起来呢？

回答是肯定的。图 8.39 显示了我们宇宙的一种模型。图中点 M 是线段 PQ 的中点。线段 MS 垂直于 PQ 且其欧氏长度等于 MQ。点 O 是线段 MS

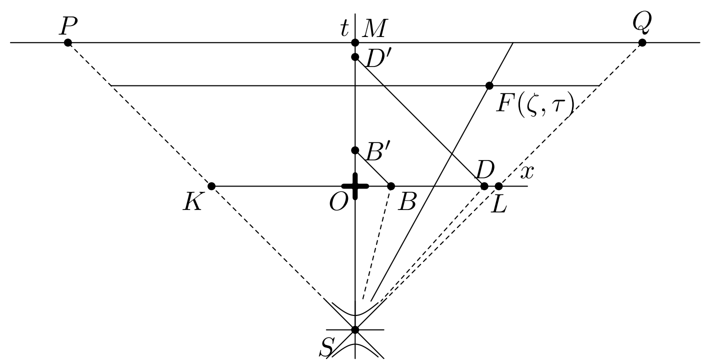

图 8.39: 直角坐标 (x, y) 和同调坐标 (ζ, τ)

的中点。我们以点 O 为原点建立直角坐标系。横轴为空间轴 x，纵轴为时间轴 t。

现在，不同寻常的是：我们取保持点 P 和点 Q 不动的双曲运动群为直线 PQ 上的空间移动群，设其参数为 ζ；取保持点 M 和点 S 不动的双曲运动群为时间移动群，设其参数为 τ。这样就唯一确定了平面上的一个同调，以直线 PQ 为轴，以点 S 为心。指定线束 (S) 中同一条直线上的点的 ζ 值都相等，同一条水平直线上的点的 τ 值都相等。两个参数 (ζ, τ) 本来张成一个完全的欧氏平面 $\mathbf{R}^2$，纵横两个方向上的取值范围都是从负无穷大到正无穷大，如今被映射到等腰直角三角形 $\triangle PSQ$ 之中。

这样，三角形 $\triangle PSQ$ 内部任意一点 F 就有了两套坐标：直角坐标 (x, t)，和同调坐标 (ζ, τ)。设图 8.39 中线段 OL 的欧氏长度为 R，则显见有坐标变换关系：

$$\frac{x}{R} = (1 + \tanh\frac{c\tau}{R})\tanh\frac{\zeta}{R}, \quad \frac{ct}{R} = \tanh\frac{c\tau}{R}$$

在点 O 的邻域内，横轴上，$\tau = 0$，有 $\mathrm{d}x = \mathrm{d}\zeta$；纵轴上，有 $\mathrm{d}t = \mathrm{d}\tau$。可见，该邻域内两套坐标是一致的。

在横轴上其他地方，因为 $\tau = 0$，所以有 $\frac{x}{R} = \tanh\frac{\zeta}{R}$。由洛伦兹变换，我们知道双曲旋转的双曲角度 ζ 是与惯性系的相对运动速度 v 相联系的。两者关系为 $\frac{v}{c} = \tanh\frac{\zeta}{R}$。带入前式，我们有 $v = \frac{c}{R}x$。这便是哈勃定律。哈勃常数 $H = \frac{v}{x} = \frac{c}{R}$ 是个定值。在图 8.39 中，点 D 处的星系的退行速度将数倍于点 B 处的星系，

有人说，哈勃常数只在一定范围内是个定值，在‘宇宙大爆炸初期’，很可能与现在的不同。其实，哈勃常数不是定值的原因在于天文学家使用的距离是 ζ 而不是 x，速度是 $\nu = \frac{\mathrm{d}\zeta}{\mathrm{d}\tau}$ 而不是 $v = \frac{\mathrm{d}x}{\mathrm{d}t}$。这样定义的哈勃常数 $H' = \frac{\nu}{\zeta} = H\frac{\sinh(2\zeta/R)}{2\zeta/R}$。当 ζ 很小时，$H' \to H$，可以看做是个定值。当 ζ 很大时，比如所谓的‘宇宙大爆炸初期’，H' 可以远大于 H。

推导过程是这样的：因为对于横轴上的点，$\tau = 0$，所以 $\dfrac{x}{R} = \tanh\dfrac{\zeta}{R}$。那么 $\mathrm{d}x = (1-(x/R)^2)\,\mathrm{d}\zeta$。又 $\mathrm{d}t = \mathrm{d}\tau$，所以 $v = \dfrac{\mathrm{d}x}{\mathrm{d}t} = (1-(x/R)^2)\dfrac{\mathrm{d}\zeta}{\mathrm{d}\tau} = (1-(x/R)^2)\nu$。代入 $v = Hx$，得 $H' = \dfrac{\nu}{\zeta} = H\dfrac{x}{(1-(x/R)^2)\zeta}$。把 $\dfrac{x}{R} = \tanh\dfrac{\zeta}{R}$ 代入即可得证。

参数 R 常被称为宇宙半径，$\dfrac{R}{c}$ 常被称为宇宙年龄。这些名称都是很不恰当的。只要做个简单的变量代换 $x' = x/R, t' = t/R, \zeta' = \zeta/R, \tau' = \tau/R$，我们就可以得到一个‘单位宇宙’，即：

$$x' = (1+\tanh c\tau')\tanh\zeta', \quad ct' = \tanh c\tau'$$

该式根本不含 R。这种标准化的‘单位宇宙’的性质与 R 完全无关。

R 的实际取值，很可能会随着不同地点邻域内物质和能量分布的不同而不同。因此，R 不是‘宇宙’的半径，而是处于该地的观察者用给定的手段可以加以考察的范围的半径。如果观察者移动了地点，或者更换了手段，那么这个范围当然是可以改变和突破的。特别是，我们只要做匀速直线运动，一直进行下去，就总可以走出‘宇宙’之外。

由于宇宙半径、宇宙年龄这些词用的实在太多了，出于习惯，我们也用它们表示模型中用 (x,t) 坐标表示的 OL 的空间长度和 OS 的时间间隔。但是，这样做的时候，我们心里必须清楚这些词的确切含义。

8.12.14　光速的变与不变

狭义相对论对直角坐标系 (x,t) 成立。换算到同调坐标 (ζ,τ) 一定会有一些‘新奇’的结果。比如，光的世界线 $x = \pm ct$ 映射到 (ζ,τ) 坐标系中是个很复杂的曲线。这意味着光速不再是固定的三十万公里每秒。比如在图 8.39 中，点 D 位置有个星系，如果我们使用 ζ 和 τ 来描述那个星系周围的物质运动情况，那么我们需要采用的光速的值将远大于三十万公里每秒。

定量的，由于在横轴上 $\mathrm{d}x = (1-(x/R)^2)\,\mathrm{d}\zeta$，即 $\mathrm{d}\zeta = \cosh^2(\zeta/R)\,\mathrm{d}x$。另有 $\mathrm{d}\tau = \mathrm{d}t$，代入 $\mathrm{d}x = \pm c\,\mathrm{d}t$，得 $\mathrm{d}\zeta = \pm\cosh^2(\zeta/R)\,c\,\mathrm{d}\tau$。可见光速被‘放大’了 $\cosh^2(\zeta/R)$ 倍。当 ζ 足够大时，光速可以数倍、数百倍于三十万公里每秒。

纵轴上，用 τ 刻画时间，那么，由于是在纵轴附近，ζ 取值微小，所以有 $\mathrm{d}x = (1+\tanh\dfrac{c\tau}{R})\,\mathrm{d}\zeta$。另有 $\mathrm{d}\tau = \cosh^2\dfrac{c\tau}{R}\,\mathrm{d}t$。代入 $\mathrm{d}x = \pm c\,\mathrm{d}t$，得

$$\mathrm{d}\zeta = \pm\frac{1}{(1+\tanh\frac{c\tau}{R})\cosh^2\frac{c\tau}{R}}c\,\mathrm{d}\tau = \pm\frac{2}{\exp\frac{2c\tau}{R}+1}c\,\mathrm{d}\tau$$

在无限久远的过去，光速是现在的二倍；在无限遥远的未来，光速将趋于零。这不是说物理的光不再行进了，而是说我们采用的数学方法在那种情形下不适合使用了。改换到 (x,t) 坐标，这些‘奇怪’的问题就自然化解了。

为何这么多年来都没有人提光速变化的事儿呢？那是因为‘全程平均光速’即便是在 (ζ,τ) 坐标系中，也仍然是三十万公里每秒。所谓‘全程’，是指从横轴某点出发直到抵达纵轴的光的世界线。比如图 8.39 中的 BB' 和 DD'。这种情况下，$\dfrac{x_D}{R}=\tanh\dfrac{\zeta_D}{R}$, $\quad \dfrac{ct_{D'}}{R}=\tanh\dfrac{c\tau_{D'}}{R}$，因为 $x_D=ct_{D'}$，所以 $\tanh\dfrac{\zeta_D}{R}=\tanh\dfrac{c\tau_{D'}}{R}$，进而 $\zeta_D=c\tau_{D'}$，即全程平均光速 $\zeta_D/\tau_{D'}$ 仍旧等于 c。天文上实际使用的光速，基本上都是这种意义的光速。

8.12.15 暗物质与暗能量问题

与‘瞬时光速’或者‘局域光速’本质上密切相关的，是暗物质与暗能量问题。用 (x,t) 坐标，在图 8.39 中，设点 B 处有一星系，直径为 $\mathrm{d}x$；在点 D 处有一物质密度相差不多的星系，直径也为 $\mathrm{d}x$。它们的结团状况类似是很自然的事情。但是，换算到 (ζ,τ) 坐标系中，点 D 处星系的直径 $\mathrm{d}\zeta_D=\cosh^2(\zeta_D/R)\,\mathrm{d}x$，点 B 处星系的直径 $\mathrm{d}\zeta_B=\cosh^2(\zeta_B/R)\,\mathrm{d}x$，如果 $\zeta_D\gg\zeta_B$，那么 $\mathrm{d}\zeta_D$ 也将远大于 $\mathrm{d}\zeta_B$。因而点 D 处星系的平均物质密度将远小于点 B 处的星系。这就产生了远处一个物质密度很低的星系，表现出的结团状况却与近处一个物质密度高得多的星系相类似的现象。

8.13 配极与对合

8.13.1 射影平面上的自配极三点形

如图 8.40 所示。 这是个球极投影的剖面图。在三维空间中，以点 S 为

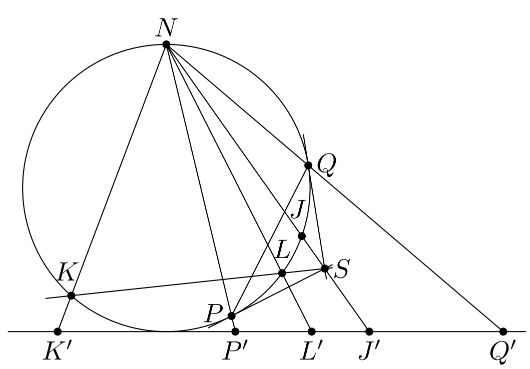

图 8.40: 球极投影与反演变换

‘线把’中心。点 S 在底面上的投影称为反演中心。‘线把’中的直线如果与球面有两个交点，那么这两个点在底面上的投影互为反演。从其中一个到另一个的变换称为反演变换。基准圆是自反演点的集合，其上每个点都是反演变换

的不动点。基准圆的圆心与复平面无穷远点互为反演。图 8.40 上，点 K' 和点 L' 互为反演。点 P' 和点 Q' 都为自反演点。

图 8.41 是同一个剖面图。 但是，这一次我们以三维空间中，点 T 为‘线

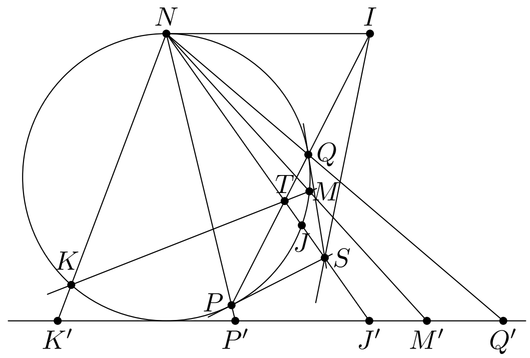

图 8.41: 球极投影与对径变换

把’中心。点 T 在底面上的投影称为对径中心。‘线把’中的任何直线都与球面有两个交点。这两个点在底面上的投影互为对径。从其中一个到另一个的变换称为对径变换。垂直平分线过对径中心的所有对径点对构成基准圆。图 8.41 上点 K' 与点 M' 互为对径点。

同一个基准圆的反演变换与对径变换的合成，称为中心对称变换。这里的中心，指基准圆圆心，既是反演中心，又是对径中心。图 8.40 和图 8.41 是同一个基准圆。点 L' 先反演到点 K'，再对径到点 M'，所以点 L' 与点 M' 中心对称。

我们在前面讲过射影空间中给定二次型的配极。在图 8.41 中，直线 TI 是点 S 的配极，直线 TS 是点 I 的配极，直线 IS 是点 T 的配极，称这样的三点形 ITS 为自配极三点形。

8.13.2 二次曲线的中心与直径

图 8.41 和图 8.26 本质上是一样的。图 8.26 中的 l_∞ 就是图 8.41 中的 PQ。前者的点 C 就是后者的点 S。它们在各自的图上都是相互共轭的极点和极线。

图 8.26 与图 8.27 的区别是无穷远直线与对分轴交换了角色，图 8.40 与图 8.41 的区别是以点 S 还是点 T 作为二次曲线的‘中心’。这个‘中心’是加引号的，因为这是个仿射几何的概念。只有给定了绝对直线，即‘无穷远直线’，才能定义给定二次曲线的‘中心’，即：绝对直线关于该二次曲线的配极点，称为该二次曲线的中心。

过二次曲线中心的直线称为该二次曲线的直径。因此，反演其实也是‘对径’，是双曲型对径。与椭圆型对径不同的是，双曲线的直径与双曲线可以有

两个实交点，一个二重实交点，两个虚交点，而椭圆的直径与椭圆恒有两个实交点。

8.13.3 反演与对径统称对合

反演与对径可以重述为：双曲线任何直径的两端点在南极投影下互为反演，椭圆任何直径的两端点在南极投影下互为对径。

如图 8.42 所示。 图中 RQ 是单位双曲线的一条直径。点 P 是点 Q 的投

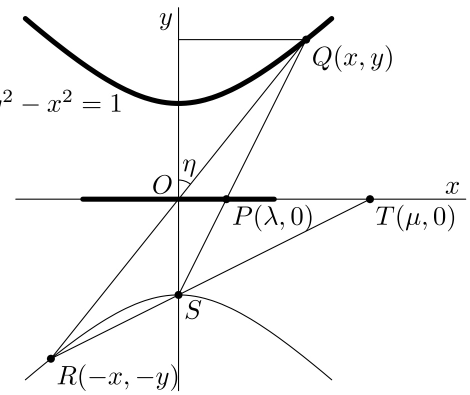

图 8.42: 双曲线任何直径的两端点在南极投影下互为反演

影，点 T 是点 R 的投影。由三角形相似立得 $\dfrac{1+y}{1}=\dfrac{x}{\lambda}$ 以及 $\dfrac{\mu}{1}=\dfrac{\mu+x}{y}$，即

$$\mu=\frac{x}{y-1},\quad \lambda=\frac{x}{y+1}$$

两式相乘得 $\mu\lambda=\dfrac{x^2}{y^2-1}=1$，这清楚地表明点 P 和点 T 互为反演。

注意到，$\dfrac{\lambda}{1}=\dfrac{1}{\mu}$ 意味着当 SP 从竖直方向顺时针偏转一定角度的时候，ST 从水平方向逆时针偏转了相同的角度，因此 SP 与 ST 双曲正交。

如图 8.43 所示。 图中 RQ 是单位圆的一条直径。点 P 是点 Q 的投影，点 T 是点 R 的投影。由三角形相似立得

$$\mu=\frac{-x}{1-y},\quad \lambda=\frac{x}{y+1}$$

两式相乘得 $\mu\lambda=\dfrac{-x^2}{1-y^2}=-1$，这意味着点 P 和点 T 互为对径。

注意到，$\dfrac{\lambda}{1}=\dfrac{1}{-\mu}$ 意味着当 SP 从竖直方向顺时针偏转一定角度的时候，ST 从水平方向也顺时针偏转了相同的角度，因此 SP 与 ST 是普通的正交。

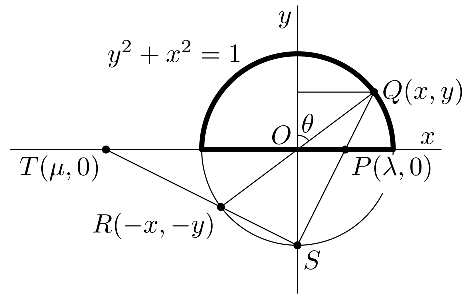

图 8.43: 椭圆任何直径的两端点在南极投影下互为对径

射影几何中，直线是闭合的。在图 8.42 中，从点 O 一直向左可以到达无穷远点，而不经过点 P 和点 T。这称为点 O 和无穷远点‘不分离’点 P 和点 T。实际上，任何一对互为反演的点都不分离另外一对互为反演的点。但是，在图 8.43 中，从点 O 要想到达无穷远点，要么经过点 P，要么经过点 T。这称为点 O 和无穷远点‘分离’点 P 和点 T。实际上，任何一对互为对径的点都分离另外一对互为对径的点。

图 8.40 中的直线 PQ 对应于图 8.42 中的无穷远直线。图 8.41 中的直线 IS 对应于图 8.43 中的无穷远直线。反演和对径统称为对合。反演称为双曲型对合，对径称为椭圆型对合。在中心投影下，任何直径的两个端点被投影为一点。因为这一点是相互对合的两点合成的，所以称为对合合点。图 8.42 中的无穷远直线被分成三类区域，一个实合点区，即类时区域；两个二重合点区，即类光区域；一个虚合点区，即类空区域。图 8.43 中的无穷远直线整个都是实合点区。

8.13.4 两个有趣的自对偶图形

如图 8.44 所示。 这是一个自对偶的图形，有两种对偶的解释。一个是：

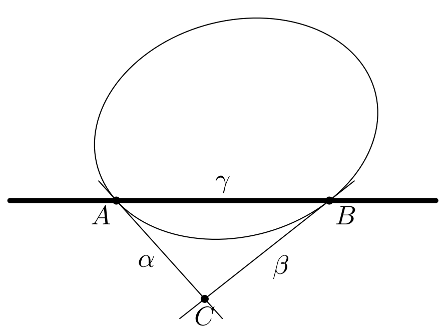

图 8.44: 互为配极的一个二自共轭直线点 C 和一条二自共轭点直线 γ

二次曲线由切点组成，其上两个切点有一连线，两个切点处的切线有一交点。另一个是：二阶曲线由切线包络而成，其上两个切线有一交点，两个切线的切点有一连线。前面已经介绍过，该连线和该交点互为配极。该连线称为极线，该交点称为极点。不难看出，二次曲线与二阶曲线对偶，切点与切线对偶，极点与极线对偶。

如果在图 8.44 上再加一套这样的点和线的组合，就可以构成另一种类型的极点和极线。如图 8.45 所示。两个二自共轭点直线的交点，与它们的极点的

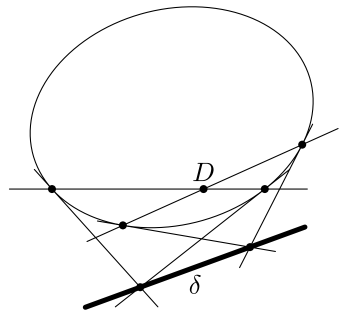

图 8.45: 互为配极的一个无自共轭直线点 D 和一条无自共轭点直线 δ

连线，互为配极。这个新的极线与二次曲线没有实交点，或者说，有两个虚交点。显然，这个图也是自对偶的。

配极有个奇妙的性质：过极点的任何直线都与极线‘正交’。这表明极点在某种意义上也是‘奇点’。比如，椭圆几何的模型之一是对径粘合的球面的几何。球面几何的‘直线’就是大圆。那么，南极和北极粘合在一起，与赤道互为配极。南北极是极点，赤道是极线。过南北极的任何大圆都与赤道正交。也只有南北极才有这个性质。所以说，南北极是球面上的‘奇点’。

注意，我们给‘正交’加了引号，因为不同的几何对直线和正交有不同的定义。复平面上的圆几何中，‘直线’是测地圆，‘正交’是指两测地圆在交点处的切线相互垂直；仿射几何中，‘直线’是仿射直线，‘正交’是指两仿射直线适合添加的无穷远点是绝对对合的一对对应点。当然，在适当的解释下，或者说‘翻译’下，不同几何中的‘正交’实际上是一致的。

8.13.5 对射变换与配极共轭

射影平面上可做两类一般线性变换，一类是从点到点的，称为直射变换；另一类是从点到直线的，称为对射变换，也称逆射变换。由于一般线性变换保持点与直线的接合关系不变，因此直射把点列的底变换成点列的底，即把直线变换成直线；对射则把点列的底变换成线束的底，即把直线变换成点。

如果用矩阵表示点或直线的齐次坐标之间的线性变换，那么直射的从点到

点（或者对射的从点到直线）的变换可表示为一个矩阵，同一直射的从直线到直线（或者同一对射的从直线到点）的变换可表示为其代数余子式矩阵的转置。

如果一个变换 P 满足 $P^2 = 1$，则称之为对合。直射中的对合称为调和同调，或者调和透视。对射中的对合称为配极。配极中相对应的点和直线称为相配的极点和极线。由上一段易知，一个对射是配极的充要条件是：该对射的表示矩阵为对称矩阵。

配极把极线上任何一点变换成一条过极点的直线，而这点和这直线又是相配的极点和极线，因此，射影平面上任何两点，如果一点在另一点的极线上，那么另一点也必定在这一点的极线上。称这样的两个点是配极共轭的，简称共轭的。

设共轭两点的坐标分别为 x^i 和 y^j，所论配极 a_{ij} 将 x^i 变换为极线 $\xi_j = a_{ij}x^i$，其中 $a_{ij} = a_{ji}$，如上一段所述，点 y^j 必在该直线上，因此 $\xi_j y^j = a_{ij}x^i y^j = 0$。这相当于 x^i 和 y^j 的双线性型。反之，如果两个点的坐标满足该方程，那么两点必配极共轭。

8.13.6 自共轭点和自共轭直线

如果所论配极将一个点变换成过该点的直线，即极点在相配的极线上，那么称这个点为自共轭的。自共轭点满足约束条件 $a_{ij}x^i x^j = 0$。反之，坐标满足该方程的点，必是自共轭的。显见，所有自共轭点构成二次曲线。

对偶地，如果一直线过另一直线的极点，那么另一直线也必过这一直线的极点。称这样的两直线共轭。如果一直线的极点在该直线上，那么称该直线为自共轭的。所有自共轭直线包络成二阶曲线。

根据矩阵理论，实对称矩阵必可对角化。这意味着，对于任何配极，经适当选择坐标系后，总可表示为 a_{ii} 非零，其他元素都为零的矩阵。此时，所有自共轭点满足的方程可表示为 $b_1{x_1}^2 + b_2{x_2}^2 + b_3{x_3}^2 = 0$。

这样，射影平面上的配极就只有两类了，一类是三项同号的，称为椭圆型配极；另一类是有一项与另外两项不同号的，称为双曲型配极。双曲型配极有实的自共轭点和实的自共轭直线，构成实二次曲线和实二阶曲线；而椭圆型配极没有实的自共轭点，必须要容许某些坐标取复数，自共轭条件才会有解。在这种意义上，我们说椭圆型配极有虚的自共轭点和虚的自共轭直线，构成虚二次曲线和虚二阶曲线。

前面讲过，射影直线上的反演和对径统称对合。反演是双曲型对合，有两个不动点；对径是椭圆型对合，没有不动点。直线上的对合要么是反演，要么是对径，没有其他可能。

给定一条非自共轭直线，其上任意一点的极线与该直线有交点，从该点到该交点的射影变换满足 $P^2 = 1$，因此构成对合。如果该对合是反演，那么两个

不动点就是配极的自共轭点。这意味着该非自共轭直线与配极的二次曲线有两个交点。如果该对合是对径，那么该非自共轭直线上就没有自共轭点。这意味着该直线与配极的二次曲线没有交点。

对偶地，过任意一个非自共轭点，要么有配极的二次曲线的两条切线，要么没有切线。如果是前者，则称该非自共轭点为二自共轭直线点，也称二切线点；如果是后者，则称该点为无自共轭直线点，也称无切线点。所有二切线点构成一个区域，称为外域。所有无切线点也构成一个区域，称为内域。

给定一个双曲型配极，可把实射影平面划分为三个部分：自共轭点构成的二次曲线，内域，和外域。这种划分非常接近于相对论中对时空点的划分。内域对应类时区域，外域对应类空区域，二次曲线对应类光区域。如果取一条二自共轭点直线为无穷远直线，并且让双曲线的‘半径’无限小，直至退化为两条相交直线，那么这种对应就是精确的了。

二次曲线将内域完全包围。因此，过内域中任何一点的直线，其上必有两个自共轭点。如果存在一条过无切线点的直线，其上没有自共轭点，那么这个配极必定不是双曲型配极，因此必定是椭圆型配极。

8.13.7 非齐次坐标与仿射变换

在仿射几何中，选择绝对直线作为坐标三点形的第三线 $\rho[0,0,1]$，则该绝对直线的方程为 $x_3=0$。也就是说，不在绝对直线上的点，都有 $x_3\neq 0$。这样，我们就可以对不在绝对直线上的点舍弃齐次等价条件，统一选择 $x_3=1$ 的解析点作为几何点。此时 x_1,x_2 就不再是齐次坐标了，称为非齐次坐标。

用非齐次坐标，我们可以把仿射变换群的一般形式表述为：

- 对于不在绝对直线上的点 $(x_1,x_2,1)$，有

$$
\begin{aligned}
x_1' &= a_{11}x_1+a_{12}x_2+a_{13}\\
x_2' &= a_{21}x_1+a_{22}x_2+a_{23}\\
\Phi &= \begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}, \quad |\Phi|\neq 0
\end{aligned}
$$

- 对于在绝对直线上的点 $(x_1,x_2,0)$，有

$$
\begin{aligned}
x_1' &= a_{11}x_1+a_{12}x_2\\
x_2' &= a_{21}x_1+a_{22}x_2\\
\Phi &= \begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}, \quad |\Phi|\neq 0
\end{aligned}
$$

注意，对于后者，因为 $x_3=0$，所以 x_1,x_2 不能都为零，而且齐次等价条件仍可适用。这样一来，仿射几何中的绝对直线实际上还是可以作为射影几何中的

一条射影直线来处理，其上的点的非齐次坐标可以看作是射影直线上的齐次坐标，平面上的仿射变换对绝对直线的作用等价于绝对直线上的射影变换。因此，我们说，绝对直线本身不含仿射性质。仿射性质都是关于不在绝对直线上的点 $(x_1, x_2, 1)$ 的。这些点就构成了仿射平面。换句话说，从射影平面上除去一条绝对直线，就是仿射平面，记为 $\mathbf{A}^2$。这样，每条直线都被除去了一个点，即它与绝对直线的交点，称这样的直线为仿射直线。

虽然从概念上讲，仿射直线与绝对直线没有交点。但是，在许多实际的讨论中，说'仿射直线与绝对直线的交点'并没有什么坏处，它可以从射影几何中借来精确的定义，而且用它来讨论问题非常方便。所以，在不致混淆的情况下，我们将继续使用这种说法。

以仿射变换群为主变换群的几何称为仿射几何。仿射几何中，如果两直线与绝对直线三线共点，则称这两直线平行。由于所有射影变换都保持接合关系不变，因此仿射变换保持直线之间的平行关系不变。

8.13.8 相似变换、绝对对合与正交

如果上述仿射变换满足 $\Phi\Phi^T = \rho^2 I, \quad \rho \neq 0$，则称之为相似变换。称 $|\rho|$ 为相似比。相似变换把图形变成相似图形。由于 $|\Phi^T| = |\Phi|$，所以对相似变换，有 $|\Phi| = \pm|\rho|$。所有相似变换构成群。该群是仿射变换群的子群。相似比为 1 的相似变换称为正交变换。正交变换把图形变成全等图形。所有正交变换构成群。该群是相似变换群的子群。

可证相似变换群保持绝对直线上的一个椭圆型对合不变，即相似变换把该对合的一对对应点变成另一对对应点。设 $(x_1, x_2, 0)$ 和 $(y_1, y_2, 0)$ 是该对合的一对对应点，则必有 $x_1y_1 + x_2y_2 = 0$，简记为 $\delta_{ij}x^iy^j = 0$。反之，如果绝对直线上两个点的坐标满足该方程，则必是该对合的一对对应点。由 $\Phi\Phi^T = \rho^2 I, \quad \rho \neq 0$ 可证，如果设相似变换把 $(x_1, x_2, 0)$ 变到 $(x_1', x_2', 0)$，把 $(y_1, y_2, 0)$ 变到 $(y_1', y_2', 0)$，那么 $x_1'y_1' + x_2'y_2' = \rho^2(x_1y_1 + x_2y_2)$。因此，如果这两个点原先是对合对应的，那么变换后仍然是对合对应的。显见，该椭圆型对合是相似变换群的绝对形。故而称该对合为绝对对合。

如果两条直线与绝对直线的交点是绝对对合的一对对应点，则称这两条直线正交。相似变换保持直线之间的正交关系不变。

前面说过，给定双曲型配极，那么，过极点的任何直线都与极线'正交'。现在可以给出'正交'的确切含义了。如图 8.46 所示。选择一条过极点的非自共轭直线为绝对直线。我们已经知道，双曲型配极在任何非自共轭直线上都导出一个对合。显见极线与绝对直线的交点与极点是该对合的一对对应点。极线当然过该交点，而任何其他过极点的直线当然过极点，也就是说，两者分别过绝对对合的一对对应点，因此，所有过极点的直线都与极线正交。

图 8.46 中 AC 是所选择的绝对直线。给定配极在该直线上导出一个椭圆型

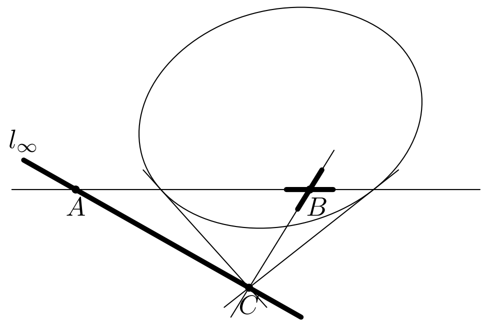

图 8.46: 绝对直线上的椭圆型对合确定直线之间的正交关系

对合，点 A 和点 C 恰为一对对应点，因此，直线 AB 与直线 BC 正交。

可是，这种意义上的‘正交’并没有要求对合必须是椭圆型的。如图 8.47 所示。 直线 AC 被选定为绝对直线。它是给定配极的二自共轭点直线，因此，

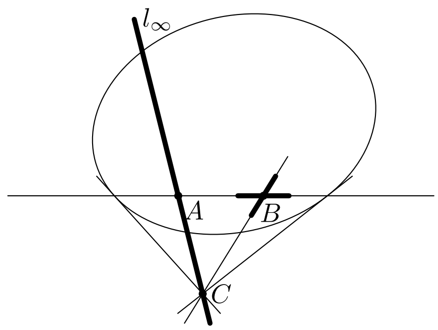

图 8.47: 绝对直线上的双曲型对合确定直线之间的正交关系

给定配极在该直线上导出一个双曲型对合。点 A 和点 C 是该对合的一对对应点。那么，我们也可以说，实际上，也应该说，直线 AB 与直线 BC 是正交的。可见，绝对直线上的双曲型对合同样可以用来确立直线之间的‘正交’关系。当然，这样的‘正交’也会有些不一样的特点，我们称之为闵氏正交。

如果前述仿射变换满足 $\Phi\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\Phi^T=\rho^2\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad \rho\neq 0$，则称之为闵氏相似变换。所有闵氏相似变换构成群。该群是仿射变换群的子群。相似比为 1 的闵氏相似变换称为闵氏正交变换，也称洛伦兹变换。所有闵氏正交变换构成群。该群是闵氏相似变换群的子群。

闵氏相似变换群保持绝对直线上的一个双曲型对合不变。设 $(x_1,x_2,0)$ 和 $(y_1,y_2,0)$ 是该对合的一对对应点，则必有 $x_1y_1-x_2y_2=0$，简记为 $\eta_{ij}x^iy^j=0$。η_{ij} 的矩阵表示为 $\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$。反之，如果绝对直线上两个点的坐标满足该方程，则必是该对合的一对对应点。该双曲型对合是闵氏相似变换群的绝对形。

图 8.34 显示了上述两种典型的正交。图 8.42 和图 8.43 中 SP 和 ST 正交也说明了正交与对合的关系。

8.13.9 度规与一般相似变换

相关概念可以进一步一般化。设 $G=\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$, $\quad g_{12}=g_{21}$, $\quad |G|\neq 0$，或者记为更一般的形式 $g_{ij}=g_{ji}$, $\quad |g_{ij}|\neq 0$。这个矩阵称为度规。如果前述仿射变换满足 $\Phi G\Phi^T=\rho^2 G$, $\quad \rho\neq 0$，则称之为给定度规 G 的一般相似变换。同一度规的所有一般相似变换构成群。该群是仿射变换群的子群。给定度规的相似比为 1 的一般相似变换称为该度规的一般正交变换。同一度规的所有一般正交变换构成群。

给定度规 G 的一般相似变换群保持绝对直线上的一个对合不变。设 $(x_1,x_2,0)$ 和 $(y_1,y_2,0)$ 是该对合的一对对应点，则必有 $g_{ij}x^iy^j=0$。反之，如果绝对直线上两个点的坐标满足该方程，则必是该对合的一对对应点。称该对合为绝对对合。称该双线性型是该一般相似变换群的绝对双线性型。如果两直线分别过绝对对合的一对对应点，则称这两直线正交。

8.13.10 仿射向量与保形变换

对于仿射平面 $\mathbf{A}^2$ 上的两个流动点 $(x_1,x_2,1)$ 和 $(y_1,y_2,1)$，称 $(y_1-x_1,y_2-x_2,\hat{0})$ 为一个从起点 x^i 到终点 y^i 的向量，记为 $\vec{a}$。前述仿射变换把一个向量的起点和终点变换到新的起点和终点，因而就把一个向量变换到一个新的向量 $\vec{a}\,'=\Phi\vec{a}$。

注意，虽然形式上向量的第三坐标为零，很像是绝对直线上的点，甚至在仿射变换下的变换关系也一样，但向量的前两个坐标分量都是非齐次坐标，因而向量的前两个分量是可以同时取零的，称向量 $(0,0,\hat{0})$ 为零向量。即使不同时取零，$(1.1,1.2,\hat{0})$ 和 $(3.3,3.6,\hat{0})$ 也被认为是两个不同的向量，也就是说，向量之间不存在齐次等价关系。而绝对直线上的点的坐标仍旧被认为是齐次坐标，存在着齐次等价关系。实际上，从所有向量中除去零向量，再加上齐次等价关系，以等价类为基本元素，确实可以构成射影空间。为了把向量和绝对直线上的点区分开，我们给向量的第三坐标的零加上标记‘ˆ’。

仿射平面上有一个特殊点 $(0,0,1)$，称为坐标原点。任何一点 $(x_1,x_2,1)$ 都可以与坐标原点定义一个向量 $(x_1,x_2,\hat{0})$，起点为坐标原点，终点为该点，称该向量为矢径。显然，给定坐标原点之后，矢径与点一一对应。所以，矢径与点常被混用，其确切含义需要根据上下文来做判断。

仿射平面上有两个特殊的向量 $\vec{e}_1=(1,0,\hat{0})$ 和 $\vec{e}_2=(0,1,\hat{0})$，任何向量 $\vec{a}=(a_1,a_2,\hat{0})=a_1\vec{e}_1+a_2\vec{e}_2$，简记为 $\vec{a}=a^i\vec{e}_i$。称这样的一组向量 $\vec{e}_i$ 为一个基底，或坐标基，也称坐标框架，简称标架。基底中的任何一个向量 $\vec{e}_i$ 都称为基矢。

设有任意两个向量 $\vec{a}=(a_1,a_2,\hat{0})$ 和 $\vec{b}=(b_1,b_2,\hat{0})$，称 $\vec{a}\cdot\vec{b}=g_{ij}a^ib^j$ 为这两个向量的内积。如果两向量的内积为零，则称这两向量正交，因为此时它们起点和终点的连线正交。一个向量 $\vec{a}$ 与自己的内积，称为该向量的模长平方，再开方，就得到该向量的模长，也称长度，记为 $|\vec{a}|$。

设给定度规的一般相似变换把向量 $\vec{a}$ 变换成 $\vec{a}\,'$，把向量 $\vec{b}$ 变换成 $\vec{b}\,'$，那么 $\vec{a}\,'\cdot\vec{b}\,'=\rho^2\vec{a}\cdot\vec{b}$，同时，$|\vec{a}\,'|=\rho|\vec{a}|$，　$|\vec{b}\,'|=\rho|\vec{b}|$，因此 $\dfrac{\vec{a}\cdot\vec{b}}{|\vec{a}|\,|\vec{b}|}$ 是一个不变量，可以作为对两向量起点和终点的连线所交角度的刻画。按照这个定义，一般相似变换保持两相交直线的交角不变，故而又称保角变换。另一方面，一般相似变换对任何向量的模长做同比例的缩放，因此，它把图形变成相似图形，故而又称保形变换。

由于一般正交变换的相似比为 1，所以一般正交变换保持任意两向量的内积不变，保持任意向量的模长不变。因此，一般正交变换又称为保长变换。

8.13.11　一般相似变换的绝对形

给定度规 g_{ij}，一般相似变换群保持绝对空间上一个绝对双线性型 $g_{ij}x^iy^j=0$ 不变。注意，绝对空间仍被视为射影空间。由于度规是对称的，因此该绝对双线性型实际上定义了射影空间中的一个配极。称满足绝对双线性型的点 x^i 和点 y^j 配极共轭，简称共轭。与给定点共轭的所有点构成与该点相配极的对偶元素。比如，在三维仿射空间 $\mathbf{A}^3$ 中，有一个选定的绝对平面。给定度规在绝对平面上导出一个配极。该配极中，与给定点共轭的所有点构成与该点相配极的极线，而该点则称为与该极线相配极的极点。

因此我们说，仿射空间 $\mathbf{A}^n$ 中，给定度规 g_{ij} 的一般相似变换群保持绝对空间中的配极 $g_{ij}x^iy^j=0$ 不变。如果仿射空间中两条直线与绝对空间的交点是该配极的一对共轭点，则称这两条直线正交。

该配极下的所有自共轭点构成绝对空间中的一个不变子流形 $g_{ij}x^ix^j=0$。在不同维度时，该不变子流形可以是二次孤点、二次曲线、二次曲面、二次超曲面等等。它又被称为一般相似变换群的绝对形。

8.14　仿射联络系数和黎曼曲率张量

8.14.1　仿射联络系数

前面讲过，短程线微分方程组是

$$g_{ij}\ddot{x}^j+\frac{\partial g_{in}}{\partial x^m}\dot{x}^m\dot{x}^n-\frac{1}{2}\frac{\partial g_{mn}}{\partial x^i}\dot{x}^m\dot{x}^n=0$$

后两项又可写成

$$\frac{1}{2}\left(\frac{\partial g_{in}}{\partial x^m}+\frac{\partial g_{im}}{\partial x^n}-\frac{\partial g_{mn}}{\partial x^i}\right)\dot{x}^m\dot{x}^n=\Gamma_{mni}\dot{x}^m\dot{x}^n$$

其中的 Γ_{mni} 就是仿射联络系数。这个记号称为第一类 Christoffel 记号。更常使用的是第二类 Christoffel 记号 $\Gamma^i_{mn} = g^{ik}\Gamma_{mnk}$。用这种符号可以把测地线微分方程组表示为

$$\ddot{x}^i + \Gamma^i_{mn}\dot{x}^m\dot{x}^n = 0$$

Γ^i_{mn} 参与运算时完全符合张量简记法，指标升降也可通过 g_{mn} 或 g^{mn} 来实现。但是，它在坐标系变换时并不符合张量的变换法则。因此，它并不是张量。

由于黎曼流形的度规都是对称的，因此仿射联络系数关于两个下标也是对称的，即 $\Gamma^i_{mn} = \Gamma^i_{nm}$。

来看一个实例，获得一些感性认识。前面讲过的平面共点圆几何，有度规 $g_{ij} = \dfrac{1}{(u^2+v^2)^2}\delta_{ij}$。记 $\rho = \dfrac{1}{(u^2+v^2)^2}$，则有

$$\frac{\partial\rho}{\partial u} = \frac{-4u}{(u^2+v^2)^3}, \qquad \frac{\partial\rho}{\partial v} = \frac{-4v}{(u^2+v^2)^3}$$

另外，$g^{ij} = (u^2+v^2)^2\delta^{ij}$，代入仿射联络系数的表达式

$$\Gamma^k_{mn} = \frac{1}{2}g^{ik}\left(\frac{\partial g_{im}}{\partial x^n} + \frac{\partial g_{in}}{\partial x^m} - \frac{\partial g_{mn}}{\partial x^i}\right)$$

可以得到仿射联络系数的各分量为

$$\Gamma^u_{uu} = \frac{-2u}{u^2+v^2}, \quad \Gamma^v_{uu} = \frac{2v}{u^2+v^2}, \quad \Gamma^u_{uv} = \Gamma^u_{vu} = \frac{-2v}{u^2+v^2}$$

$$\Gamma^v_{uv} = \Gamma^v_{vu} = \frac{-2u}{u^2+v^2}, \quad \Gamma^u_{vv} = \frac{2u}{u^2+v^2}, \quad \Gamma^v_{vv} = \frac{-2v}{u^2+v^2}$$

代入短程线微分方程组：

$$\begin{aligned} \ddot{u} + \Gamma^u_{uu}\dot{u}\dot{u} + 2\Gamma^u_{uv}\dot{u}\dot{v} + \Gamma^u_{vv}\dot{v}\dot{v} &= 0 \\ \ddot{v} + \Gamma^v_{uu}\dot{u}\dot{u} + 2\Gamma^v_{uv}\dot{u}\dot{v} + \Gamma^v_{vv}\dot{v}\dot{v} &= 0 \end{aligned}$$

就得到了前面讲过的短程线方程组。

8.14.2 黎曼曲率张量

黎曼曲率张量可以表示为

$$R^l_{kij} = \partial_i\Gamma^l_{kj} - \partial_j\Gamma^l_{ki} + \Gamma^p_{kj}\Gamma^l_{pi} - \Gamma^p_{ki}\Gamma^l_{pj}$$

既然是张量，就可以做角标下降，即 $R_{lkij} = g_{lm}R^m_{kij}$。

显见 R^l_{kij} 关于 i,j 反对称。还可证 R_{lkij} 不仅关于 i,j 反对称，而且关于 l,k 反对称，以及 $R_{lkij}=R_{ijlk}$。另外还有一组恒等式，称为第一 Bianchi 恒等式：

$$R_{lijk}+R_{lkij}+R_{ljki}=0$$

这里可以再引入一种简记法：用 $\overline{+ijk}$ 表示对其左边的表达式中的三个角标 i,j,k 取轮排求和。这样，上式可以表示为

$$R_{lijk}\overline{+ijk}=0$$

n 维黎曼流形的 R_{ijkl} 有 n^4 个分量，据说在上述条件的约束下，只有 $\frac{1}{12}n^2(n^2-1)$ 个分量是独立的。对于二维黎曼流形，只有一个分量是独立的。对于三维黎曼流形，有六个分量是独立的。对于四维黎曼流形，有二十个分量是独立的。

继续用平面共点圆几何作例子。因为 R_{lkij} 关于 l,k 反对称，所以，首先是 $R_{uuij}=R_{vvij}=0$，其中的 ij 涵盖了 uu,uv,vu,vv 四种组合，因此该式决定了 R_{lkij} 的八个分量为零。其次是 $R_{uvij}=-R_{vuij}$，也就是说，剩下的八个分量中，又只有一半是独立的。因此，我们只需要关注 R_{uvuu}， R_{uvuv}， R_{uvvu}，R_{uvvv} 即可。又因为 R_{lkij} 关于 i,j 也是反对称的，因此 $R_{uvuu}=R_{uvvv}=0$，$R_{uvuv}=-R_{uvvu}$。这样，我们只需要计算 R_{uvuv}，就能够用上述关系简单地推算出整个 R_{lkij}。当然，这只是对二维黎曼流形而言的。对四维黎曼流形就得计算二十个独立分量了。

利用上面的定义，考虑到对于共点圆几何，有 $g_{uv}=g_{vu}=0$，代入，得

$$\begin{aligned}
R_{uvuv} &= g_{uu}R^u_{vuv}\\
&= g_{uu}[\partial_u\Gamma^u_{vv}-\partial_v\Gamma^u_{vu}+(\Gamma^u_{vv}\Gamma^u_{uu}+\Gamma^v_{vv}\Gamma^u_{vu})-(\Gamma^u_{vu}\Gamma^u_{uv}+\Gamma^v_{vu}\Gamma^u_{vv})]\\
&= \frac{1}{(u^2+v^2)^2}\left[\frac{2(v^2-u^2)}{(u^2+v^2)^2}+\frac{2(u^2-v^2)}{(u^2+v^2)^2}\right.\\
&\quad\left.+\frac{(2u)(-2u)+(-2v)(-2v)}{(u^2+v^2)^2}-\frac{(-2v)(-2v)+(-2u)(2u)}{(u^2+v^2)^2}\right]\\
&= 0
\end{aligned}$$

这意味着共点圆几何的黎曼曲率张量所有分量全为零。

8.14.3 里奇曲率张量与标量曲率

R_{lkij} 的第一和第三角标缩并就是 Ricci 曲率张量，即 $R_{kj}=g^{li}R_{lkij}=R^i_{kij}$。因为 $R_{lkij}=R_{ijlk}$，所以 Ricci 曲率张量显然是对称的。

Ricci 曲率张量再缩并，就是标量曲率，即 $R=g^{kj}R_{kj}=R^j_j$。对二维黎曼流形来说，标量曲率是高斯曲率的二倍。

既然共点圆几何的黎曼曲率张量所有分量全为零，那么无论如何缩并，结果必定都还是零。所以，共点圆几何刻画的曲面，其高斯曲率为零。也就是说，这个曲面是平直的，实际上是个平面。

这一结果并非偶然。共点圆几何本来就是欧氏平面几何经过反演变换而做成的，其度规 $g_{ij} = \dfrac{1}{(u^2+v^2)^2}\delta_{ij}$ 也是欧氏平面几何的度规 $h_{ij} = \delta_{ij}$ 在保持微小向量模长平方不变的条件下，做了一个变量代换而得到的。在这种情况下，共点圆几何的黎曼曲率张量 R_{lkij} 与欧氏平面几何的黎曼曲率张量 R_{abcd} 有简单的坐标变换关系 $R_{lkij} = \partial_l^a \partial_k^b \partial_i^c \partial_j^d R_{abcd}$。这其实是张量在坐标变换下的变换关系。由于欧氏平面几何的度规各分量皆为常数，因此其仿射联络系数各分量皆为零，进而黎曼曲率张量各分量也都为零。按照张量在坐标变换下的变换关系，共点圆几何的黎曼曲率张量各分量必定都为零。

8.14.4 测地极坐标系

短程线是正规测地线，即以弧长为参数的测地线。象光锥这样的迷向子空间，也称各向同性子空间，其上任意两点之间‘长度’为零，因此无法以弧长为参数。但它也是测地线。从同一点出发的测地线称径向测地线。与给定点测地距离相等的所有点构成局地球面。显然，径向测地线与局地球面正交。这就是我们在图 8.34 中所看到的。不光是圆和球面，闵氏平面上的单位双曲线和闵氏空间中的单位双叶双曲面，也都是局地球面的例子。

在给定解析点的邻域内，用径向测地线和它们之间的夹角，可以建立局部测地极坐标系。二维黎曼流形的局部测地极坐标系可以分为三类：零曲率的，正曲率的，负曲率的。其实就是我们前面已经讲过的平面极坐标系，球面极坐标系，双叶双曲面上的极坐标系，统称测地极坐标系。我们来逐一计算它们的曲率张量。

8.14.5 平面极坐标系

平面极坐标系既然是刻画平面的，其曲率张量各分量必定全为零。为了确信这一事实，我们还是按部就班地计算一下。用平面极坐标系表达的平面上的微小向量模长平方为 $\mathrm{d}l^2 = \mathrm{d}r^2 + r^2\,\mathrm{d}\phi^2$，即度规的各分量为

$$g_{rr} = 1, \quad g_{\phi\phi} = r^2, \quad g_{r\phi} = g_{\phi r} = 0; \quad g^{rr} = 1, \quad g^{\phi\phi} = \frac{1}{r^2}, \quad g^{r\phi} = g^{\phi r} = 0$$

仿射联络系数各分量为：

$$\Gamma^r_{rr} = 0, \quad \Gamma^r_{r\phi} = \Gamma^r_{\phi r} = 0, \quad \Gamma^r_{\phi\phi} = -r, \quad \Gamma^\phi_{rr} = 0, \quad \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = \frac{1}{r}, \quad \Gamma^\phi_{\phi\phi} = 0$$

黎曼曲率张量的独立分量为

$$\begin{aligned}
R_{r\phi r\phi} &= g_{rr}R^r_{\phi r\phi} \\
&= g_{rr}[\partial_r\Gamma^r_{\phi\phi} - \partial_\phi\Gamma^r_{\phi r} + (\Gamma^r_{\phi\phi}\Gamma^r_{rr} + \Gamma^\phi_{\phi\phi}\Gamma^r_{\phi r}) - (\Gamma^r_{\phi r}\Gamma^r_{r\phi} + \Gamma^\phi_{\phi r}\Gamma^r_{\phi\phi})] \\
&= -1 - \frac{1}{r}(-r) \\
&= 0
\end{aligned}$$

可见我们的预想没有错。

8.14.6 球面极坐标系

用球面极坐标系表达的球面上的微小向量模长平方为 $\mathrm{d}l^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$，即度规的各分量为

$$g_{\theta\theta} = 1, \quad g_{\phi\phi} = \sin^2\theta, \quad g_{\theta\phi} = g_{\phi\theta} = 0;$$

$$g^{\theta\theta} = 1, \quad g^{\phi\phi} = \frac{1}{\sin^2\theta}, \quad g^{\theta\phi} = g^{\phi\theta} = 0$$

仿射联络系数各分量为：

$$\Gamma^\theta_{\theta\theta} = 0, \quad \Gamma^\theta_{\theta\phi} = \Gamma^\theta_{\phi\theta} = 0, \quad \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta,$$

$$\Gamma^\phi_{\theta\theta} = 0, \quad \Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi\theta} = \frac{\cos\theta}{\sin\theta}, \quad \Gamma^\phi_{\phi\phi} = 0$$

黎曼曲率张量的独立分量为

$$\begin{aligned}
R_{\theta\phi\theta\phi} &= g_{\theta\theta}R^\theta_{\phi\theta\phi} \\
&= g_{\theta\theta}[\partial_\theta\Gamma^\theta_{\phi\phi} - \partial_\phi\Gamma^\theta_{\phi\theta} + (\Gamma^\theta_{\phi\phi}\Gamma^\theta_{\theta\theta} + \Gamma^\phi_{\phi\phi}\Gamma^\theta_{\phi\theta}) - (\Gamma^\theta_{\phi\theta}\Gamma^\theta_{\theta\phi} + \Gamma^\phi_{\phi\theta}\Gamma^\theta_{\phi\phi})] \\
&= (\sin^2\theta - \cos^2\theta) - \frac{\cos\theta}{\sin\theta}(-\sin\theta\cos\theta) \\
&= \sin^2\theta
\end{aligned}$$

由此得 $R_{\theta\phi\theta\phi} = R_{\phi\theta\phi\theta} = \sin^2\theta, \quad R_{\theta\phi\phi\theta} = R_{\phi\theta\theta\phi} = -\sin^2\theta$。其他分量皆为零。代入 Ricci 曲率张量的表达式

$$R_{kj} = g^{li}R_{lkij} = g^{\theta\theta}R_{\theta k\theta j} + g^{\phi\phi}R_{\phi k\phi j}$$

得 Ricci 曲率张量的各分量为

$$R_{\theta\theta} = \frac{1}{\sin^2\theta}R_{\phi\theta\phi\theta} = 1, \quad R_{\theta\phi} = R_{\phi\theta} = 0, \quad R_{\phi\phi} = R_{\theta\phi\theta\phi} = \sin^2\theta$$

再缩并，得标量曲率 R 为

$$R = g^{kj}R_{kj} = g^{\theta\theta}R_{\theta\theta} + g^{\phi\phi}R_{\phi\phi} = 1 + \frac{1}{\sin^2\theta}\sin^2\theta = 2$$

高斯曲率是 $\kappa^2 = R/2 = 1$。

8.14.7 双曲极坐标系

用双叶双曲面上的双曲极坐标系表达的微小向量模长平方为 $-\mathrm{d}l^2 = \mathrm{d}\zeta^2 + \sinh^2\zeta\,\mathrm{d}\phi^2$，即度规的各分量为

$$g_{\zeta\zeta} = 1, \quad g_{\phi\phi} = \sinh^2\zeta, \quad g_{\zeta\phi} = g_{\phi\zeta} = 0;$$

$$g^{\zeta\zeta} = 1, \quad g^{\phi\phi} = \frac{1}{\sinh^2\zeta}, \quad g^{\zeta\phi} = g^{\phi\zeta} = 0$$

仿射联络系数各分量为：

$$\Gamma^{\zeta}_{\zeta\zeta} = 0, \quad \Gamma^{\zeta}_{\zeta\phi} = \Gamma^{\zeta}_{\phi\zeta} = 0, \quad \Gamma^{\zeta}_{\phi\phi} = -\sinh\zeta\cosh\zeta,$$

$$\Gamma^{\phi}_{\zeta\zeta} = 0, \quad \Gamma^{\phi}_{\zeta\phi} = \Gamma^{\phi}_{\phi\zeta} = \frac{\cosh\zeta}{\sinh\zeta}, \quad \Gamma^{\phi}_{\phi\phi} = 0$$

黎曼曲率张量的独立分量为

$$\begin{aligned}
R_{\zeta\phi\zeta\phi} &= g_{\zeta\zeta} R^{\zeta}_{\phi\zeta\phi} \\
&= g_{\zeta\zeta}[\partial_\zeta \Gamma^{\zeta}_{\phi\phi} - \partial_\phi \Gamma^{\zeta}_{\phi\zeta} + (\Gamma^{\zeta}_{\phi\phi}\Gamma^{\zeta}_{\zeta\zeta} + \Gamma^{\phi}_{\phi\phi}\Gamma^{\zeta}_{\phi\zeta}) - (\Gamma^{\zeta}_{\phi\zeta}\Gamma^{\zeta}_{\zeta\phi} + \Gamma^{\phi}_{\phi\zeta}\Gamma^{\zeta}_{\phi\phi})] \\
&= -(\sinh^2\zeta + \cosh^2\zeta) - \frac{\cosh\zeta}{\sinh\zeta}(-\sinh\zeta\cosh\zeta) \\
&= -\sinh^2\zeta
\end{aligned}$$

由此得 $R_{\zeta\phi\zeta\phi} = R_{\phi\zeta\phi\zeta} = -\sinh^2\zeta$, $\quad R_{\zeta\phi\phi\zeta} = R_{\phi\zeta\zeta\phi} = \sinh^2\zeta$。其他分量皆为零。代入 Ricci 曲率张量的表达式

$$R_{kj} = g^{li} R_{lkij} = g^{\zeta\zeta} R_{\zeta k\zeta j} + g^{\phi\phi} R_{\phi k\phi j}$$

得 Ricci 曲率张量的各分量为

$$R_{\zeta\zeta} = \frac{1}{\sinh^2\zeta} R_{\phi\zeta\phi\zeta} = -1, \quad R_{\zeta\phi} = R_{\phi\zeta} = 0, \quad R_{\phi\phi} = R_{\zeta\phi\zeta\phi} = -\sinh^2\zeta$$

再缩并，得标量曲率 R 为

$$R = g^{kj} R_{kj} = g^{\zeta\zeta} R_{\zeta\zeta} + g^{\phi\phi} R_{\phi\phi} = -1 + \frac{1}{\sinh^2\zeta}(-\sinh^2\zeta) = -2$$

高斯曲率是 $\kappa^2 = R/2 = -1$。

球面和双叶双曲面在上述给定的度规下，都是常曲率曲面。

8.14.8 局地圆的周长

在二维黎曼流形上给定点的邻域内，以给定点为坐标原点建立测地极坐标系，那么一个半径为 ρ 的局地圆的方程为 $r = \rho$，或者 $\theta = \rho$，或者 $\zeta = \rho$。现在我们来求这个圆的周长。

这个圆的周长等于绕行一周逐段微小向量模长的积分。由于这个曲线是局地圆，因此微小向量满足 $\mathrm{d}r = 0$，或者 $\mathrm{d}\theta = 0$，或者 $\mathrm{d}\zeta = 0$。这样，微小向量模长有 $\mathrm{d}l = \rho\,\mathrm{d}\phi$，或者 $\mathrm{d}l = \sin\rho\,\mathrm{d}\phi$，或者 $\mathrm{d}l = \sinh\rho\,\mathrm{d}\phi$。对整个圆求积分，就是对变量 ϕ 从 0 到 2π 求积分，从而局地圆的周长为 $l = 2\pi\rho$，或者 $l = 2\pi\sin\rho < 2\pi\rho$，或者 $l = 2\pi\sinh\rho > 2\pi\rho$。也就是说，对于零曲率、正曲率和负曲率的曲面，局地圆的周长直径比分别等于、小于和大于 π。

8.14.9 球对称外引力场的史瓦西解

已知球对称外引力场的史瓦西解是

$$\mathrm{d}s^2 = -(1-\frac{R_g}{r})c^2\,\mathrm{d}t^2 + (1-\frac{R_g}{r})^{-1}\,\mathrm{d}r^2 + r^2(\,\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)$$

其中 $R_g = \dfrac{2GM}{c^2}$。只考虑径向运动，则 $\mathrm{d}\theta = 0,\quad \mathrm{d}\phi = 0$，问题可简化为

$$\mathrm{d}s^2 = -(1-\frac{R_g}{r})c^2\,\mathrm{d}t^2 + (1-\frac{R_g}{r})^{-1}\,\mathrm{d}r^2$$

即该史瓦西解所定义的度规为

$$g_{tt} = -(1-\frac{R_g}{r})c^2,\quad g_{rr} = (1-\frac{R_g}{r})^{-1},\quad g^{tt} = -(1-\frac{R_g}{r})^{-1}\frac{1}{c^2},\quad g^{rr} = 1-\frac{R_g}{r}$$

其他分量皆为零。代入仿射联络系数的公式，得仿射联络系数各分量为：

$$\Gamma^t_{tt} = 0,\quad \Gamma^t_{rt} = \Gamma^t_{tr} = \frac{1}{2}(1-\frac{R_g}{r})^{-1}\frac{R_g}{r^2},\quad \Gamma^t_{rr} = 0,$$

$$\Gamma^r_{tt} = \frac{1}{2}(1-\frac{R_g}{r})\frac{R_g}{r^2}c^2,\quad \Gamma^r_{rt} = \Gamma^r_{tr} = 0,\quad \Gamma^r_{rr} = -\frac{1}{2}(1-\frac{R_g}{r})^{-1}\frac{R_g}{r^2}$$

代入短程线微分方程组的公式，得

$$\begin{aligned} \ddot{t} + (1-\frac{R_g}{r})^{-1}\frac{R_g}{r^2}\dot{r}\dot{t} &= 0 \\ \ddot{r} - \frac{1}{2}(1-\frac{R_g}{r})^{-1}\frac{R_g}{r^2}\dot{r}\dot{r} + \frac{1}{2}(1-\frac{R_g}{r})\frac{R_g}{r^2}c^2\dot{t}\dot{t} &= 0 \end{aligned}$$

牛顿的万有引力定律可以表示为：

$$\ddot{r} - r\dot{\theta}^2 = -G\frac{M}{r^2},\qquad r\ddot{\theta} + 2\dot{r}\dot{\theta} = 0$$

考虑到 $\dot{t}\equiv 1,\quad \ddot{t}\equiv 0$，并代入 $\dot{\theta} = 0$，该定律就是

$$\ddot{r} + G\frac{M}{r^2}\dot{t}\dot{t} = 0.$$

在弱场近似条件下，$\dfrac{R_g}{r} \ll 1$，史瓦西解趋于万有引力定律。史瓦西解的 $\dot{r}\dot{r}$ 和 $\dot{r}\dot{t}$ 项的系数与 $\dot{t}\dot{t}$ 的系数相比，相差 c^2 倍的量级，从而 $\ddot{t}$ 也十分微小，要在弱引力场中观测到相关效应并不容易。

8.14.10 球对称外引力场的新解

前面讲了那么多不同的度规，出于美学的考虑，我们可以合理假设度规具有 $g_{ij} = \rho(t,r)\eta_{ij}$ 的形式，其中的 $\rho(t,r)$ 是关于 t 和 r 的某种函数。展开来写，就是

$$\mathrm{d}s^2 = \rho(t,r)(-c^2\,\mathrm{d}t^2 + \,\mathrm{d}r^2)$$

度规各分量为

$$g_{tt}=-\rho c^2,\quad g_{rr}=\rho,\quad g^{tt}=\frac{-1}{\rho c^2},\quad g^{rr}=\frac{1}{\rho}$$

其他分量皆为零。代入仿射联络系数的公式，得仿射联络系数各分量为

$$\Gamma^t_{tt}=\frac{1}{2\rho}\frac{\partial\rho}{\partial t},\quad \Gamma^t_{rt}=\Gamma^t_{tr}=\frac{1}{2\rho}\frac{\partial\rho}{\partial r},\quad \Gamma^t_{rr}=\frac{1}{2c^2\rho}\frac{\partial\rho}{\partial t}$$

$$\Gamma^r_{tt}=\frac{c^2}{2\rho}\frac{\partial\rho}{\partial r},\quad \Gamma^r_{rt}=\Gamma^r_{tr}=\frac{1}{2\rho}\frac{\partial\rho}{\partial t},\quad \Gamma^r_{rr}=\frac{1}{2\rho}\frac{\partial\rho}{\partial r}$$

代入短程线微分方程组，得

$$\begin{aligned}\ddot{t}+\frac{1}{2\rho}\frac{\partial\rho}{\partial t}\dot{t}\dot{t}+\frac{1}{\rho}\frac{\partial\rho}{\partial r}\dot{r}\dot{t}+\frac{1}{2c^2\rho}\frac{\partial\rho}{\partial t}\dot{r}\dot{r}\ &=0\\ \ddot{r}+\frac{c^2}{2\rho}\frac{\partial\rho}{\partial r}\dot{t}\dot{t}+\frac{1}{\rho}\frac{\partial\rho}{\partial t}\dot{r}\dot{t}+\frac{1}{2\rho}\frac{\partial\rho}{\partial r}\dot{r}\dot{r}\ &=0\end{aligned}$$

与万有引力定律比较，可以做出一个‘合理的’，看着舒服的假设：

$$\frac{\partial\rho}{\partial t}=0,\qquad \frac{c^2}{2\rho}\frac{\partial\rho}{\partial r}=G\frac{M}{r^2}$$

解得

$$\rho(t,r)=\exp(-R_g/r)$$

代回去，得短程线微分方程组

$$\begin{aligned}\ddot{t}+\frac{R_g}{r^2}\dot{r}\dot{t}\ &=0\\ \ddot{r}+\frac{1}{2}\frac{R_g}{r^2}\dot{r}\dot{r}+\frac{c^2}{2}\frac{R_g}{r^2}\dot{t}\dot{t}\ &=0\end{aligned}$$

这个结果在弱场近似条件下，与史瓦西解的差别是增强了吸引力，增加的量为 $\frac{R_g}{r^2}\dot{r}\dot{r}$。换句话说，史瓦西解含有‘斥力项’，而我们的解没有‘斥力项’，全都是‘吸引力项’。不知道这能否消除黑洞问题。毕竟，到目前为止，还没有确凿的证据证明黑洞真的存在，尽管大家都热火朝天地讨论它，就像当年热火朝天地讨论以太一样。

8.14.11 两种解所决定的空间部分的曲率

为了考察‘同一时刻’的空间各部分之间的关系，我们假设某种‘虚曲线’不经历时间而有逐段的微小间隔模长，这意味着在该曲线上 $\mathrm{d}t\equiv 0$。为了进一步简化，再取 $\theta\equiv\pi/2$。这样，对于史瓦西解，有

$$\mathrm{d}s^2=(1-\frac{R_g}{r})^{-1}\,\mathrm{d}r^2+r^2\,\mathrm{d}\phi^2$$

即度规的各分量为

$$g_{rr} = (1-\frac{R_g}{r})^{-1}, \quad g_{\phi\phi} = r^2, \quad g^{rr} = 1-\frac{R_g}{r}, \quad g^{\phi\phi} = \frac{1}{r^2}$$

其他分量皆为零。仿射联络系数各分量为

$$\Gamma^r_{rr} = \frac{-1}{2}(1-\frac{R_g}{r})^{-1}\frac{R_g}{r^2}, \quad \Gamma^r_{r\phi} = \Gamma^r_{\phi r} = 0, \quad \Gamma^r_{\phi\phi} = -1(1-\frac{R_g}{r})r$$

$$\Gamma^\phi_{rr} = 0, \quad \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = \frac{1}{r}, \quad \Gamma^\phi_{\phi\phi} = 0$$

黎曼曲率张量各分量为

$$R_{r\phi r\phi} = R_{\phi r\phi r} = \frac{-1}{2}(1-\frac{R_g}{r})^{-1}\frac{R_g}{r}, \quad R_{r\phi\phi r} = R_{\phi rr\phi} = \frac{1}{2}(1-\frac{R_g}{r})^{-1}\frac{R_g}{r}$$

其他分量皆为零。Ricci 曲率张量各分量为

$$R_{rr} = \frac{-1}{2}(1-\frac{R_g}{r})^{-1}\frac{R_g}{r^3}, \quad R_{r\phi} = R_{\phi r} = 0, \quad R_{\phi\phi} = \frac{-1}{2}\frac{R_g}{r}$$

标量曲率为

$$R = \frac{-R_g}{r^3} = \frac{-1}{R_g^2}\left(\frac{R_g}{r}\right)^3$$

可见，史瓦西解刻画的时空的空间部分是负曲率的。

对于我们给出的新解，有

$$\mathrm{d}s^2 = \exp(-\frac{R_g}{r})\,(\,\mathrm{d}r^2 + r^2\,\mathrm{d}\phi^2)$$

即度规的各分量为

$$g_{rr} = \exp(-\frac{R_g}{r}), \quad g_{\phi\phi} = r^2\exp(-\frac{R_g}{r}), \quad g^{rr} = \exp(\frac{R_g}{r}), \quad g^{\phi\phi} = \frac{1}{r^2}\exp(\frac{R_g}{r})$$

其他分量皆为零。仿射联络系数各分量为

$$\Gamma^r_{rr} = \frac{1}{2}\frac{R_g}{r^2}, \quad \Gamma^r_{r\phi} = \Gamma^r_{\phi r} = 0, \quad \Gamma^r_{\phi\phi} = \frac{-1}{2}(R_g + 2r)$$

$$\Gamma^\phi_{rr} = 0, \quad \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = \frac{1}{2r^2}(R_g + 2r), \quad \Gamma^\phi_{\phi\phi} = 0$$

黎曼曲率张量各分量为

$$R_{r\phi r\phi} = R_{\phi r\phi r} = \frac{1}{2}\frac{R_g}{r}\exp(-\frac{R_g}{r}), \quad R_{r\phi\phi r} = R_{\phi rr\phi} = \frac{-1}{2}\frac{R_g}{r}\exp(-\frac{R_g}{r})$$

其他分量皆为零。Ricci 曲率张量各分量为

$$R_{rr} = \frac{1}{2}\frac{R_g}{r^3}, \quad R_{r\phi} = R_{\phi r} = 0, \quad R_{\phi\phi} = \frac{1}{2}\frac{R_g}{r}$$

标量曲率为

$$R = \frac{R_g}{r^3}\exp(\frac{R_g}{r}) = \frac{1}{R_g^2}\left(\frac{R_g}{r}\right)^3\exp(\frac{R_g}{r})$$

可见，我们的新解刻画的时空的空间部分是正曲率的。

光线在引力场中发生偏折的现象并不能判定空间曲率的正负。新解的曲率有个 $\exp(\frac{R_g}{r})$ 因子，可以用来对两个解孰是孰非做出判断。在太阳的边缘，$R_g/r \sim 10^{-6}$，如果我们的新解是对的，那么太阳边缘处的曲率应该比史瓦西解所预言的大百万分之几，要检验出来并不容易。但是，对白矮星边缘，$R_g/r \sim 10^{-4}$，对中子星边缘，$R_g/r \sim 1/3$，如果真的有差异，应该是可以检测到的。

Chapter 9

楚忠厚的幸福生活

今天就讲到这里吧。”楚老师看天色已晚，就准备收摊回家了。三个代表重要学生却依依不舍，仿佛还有无数的问题要问似的。

“那就到我家去继续聊吧，”楚老师一边说，一边从三轮车的车帮上，解下一把两米多长的折叠伞，竖着放进车座下面一个什么位置安装好，然后打开。大家惊奇地发现，这个伞展开后，直径竟然超过十米。

“这是城管同志发的，”楚老师看大家目瞪口呆的样子，就解释说。“摆摊卖菜、卖水果，人人必备，否则不让卖。所以，大家都把这个伞叫保护伞。”

等大家都上车坐定了，楚老师就奋力地蹬起了三轮车。车子并不动，但是那个伞却象鼓起的风帆一样，列列作响。不一会儿，车子居然飞了起来。

“飞碟！”小明立刻惊呼起来。

“而且是个三轮飞碟，”小娟跟着说。

“而且是个人力三轮飞碟，”小亮也跟着说。

“拜你科幻科一点好不好，”这位看官说了。“搞成魔幻、玄幻就没意思啦！”

“你是问这个飞伞的科学原理吗？”楚老师和蔼可亲、平易近人、循循善诱地谆谆教导说。“这车座底下是个空气压缩机，产生的高压空气被输送到伞的上表面，沿着表面向四周喷射。根据流体力学中的伯努利方程，流速越高，横向压强就越低。这样，伞的上表面压强大大低于大气压，而下表面压强基本等于大气压，于是，连伞带车就被大气压托起来了。”

“哦，”三个代表重要学生和那位看官都恍然大悟，豁然开朗了。

师徒四人一路西行，不多时来到一处黄沙万里之地。

“到家了，”楚老师找了个停车位，把车子停好。“这里就是平昌。当年我没有暂住证，经常被抓到这里来挖沙子，就像城里人经常到郊外度周末一样。”

“您说的是昌平吧，”小明不解地问。

“是吗？”这下楚老师也拿不准了。“你知道，这世上很多事情是颠倒的，主仆颠倒、黑白颠倒、是非颠倒、亲仇颠倒、次序颠倒……”

“颠倒众生啊，”如来佛哀叹道。

“是啊，”楚老师却只是微微一笑，泰然自若地说：“后来慢慢地熟了，就干脆住在这里了，免得总麻烦人家接送。那可是公务用车啊，怪腐败的。”

三个代表重要学生都开心地笑了。

大家进屋，洗了手脸，坐定了。楚老师端出一个果盘，上面整整齐齐摆了一层，大约八、九个苹果。

小明伸手想拿起一个，却发现这苹果似乎还长在盘子上。

“你得把它剪下来，”楚老师递过来一把剪刀，说：“这盘子是一颗人造苹果树。”

“人造苹果？”小娟愣了一下，喃喃自语地说。

“苹果是货真价实的植物苹果，”楚老师解释说。“但是，树是人造树。”

“这个盘子是一棵树？”小明接过剪子，随手放在一边，却端起盘子上下左右地看个没完。“它的根呢？它的叶呢？它怎么活啊？”

“准确地讲，它是一个人造结果枝，”楚老师说。“它的叶和根都在海南。”

“海南？”三个代表重要学生都糊涂了。“海南岛的那个海南？离这里有十万八千里，这棵树是什么神树啊，能离这么远隔空结果！”

楚老师拿出一个大可乐瓶，上面写着“某某牌苹果树液”，“海南岛某某公司制造”。大家都恍然大悟、豁然开朗了。

“果树枝条主要分营养枝和结果枝两类，”楚老师说。“营养枝和根都是制造和吸收营养的，结果枝则只负责结果。这样，海南的这家公司利用当地的阳光、气温、土质等有利的自然条件，种了大量只有营养枝的果树，把制造的营养液发售到世界各地。人们在家里，象打点滴一样把营养液输送给这些人造结果枝，就可以得到鲜活的苹果了。”

“那应该用真正的果树的结果枝啊，”小娟问。“为什么要用人造的结果枝呢？”

“真正的果树的结果枝支持不了这么多的果子，”楚老师说。“这就是果树要疏花疏果的道理。开三千朵可以结果的花，最后留不到三百个果。而人造结果枝就没有这个问题。三千朵可以结果的花，都移植到人造结果枝上，就能够结三千个果。这样我们种十分之一的树，就能够提供相同的产量，从而大大节省了土地。”

小明剪下一个苹果，尝了一下，味道真是好。于是大家就狼吞虎咽地吃起来。

“看来大家真的是饿了，”楚老师和蔼可亲、平易近人地说。他走到“饮水机”前，接了几杯“牛奶”，给大家喝。

“这是什么啊？”小亮闻了一下，没有奶香，倒是有一种面粉的香气。

“这是鲜麦浆，很好消化的，”楚老师一边说，一边打开“饮水机”的面板，给大家看里面一层一层的玻璃板。“每层玻璃板内部是两层人造叶片，中间夹一层发光板。这样就不再需要土地和阳光了。只要联合国主管的那几座核聚变反应堆不熄火，全人类就不会饿肚子。”

“早就听说人类吃粮食，就是吃阳光，就是吃太阳的奶，”小明说。“现在人类不吃阳光，改吃核聚变燃料了，就象孩子断奶，开始吃饭一样。”

“这样人类就可以脱离太阳妈妈的照顾，到宇宙深处去打工挣钱了。”小娟说。

“没错，”楚老师说。“这些缩微生态技术当中，有很多是首先为探索太空而开发的。当然，探索深海也是一样。只要在太空基地或者深海基地装两座核聚变反应堆，就可以驱动缩微的生态环境运转，让人类过上富足、愉快的生活。”

“我也想去那里生活啊，”小亮深情地说。

“那你现在就要好好学习，将来在那里才能找到合适的工作岗位，”楚老师说。“如果你只是想到那里去度假、享福，那可要花大价钱才行。”

同学们都开心地笑了。

小娟尝了一口鲜麦浆，说：“总听人家说小麦灌浆期，难道灌浆灌浆，指的就是这种浆？”

“对，”楚老师说。“营养成分已经进入小麦籽粒之中，但尚未固化，就是麦浆。”

楚老师又接了大半盆麦浆，加了些佐料拌匀，放进低压除水机里除去一些水分，麦浆就变成了面团。再放进烤箱，十几分钟出炉，就是香喷喷的面包。

“这种低压除水机就是抽气机，一直抽到水在常温下沸腾。这样，麦浆就可以在常温下快速失水，变成面团了。做面包，压面条，都可以。”楚老师说。

“可惜都是素的，没有牛奶和牛肉，”小亮一边吃，一边怅然若失地说。

“有啊，”楚老师起身到另一台“饮水机”前，为大家接了几杯牛奶。

小亮闻了闻，又尝了尝，点点头说：“嗯，是牛奶。”

小明却呆呆地望着那台出牛奶的“饮水机”，忽然问：“楚老师，您不会告诉我们那台饮水机是一头人造母牛吧！”

“猜对了，那就是一头人造母牛，”楚老师拿来一个烤盘，打开人造母牛的面板，取出一块大约有普通杂志那么大的板子，把上面约两厘米厚的牛肉揭下来放到烤盘上。

“这块板子就是人造牛骨，”楚老师一边说，一边拿出一张“保鲜膜”一样的东西，铺在人造牛骨上。“这是人造骨膜，上面有肌肉干细胞。一旦附着在人造牛骨上，就会恢复生物活性，不断生长，长成大块的牛肉组织。”

楚老师把人造牛骨放回去，合好人造母牛的面板。然后，在刚才揭下来的牛肉上撒些佐料，放进烤箱。十几分钟出炉，就是香喷喷的烤牛肉。

“这牛肉安全吗？”小娟狐疑不信地说。

“该技术一零年代就在实验室里做成功了，”楚老师回答。“这些年只是在扩大规模而已。安全方面，无论是理论上还是实践上，都没有发现任何问题。”

“照这么说，这牛奶也是从人造牛乳房里面挤出来的？”小亮心里五味杂陈，不知道自己该不该继续喝了。

“不需要挤，”楚老师一边说，一边打开人造母牛的另一块面板。大家看到一个方方正正的玻璃缸中，竖着放了几块象蜂箱的巢皮一样的东西，时不时地有乳汁从上面淌下来。

“那几块巢皮一样的东西就是牛乳腺组织，”楚老师说。“分泌的乳汁就直接收集起来了，根本不需要挤。”

“可是……”小亮越发地不自在了。“这样的牛奶真的能喝吗？”

“当然，”楚老师说着，自己端起杯子喝了一大口。“人造设备只是提供营养支持，这些牛肉组织、牛乳腺组织，是完全正常的动物组织，甚至可以向活牛体内移植，没有不良反应。这是很多实验确认过的事情。”

小亮壮起胆子又尝了一口，说：“尝是尝不出福尔马林的味道。可是，我总是会想到福尔马林。”

“我们到院子里看看这头人造母牛吃什么，你就放心了。”于是楚老师带大家来到院子里。

院子里百草丛生，蝴蝶和蜜蜂忙个不停。

楚老师找了一块草长得比较茂盛的地方，用割草机割了一大捆，用粉碎机粉碎，把一些胡萝卜和玉米也粉碎了，一起放到搅拌机里，一边搅拌，一边添加一种液体。

“这种液体是牛的唾液，是人造母牛的唾液腺组织分泌的。”楚老师解释说。“这几道工序实现的是牛的嘴巴的功能。”

然后，楚老师把拌好的牛饲料倒进一台机器的进料口。“这台机器实现的是牛胃肠的功能。出料口直通沼气池。在那里腐熟之后，就还田养草了。”

“那么，牛心牛肺牛肝牛肾都在哪里？”小亮问。

“尽量改成机械、半机械的装置，改不了的，就做成单独的功能组织，就象你们刚才看到的牛乳腺组织那样。”

“把好好一头牛，拆成这样七零八落的，各种激素水平还能正常吗？各种

肌肉组织、腺体组织还能正常吗？乳汁的成分还能正常吗？”小亮狐疑不信地说。

“所谓的驯化动物、培育新的生物品种，其实就是在做拆解和重组的工作，”楚老师说。“新品种与旧品种不一样了，某种意义上也是不‘正常’了。只是这种不‘正常’对人类不仅无害，有时还是有益的。所以我们就接受新品种。在一零年代，动物的生长对人类来说，还不是完全可控的，象这样的人造母牛，在当时做，风险比较大。但是这些年来，相关技术越来越成熟，原来不可能、不可靠的事情，现在已经变得可能而且可靠了。所以，现在这种人造母牛在超市就能买到，厂家上门安装，就象装空调一样平常。”

“那么人类就不再养牛了？”小娟问。

“这情况与当年马的情况类似，”楚老师说。“马肉并不好吃，所以古人养马，主要是打仗和拉车。但是，后来打仗开坦克，交通运输就开卡车、骑摩托，养马就没用了，所以规模越来越小。最后，主要是为了留种育种、表演和竞技而少量养马。类似的，在人造畜禽问世之后，大牧场和大养殖场就逐渐消失了。人们只为了留种育种和旅游观光而养殖畜禽，规模都不大。”

“畜禽？”小娟追问道。“鸡蛋也是人造鸡下的啦？”

“对，”楚老师说。“不过，人造鸡下的蛋，是真正的鸡蛋，与那种弄虚作假的人造鸡蛋可不一样啊。”

“那么，水产品呢？鱼鳖虾蟹，也可以人造了？”小明问。

“对，”楚老师说。“有人造鲑鱼，揭下来的鱼肉可以生吃。有人造雌蟹，一年四季源源不断地出蟹黄。”

三个代表重要学生都心潮澎湃、激动不已。

“那些蛇毒、蝎毒、蜂毒等过去十分稀少、珍贵的药材，都因为毒腺组织脱离机体单独培养，得以大规模生产，造成价格跳水，消费者受益。”楚老师说。“这正是技术进步改善人类生活水平的题中之义。”

这时，一只兔子从前面蹦蹦跳跳地跑过。

“兔子！您养了……”小明迅速扫视了一下院子。“五只兔子？”

“不，四只。”楚老师说。“有一只是傀儡兔。”

“傀儡兔？”大家都饶有兴趣地竖起了耳朵。

“就是有一定自主行动能力的智能机器兔，必要时，人也可以干预，精确地控制其行为，”楚老师说。“雄兔和雌兔发情不一定同步，有时双方会为此断打而两败俱伤。傀儡兔在雄兔发情时，就装扮成雌兔来交配，并通知主人及时取走精液，妥善保存。当雌兔发情时，傀儡兔就通知主人，请求装备一定的精液，扮成雄兔去交配。这样做比直接抓兔子做人工授精要温和得多，几乎没有应激反应。

“类似的技术应用广泛。比如，有时人们希望体型大的品种与体型小的品种杂交，以获得杂交优势，比如山东大黑驴的驴皮经济价值高，但是大黑驴数

量少，成熟慢，难以扩繁，而云南矮种驴则数量多，成熟快，容易扩繁。但是两者不可能自然交配。这时可以派傀儡黑母驴去收集黑公驴的精液，装备给傀儡矮公驴，去和矮母驴交配，间接地完成杂交。”

“原来就是人工授精的改进版本啊，”小亮说。

“除了人工授精，还可用于其他目的。比如大熊猫扩繁。人工繁育的大熊猫常常缺乏交配技巧，两情相悦，却无法成功。对这种情况，可以派傀儡大熊猫去培训它们，提高它们的交配技巧，最终实现自然繁育和野化的目标。

“再比如候鸟迁飞。由于人类活动的影响，有些候鸟改变了迁飞习性，长期持续下去就会灭绝。这时，可以派傀儡鸟去带领鸟群恢复迁飞的本性。一零年代就有人用小飞机成功地带领一群鹤恢复了迁飞。

“再比如，随着生态环境的改善，野生狼群开始壮大，对牧场、农庄、景区的人类生活构成威胁。这时，可以派傀儡狼去夺权，在狼群中建立傀儡政府，把狼群的活动控制在安全范围之内。”

“有创意！”三个代表重要学生都啧啧称奇。

这时，一只硕大的蜜蜂嗡嗡地飞过。吓得小娟不由自主地倒退了一步，说：“蜜蜂怎么会那么大个，难道是传说中的杀人蜂？”

“不是杀人蜂，是杀虫蜂，”楚老师说。“那是个小型扑翼式飞行器，装备有针孔摄像机、通讯部件、机械臂和纳米电容。它每天在园子里巡逻，把拍摄到的画面传回控制中心。控制中心能够对数千种害虫进行自动识别。一旦发现，就会派杀虫蜂去抓捕或者就地消灭。”

“它的能量来源是什么呢？”小明问。

“你看到园子两边那两串灯笼了吗？”楚老师说。“杀虫蜂在这边的一个灯笼上充好电，飞过园子，到那边的灯笼上再充。这样走之字形路线，就可以扫描整个园子。由于充一次电只飞一趟，所以它的纳米电容并不很重，不会给飞行造成困难。”

“我还以为蜜蜂都是授粉的呢，”小娟说。

“也有用于授粉的人造蜂啊，”楚老师说。“它们没有摄像机，也没有机械臂。放蜂时，要把多排充电灯笼并起来做成充电顶棚，人造蜂做一次授粉，就飞回顶棚充一次电，这样可以减小纳米电容的重量，缩小人造蜂的体积，从而扩大适用范围，能够给更多种类的花授粉。”

“它们怎么知道哪里有花呢？”小娟问。

“充电顶棚上有多架摄像机，”楚老师说。“这些摄像机把画面传回控制中心，由控制中心整合为立体信息。每朵花什么时候现蕾，什么时候开放，什么时候由哪一只人造蜂授的粉，所授的粉来自于哪一朵花，全都记录在案。”

“那得建多大的充电顶棚啊，”小亮说。

“可以做成移动式的，”楚老师说。“三、四排就够了。在拖拉机两边，一边装一个，扫过作业区。花多就多放些蜂，或者车子走得慢些。花少就少放

些蜂，或者车子走得快些。”

“哇，太高级了！”三个代表重要学生都称羡不已。

“其实，除了杀虫蜂，还有更大的田间杀虫机器人，”楚老师说。“老百姓称之为人造螳螂，可它们长得一点也不像螳螂。它们躯干部分有一只羊那么大，细长的腿使得它们在田间移动时很少践踏作物。它们可以装备两到十只弯曲自如的触手，顶端可以装备摄像机、也可以装备抓虫或灭虫的机械手。”

“这样的大家伙，能源问题怎么解决呢？”小明问。

“躯干里面有电池啊，”楚老师说。“还有的是高高举起一块太阳能电池板，但是并不朝向太阳，而是朝向田边的激光发射器。那个发射器把电能转化成激光，发到电池板上，再转化成电能，实现了田间无线输电的目标。”

“这下可省了在田间地头扯电线的麻烦了，”小明长长地舒了口气。

“对，”楚老师说。“有时一块田的四角放四个发射器。田间作业机器人举四块电池板来接收激光。这样可以做功率更大的事情。比如采收、剪枝等。”

“要是消耗功率更大呢？”小亮刨根问底地说。

“那就开拖拉机下地啦，”楚老师说。“拖拉机两边，一边装一个长长的悬臂，上面挂十几个机器人，采果、采棉、采茶，都是一遍过。效率可高啦。”

“林地怎么办呢？有机器人吗？”小娟也跟着问。

“林地机器人有象猴子那样攀爬的，也有象鸟一样飞行的，”楚老师说。“它们都靠充电车来补充能量。当然，也有传统的象踩高跷一样的步行机器人。它们的腿可以象浴室的伸缩杆一样，根据实际作业高度灵活设定腿的长度。”

“以前只听说工业机器人，现在农林牧副渔都广泛应用机器人了哈，”小娟感慨万千地说。

“对，”楚老师说。“以前养殖都是整批进整批出，现在搞生态养殖，强调自然生长，就不再有批次的概念了。于是就诞生了捕猎和捕鱼机器人。”

“不是能组织旅游者去观光、休闲、捕鱼、捕猎吗？”小明问。

“是啊，”楚老师说。“问题是单靠旅游者是捕不完的啊。剩下的如果不及时捕获，下一年就会失去控制，破坏生态平衡。人工去捕成本太高。所以就有了捕鱼捕猎机器人的用武之地了。”

说着说着，三个代表重要学生的肚子又饿了。于是，大家回屋继续吃饭。

“人要吃饭，就象机器要吃电一样，”小明说。“可是，我并没有在周围看到高压输电线路啊，您这里的电是怎么来的呢？”

“当然是太阳能和风能啦，”小亮抢着说。

“你说对了一半，”楚老师说。“还有一半来自于地热能。”

“这沙漠地区还有地热能？”三个代表重要学生都疑惑起来。

“地球上的任何地方都有地热能啊，”楚老师说。“只要你钻探得足够深。”

“我们说的地热能，是指那种具有工业开采价值的地热能啊，”小明说。“如果钻探一千米，地温只有六十度，而地面气温四十度，那就是没有工业开采价值。因为热功效率只有百分之一左右，太低了。”

“你的评价标准不对，”楚老师说。“工业开采价值是个经济概念，不是物理概念。地热的成本主要是一次性投入的设施成本，和微薄的维护成本，热能本身是零成本的。假设我这里的电力需求是一万瓦，热功效率是百分之一，那我就让地热以一百万瓦的功率流过就行了，哪怕热功效率只有千分之一，也不过就是让地热以一千万瓦的功率流过而已，虽然投资规模会大些，但是考虑到地热本身不要钱，初期的投资大些是完全可以接受的。

“实际上，我的地热发电厂是个倒装的空调。在一千米的地下，铺设了蒸发器的管网，地面阴凉处，则装了冷凝器的管网。工作介质就是蒸馏水。在底部存储冷凝水的地方装了抽水调压阀门。整个管网内部被抽成低压，直到地下的蒸发器管网里的水能够沸腾为止。这样，水吸收地热变成五十度以上的蒸汽，回到地面，通过涡轮机发电后，变成四十度左右的蒸汽。在冷凝器管网中再变成四十度左右的水，流回地下蒸发器，进入下一个循环。注意，整个涡轮机是被封装在管道内的，只有所发的电通过接线柱输出，而蒸汽是不会泄露的。所以，这种发电厂的工作介质是一次灌装，长期循环使用。

“只要多抽一些水，管网内压强降低，蒸发器内水的沸点降低，吸收地热的速度就能够提高。另一方面，扩大管网覆盖面积也能有效地提高发电厂的输出功率。还有个办法就是换成潜热低的工作介质。或许可以考虑乙醇、或者乙醚，但是相关的防火防爆措施又会增加成本。所以，选什么样的工作介质，还是需要综合计算一下才行。”

“工业上五十度以上的废热比比皆是啊，”小亮说。“为什么一直没有人建这样的低温蒸汽发电厂呢？”

“因为工业废热虽然免费，但是供应不集中，量不够大，而且不稳定，”楚老师说。“只有在稳定可靠地大量供应免费热能的前提下，巨额的设施投资和很低的热功效率，才是经济上可行的。地热正符合这个前提。”

“这样的发电厂，环保上会不会有麻烦，”小明说。“地下管网有这么大的覆盖面积，其温度从六十度降到了五十度，地表的地温势必也会降低，植物生长和动物繁育都会受到影响。”

“如有必要，你可以把冷凝器埋一部分到地表土层之中。这样，地表的地温不仅不会下降，反而会得到提升呢。”楚老师说。

“为了发那么一丁点的电，就把那么大量的地热排入大气，”小明说。“会不会对气候有影响？”

“地热的开发减少了温室气体的排放，使得大气热量可以正常辐射到太空之中。”楚老师说。“这对全球气候实际上是有益的。”

“但是，地心呢？”小亮说。“地壳的导热率有限，实际上是保护了地心，维持地心发热与散热的平衡。现在，这些地热发电厂钻井吸热，会不会导致地心状况失衡，因散热过多而停止发热，这样地球就成冰疙瘩了。”

“地心发热与散热从来都不是平衡的，否则就不会有地震和火山喷发了，”楚老师说。“地热发电厂只是把剧烈的破坏性的散热，变成平缓的建设性的散热而已。至于会不会散热过度，只要看地震和火山喷发会不会停止就可以了。什么时候停止，就在什么时候关闭地热发电厂。”

“听说生命不只碳基蛋白质这一种形式，”小娟说。“地下岩浆当中，也可能存在以高温高压环境下微晶结合为基础的生命形式。我们开采地热，会不会破坏他们的高温高压生态环境，逼得他们穿上保温服，冲出地表，来和我们拼个鱼死网破？”

“不知者不为过，”楚老师说。“古人云，举头三尺有神明。如果你害怕得罪这些或许有、或许没有的‘神明’，那就什么都没法干了。如果我们见到了他们的谈判代表，或者至少是发现了相关的可疑迹象，那么，我相信人类一定会认真对待，谨慎行事。毕竟，我们还有很多其他的选择，不是一定要开采地热不可。但是，现在，‘高温高压生命形式’还只是科幻，我们不必杞人忧天。”

“如此说来，这种低温蒸汽地热发电厂如果推广开来，真的是利国利民啊，”三个代表重要学生都感到欢欣鼓舞了。

“水是生命之源，”小明喝了一口鲜麦浆，幽幽地说。“您在这黄沙万里之地，是怎么弄到水的呢？”

“打电话叫送水公司送的呗，”小亮打趣说。

“出市区人家就不送了，更别说这沙漠中心地带啦。”小娟说。

“黄河之水天上来，”见大家都很开心，楚老师也幽默地说。“我用的水也是从天上来的。”

“雨水？”三个代表重要学生都愣住了。“这里还经常下雨吗？”

“这里几乎不下雨，”楚老师说。“但是这里有的是电啊，有电就有水。”

“抽地下水？”大家还是不明白。“您在这里找到地下水了？”

“干嘛非得是地下水啊，”楚老师说。“用电抽天上的水不行吗？”

“伸根管子到云层里去吸水？”大家越听越糊涂。

“根本不用爬那么高啊，”楚老师说着，就带大家又来到院子里，打开“下水道”的盖板，让大家看一个正在噼里啪啦滴水的东西。

“这不就是一台壁挂式空调机吗？”小亮说。“空调不挂在高处，埋在地下干什么？”

“原来您是用空调来吸收空气中的水分啊，”小明却恍然大悟、豁然开朗了。“说来也是，空调本来就有除湿的功能。您只是把这项功能别出心裁、独具匠心地用在了另一件事情上而已。”

“过去，温室大棚除湿除雾是个老大难问题，”楚老师说。“棚膜上凝结的水珠如果滴落到植物上，很容易导致植物病变。现在，电的成本微乎其微，用空调除湿在经济上变得可行了。我这个园子里，虽然草木茂盛，但是空气却很干燥，屋顶玻璃上从来不结水珠。”

“然后您用除湿抽出来的水浇灌植物的根，形成完整的水循环，”小明说。“怪不得您的院子，虽然是个完全密闭的环境，却象外面的原野一样，空气清新。”

“说到原野，你们真该看看我开辟的绿洲，”楚老师带大家来到外面，放眼望去，不下千亩的梭梭林让大家都不敢相信自己的眼睛。

“这都是您开辟的？”小明问。“即便是梭梭这样的沙漠植物，想要种活也不容易，您怎么解决水的问题啊？”

“说千遍不如做一遍，”楚老师说。“大家跟我去开一块地，实践一下就知道了。”

大家来到绿洲边缘。楚老师挑了一块直径四米的空地，用挖掘机在中心位置挖了一个两米多深的坑，底部直径接近两米。然后在底部铺了三层防渗膜，四周留了接近一米的高度，形成一个不透水的盆子。然后回填，到把盆子刚好埋住的时候，把一根水管竖着放在盆子中心的位置，然后完全回填。

“这块地的改造就完成了，”楚老师说。

他又找来一桶水，从水管中慢慢灌下去，并拧紧管口的盖子。然后在离水管约七十公分远的四个角上，种了四棵梭梭苗，并浇了定根水。

“那一桶水会储存在盆子里，并通过毛细浸润现象慢慢释放，”楚老师说。“这四棵梭梭苗定根之后，就会向有水的地方长根，成活率相当高。我称这种技术为地盆栽培。用地盆储水相当于提高了地下水的水位。”

“我听说植物的根不仅需要水，还需要氧气，”小亮说。“花盆底部都留有沥水口，就是怕积水不透气。难道您这地盆积水不会导致问题吗？”

“矛盾处处有，在具体的事情上总有主次之分，”楚老师说。“这沙漠地带的主要矛盾是缺水，而不是积水。盆外的土壤是不积水的，地盆之上一米厚的土壤也是不积水的，这些空间足够很多植物保根活命了。大雨过后的几周，地盆内部确实会积水，但是，这里小雨都很少，大雨更罕见，绝大部分的时间里，地盆里的积水不超过半米，对植物根系没有危害。”

“您是说，因为埋得深、离得远，所以没有危害吗？”小娟问。“这是否也意味着地盆积水对植物根系是鞭长莫及、爱莫能助呢？”

“绝对不是，”楚老师说。“沙漠植物都很耐旱，有点潮气就能活。半米地盆积水，通过毛细浸润作用可以保持周围三米范围的土壤有潮气，足够这四

棵梭梭活半年。平时我给它们浇水，都是三个月浇一桶，就象刚才那样慢慢灌到水管里，然后封好口就行了。”

“这么大面积的绿洲，每个水管三个月浇一桶，也需要不少水呢，”小明说。“在这户外的开放环境中，您也用空调吸水？”

“是啊，”楚老师说。“你们看到那一排排的百叶箱了吗？每天凌晨三点到六点，它们都会开动，吸收空气中的水分。冷凝下来的水除了灌溉绿洲之外，还可以补充到水库里储存起来。”

“您这里还有水库？”大家都很惊讶。

“有啊，就在这片绿洲的最低点，”楚老师说。“我专门测绘过，下雨的时候，地表径流一定从那里经过。”

大家顺着楚老师指的方向望去，只看到一个小炮楼。正困惑的时候，楚老师说：“为了减少蒸发，水库都建在地下。那个炮楼是地下水库的入口。”

楚老师带大家来到炮楼前，打开门，下了一段楼梯，才看到一片象地下停车场一样宽敞的地方，里面是一排一排的池塘。有的池塘里长满了芦苇，有的池塘里养了鱼，有的池塘则是空的。

“这些长芦苇的池塘实际上被一系列挡板隔成了蛇形的水路，”楚老师说。“养鱼产生的含有残饵和粪便的脏水从水路的一头进入，经芦苇吸收、净化之后，从另一头流回鱼塘。”

“我发现您无论建什么，都要建成循环的小系统，”小亮说。

“是啊，流水不腐嘛，”楚老师说。“多种可控的、有益的生物，或至少是无害的生物，互助共存，抢占有利的生态位，把系统内的资源瓜分干净，可以防止不可控的、有害的微生物的侵入和蔓延。这是一种整体的免疫能力，对有关各方都有利。尤其是，对我有利。我可以得到不腐败的水。”

“那么，我们老百姓，要建立怎样的系统，才能得到不腐败的政府呢？”三个代表重要学生紧张地交换了一下眼色，都低下头不吭声了。

“您在地下挖这么大的一个空间，一定花了不少力气吧，”良久，小明抬起头，扫视了一下这个地下水库，幽幽地说。

“肯定是操纵机器人来挖，”小亮说。“反正这里有的是电。”

“没错，”楚老师说。“自从可控核聚变反应堆建成之后，人类在燃料和动力方面，乃至食物方面，都再无匮乏了。问题是，过去的石油和煤炭工业还有提供化工原料的功能。现在，化石燃料枯竭了，人们需要寻找化工原料的新的来源。这就是我大面积种植梭梭的原因。”

“梭梭？化工原料？”小明歪着脑袋想了想，说：“您指的是木焦油吗？”

“多少木材才能干馏出那么一点木焦油啊，”楚老师摇摇头说。“宝贵的纤维素都变成无机碳啦，太浪费了。”

“那就是人们找到能够用木材来酿酒的方法了，”小明说。

“人类已经不缺粮食了，还会缺酒精吗？”楚老师又摇着头说。“在动物界中，对纤维素的消化吸收能力最强的是白蚁。白蚁肠道中有一种共生的细菌，能够消化纤维素，供白蚁吸收。人们据此发明了人造白蚁肠道，提供一个特殊的生化环境，让这种细菌可以在其中生存。然后把木材磨碎，拌上白蚁唾液腺组织分泌的唾液，送到人造白蚁肠道中进行消化。消化完成后，再把其中的细菌灭活，这样制出的产品就称为白蚁肠糜。白蚁肠糜的进一步加工就非常多样化了。比如，送入人造牛肠，吸收出来的营养送往牛脂肪细胞，就可以大量生产牛油；送入人造蜘蛛肠，吸收出来的营养送往蛛丝腺组织，就可以大量生产蛛丝。”

“蛛丝？”三个代表重要学生都惊呼起来。“是那种可以做防弹衣的蛛丝吗？大量生产？如果这种蛛丝可以象尼龙丝一样大量生产，世界真的要大变样啦！”

“难道世界还没有大变样吗？”楚老师说。“简言之，生化工厂代替了化工厂，白蚁肠糜代替了原油。”

“哇，”三个代表重要学生都心潮澎湃、感慨万千起来。“谁会想到，小小的白蚁，会成为未来世界经济的……的……的什么呢？的新生命的新血液啊！”

“这只能说明生物多样性的重要，”楚老师说。“不是白蚁这种生物有多么重要，而是任何一个物种都非常非常的重要，都有自己存在的价值，都有自己的用武之地，只是需要等待合适的时机，需要人们去发现罢了。”

“木材是白蚁肠糜的原料，所以您种植梭梭，”小亮说。“可是，如果人造叶片可以生产麦浆，为什么就不可以生产木材呢？”

“当然可以生产木材啊，”楚老师说。“自然长成的树干都是圆柱形的，这是支撑树冠的需要。现在的人造树没有树冠了，人们可以把树干的生长层铺成平面，控制好各部分的生长速度，让一棵树直接长成板材。”

三个代表重要学生都被这绝妙的主意惊呆了。“可是……可是……您刚才说，您大面积种植梭梭，是为了得到木材，”小亮结结巴巴地说。

“是小明说要用木材酿酒什么的，我就说木材可以做成白蚁肠糜，”楚老师说。“实际上，我大面积种植梭梭是为了生物多样性。现在，这片绿洲已经把地面固定住，开始有杂草，昆虫和老鼠了。我打算明年就引进兔子、狐狸和羊。你们可能不知道，动物的排泄物对生态环境的形成和生长是十分必要的。”

“原来您是要建一座属于自己的伊甸园啊，”三个代表重要学生都恍然大悟、豁然开朗了。

“人的个性、人的思想有时可以差异很大，就象社会生态中的不同物种一样。这种多样性对社会的长期稳定发展同样具有基础的重要作用，我更愿意建一座精神的伊甸园，为人的个性、人的思想的多样化提供庇护。”楚老师幽幽地说。

大家又回到屋里。休息了一会儿，小明突然说：“关于地热发电厂，我有一个好主意，可以省去地面冷凝器管网和工作介质，还能大幅提高输出功率。”

“说来听听，”大家都兴致盎然地洗耳恭听。

“我先介绍一个简化的方案，”小明说。“还是一千米的深井。放两个管下去，一个走下行的冷空气，称为冷气管；一个走上行的热空气，称为热气管；井底铺设水平的、蜿蜒而行的吸收地热的管道，称为地热管。地面上三十度的空气，沿冷气管下行到井底，走地热管被加温到六十度，再沿热气管升回地面。由于冷空气和热空气的密度差，这样的管道系统铺设完成之后，只要从热气管中抽半个小时的气，使热气管和冷气管产生温度差，这种对流就会自我维持，形成永不停息的‘地热风’。用这个风去推动涡轮机，就可以发电了。

“显然，只要拉大热气管和冷气管中的温度差，就可以提高发电的输出功率。我的办法是：在井底装一台空调，蒸发器放在冷气管中，凝结器放在热气管中。这样，三十度的地面空气下行到井底，被空调蒸发器降温变成十度，再进入地热管去吸收地热，显然其吸收地热的能力就大大提高了。回到热气管时，设其温度是五十度，再被空调凝结器加热到九十度，升回地面。这样，原本三十度进、六十度出的‘地热风’，就变成三十度进、九十度出了。发电的输出功率当然也随之提高。”

“如果你多发的电还不够空调本身的耗电，那岂不是得不偿失，白白耗散了更多的地热吗？”小亮问。

“在井底热气管中装个小涡轮机，直接驱动空调的压缩机即可，”小明说。“这样，空调本身就不耗电了。”

“不耗电也要耗能啊，”小亮说。“假设我们不用空调，而是在井底热气管中装个小涡轮机发电，发的电供一段电炉丝给热空气加热，提高‘地热风’的温度，这可能吗？不可能。因为六十度的地热风，发电后剩四十度，所发的电再给地热风加热，刚好恢复到六十度。这是能量守恒定律之必然。”

“在空调方案中，能量是守恒的牙，”小明说。“出来的地热风之所以温度高，是因为吸收了更多的地热。”

“吸收热源的热量并不能使自己的温度高于热源，”小亮说。“井底地温六十度，出来的地热风就不可能高于六十度。如果用空调来强制提高，那一定需要外界给空调提供动力。如果从外部输入电力，那就是耗电多于发电，得不偿失；如果从工作介质中抽取机械动力，那就是先降温后加温，等于不升不降。”

“地核算不算外界？从地核输入热能，算不算从外界提供能量呢？”小明说。“实际上，井底岩石只是导热媒介，真正的热源是地核。地热风温度的上限是地核的温度，而不是井底岩石的温度。空调方案相当于提高了岩石的导热率，也相当于把井打得更深，因此得到的地热风温度更高。这有什么问题吗？”

“想法非常好！能把知识真正地吃透，有所创新，有所发展，非常好！”楚老师微笑着点点头，称赞道。“我再给你锦上添花一下。井的内径一米五，中间只要放一个外径一米的热气管即可。冷空气从热气管外面下行。到井底分为四支，每个分支装一个空调。每支冷空气被空调降温后，负责九十度扇形区域的地热的吸收。吸热之后回来，经空调凝结器加热，汇合到热气管升回地面。这样，某一个空调检修维修时，就不致造成整个发电系统停工。”

“楚老师想的可真周到啊，”三个代表重要学生都开心地笑了。

他们这样开心，是有很好的理由的。因为后来，这三个代表重要学生合伙创办了一家公司，修建了很多很多地热发电厂，极大地改善了不通高压输电线路的地方的人民生活。五年之后，公司股票在斯纳达克上市，公司也一跃成为全球最大的能源零售商之一，而三个代表重要学生作为创办人，个个都成了万元户。

这时，电视里播报新闻说：欧洲科学家测出中微子速度高于光速。

“相对论问世一百多年了，”小明说。“终于突破光速了。”

“要严格地‘科’，那是很难的事，”楚老师说。“但是，如果允许‘幻’，那就容易多了。

“比如，我就不喜欢宇宙大爆炸理论。那么哈勃红移该如何解释呢？我们用一个不断扩大的球面波前来描述光在太空中的传播。假设宇宙密度大体均匀，那么，这个球面所包含的体积和质量与半径的立方呈正比。这个球体对球外质点的引力等效于所有质量集中于球心而对外的引力。引力的强度与半径的平方呈反比。这样，随着光的波前半径不断扩大，光子受到的向内的引力呈正比地增大。而光在引力场中会发生红移，所以，光传播的距离越远，发生的红移越大。这就是哈勃红移。

“我也不喜欢时空膨胀的理论。按照这个理论，无法解释远处星系的结团状况，以至于不得不用暗物质、暗能量来凑数。为什么不能等效地假设时空均匀，而光速变慢呢？光速在近距离内是相对恒定的，但是随着传播距离变远，瞬时速度逐渐变慢。假设在很遥远的星系上，一件事情按照和我们这里一样的速度发生并结束。发生时产生的光和结束时产生的光先后到达地球。因为这个星系很遥远，两束光传到地球，速度已经变慢。我们接收到第一束光之后，很久才接到第二束光。按照光速不变的理论，这说明遥远星系的时空膨胀了。可是，如果我们放弃光速不变，难道不是更自然吗？而且避免了暗物质、暗能量的困扰。这不是推翻，而是发展。光速不变原理在宇宙的小尺度上成立，在大尺度上，不一定成立。

“按照电磁相互作用展开均匀的背景时空，把引力理论建立在电磁背景时空上，这就是广义相对论。反过来，按照引力相互作用展开均匀的背景时空，把电磁相互作用的理论，建立在引力背景时空上，这是完全可行的，只需做一个数学上的变换即可。两种背景时空在宇宙的小尺度上是一致的，在大尺度

上，采用引力背景时空更为方便，可以避免爆炸、膨胀、存在但不可见等等别扭、牵强的解释。

“再有第三种相互作用加入，情况就更加微妙了。中微子据说是参与弱相互作用的。现在观测到中微子速度高于光速，那么在电磁背景时空中展开弱相互作用的理论，看来会遇到一些麻烦。如果在引力背景时空中展开呢？这实际上是几种基本相互作用之间的耦合问题。

“不知道将来的发展，会被上述‘科幻’说中多少。”

“可是，您还是没说中微子的速度为什么高于光速，”小明说。

“如果光速不变原理不是‘客观’的，而是人们主观选择的一种看问题的角度，”楚老师说，“那么，超光速就无需问为什么了，因为粒子有各种不同的速度是再正常不过的事情。过去人们一直把主观偏见当客观事实，现在发现了违背偏见的事实，就觉得‘不正常’了。”

“这‘偏见’可不是没有道理的，”小明说，“有太多太多的事实与这个‘偏见’相符合，实在很难想象这个‘偏见’怎么可能不是事实。”

“有多少事实与平面几何的平行公理相符合呢，”楚老师说，“那么，平行公理是不是‘客观事实’呢？不是。改变平行公理，可以建立同样好的几何。光速不变原理也是类似的情况。打破光速不变，可以建立同样好的物理。超光速中微子的发现，把光速不变原理、相对论和爱因斯坦请下神坛，恢复了他们有贡献、也有局限的真实地位。”

“打破光速不变，可以建立同样好的物理？”三个代表重要学生都狐疑不信地说。“现在，有这样的理论吗？”

“重整河山，待后生啊，”楚老师深情地说。

三个代表重要学生都心领神会地笑了。他们每个人都暗暗下定了决心，一定要好好学习数学和物理，时刻准备着，为建立这样的理论做出自己应有的贡献。

这时，电视新闻又说：目前，卫星轨道范围内已经积聚了数万件太空垃圾，主要由火箭推进器残骸和废弃卫星组成，也有象神六宇航员出舱时，被舱内残余气体吹出来的一本操作手册这样的小件物品。

“现在的太空就象一块钢化玻璃，”小明担心地说。“一旦发生碰撞，四溅的碎片势必撞毁更多的卫星，并改变它们的轨道，继而引发更多的碰撞。这样的多米诺骨牌效应可以在几个小时内将全球的卫星系统和太空站全部摧毁。太空将变成一个大垃圾场。”

“联合国正在筹建一个类似于国际原子能机构的组织，就太空环境问题进行协调，”楚老师说。

“我支持，”小亮说。“应该对卫星和太空站的发射实行严格的审批，提高准入门槛，限制太空飞行器的数量。”

“恰恰相反，”楚老师说。“联合国正在鼓励发射市场的多元化，试图用经济手段引导市场，开发更多主动清理太空垃圾的飞行器。”

“就是高薪诚聘太空清洁工，”小娟说。

“对，”楚老师也被小娟言简意赅的精妙解读逗乐了。“联合国想建立一个太空垃圾清运市场。首先是对太空垃圾进行识别，确认责任方。然后就该垃圾的清运任务进行招标。中标公司干活，责任方出钱，联合国出面协调和监督，争取五年内把可识别的太空垃圾清理干净。”

三个代表重要学生相互递了个眼色，击掌相庆曰：“发财啦！发财啦！”

师异其言，曰：“愿闻其详。”

小明曰：“美帝不是搞过一次导弹打卫星的实验吗？现在对于太空垃圾，我们不是要用导弹去迎头撞它，而是要从后面追上它，打开舱门，伸出机械手，抱紧它，把它拖进大气层坠毁。”

于是，三个代表重要学生又合伙创办了一家公司，专门设计制造“捡垃圾”导弹。用飞机带到万米高空，从空中发射进入卫星轨道，捡拾太空垃圾。开始的时候，他们精心计算太空垃圾的运行轨道，寻找几十块垃圾集中同方向飞行的机会，用大型运输机一次发射几十枚导弹，将整批垃圾一网打尽。后来这样的机会越来越少了，才逐渐改用民航客机，直至最后用歼三十战斗机，来发射导弹。由于这些措施极大地降低了成本，所以他们的公司在招标中屡投屡中。短短五年时间，他们个个都成了万元户。

这位看官问啦，为什么这些措施可以降低成本呢？

小明说：“首先，被飞机带到万米高空，成本比增加一级助推火箭要低。这主要是因为助推火箭是一次性的，而飞机则可以反复使用。第二，飞机的飞行方向调整到与太空垃圾飞行方向一致，这样，导弹起飞时就具有了有利的初速度，从而可以少携带一些燃料。第三，万米高空空气稀薄，导弹飞行时遇到的阻力小，因而节省燃料。第四，空气稀薄，导弹飞行时表面发热很少，防热措施大大简化，因而减小了导弹的自重。第五，导弹功能专一，结构简单，所以自身体积小、重量轻。第六，一次发射多枚导弹，进一步摊薄了清运任务的固定成本。”

大家都恍然大悟、豁然开朗了。

“看到楚老师在地下水库里养鱼，”小明说。“我想到一个好主意。”

“说来听听，”大家都兴致盎然地洗耳恭听。

“南方有些城里人在家里把水盆放在架子上，摆上好几层，在水盆里养龟、养鳖。”小明说。“很显然，不用这种方式养鱼是因为浅水鱼的经济价值普遍不高，没有值得养的品种。如果是两米左右的深水鱼，那么可以选择的高价值品种就多了。问题是，在城市楼房里，蓄两米深的水，太重，面积太小，不划算。

“我的办法是：做一米长、两米宽、二十厘米高的密封舱，在边缘处加装

一个两米高、直径五厘米的管子。注满水后，密封舱内的压强就是两米深的水压，完全适合深水鱼的生长。这样就省去了百分之九十的水，而且可以一层摞一层，摞个五、六层不成问题。”

说干就干，三个代表重要学生立刻合伙创办了一家公司，在城里租了一套房子，把上下铺的下铺床板拆掉，腾出空间放置密封舱。人就住在上铺。如此苦干三年，他们个个都成了万元户。

后来，他们又在朝阳沟租了十亩地，建了十米高的厂房，把密封舱摞了四十层，当年就收获了别人四百亩深水池塘才能收获的产量，被人民公社评为建设会社主义新农村积极分子。

“你们为什么不制作密封舱，对外销售呢？”楚老师问。

一语惊醒梦中人。三个代表重要学生回想起当年在出租屋里创业的情景，都唏嘘不已、感慨万千。“我们一定要让城里人，都体验一下我们创业的艰辛。”他们说。于是，他们把密封舱做成时尚家具，有沙发，也有床；把那根管子做成艺术化的灯柱。有时还用玻璃来做，里面养上金鱼或者水母，灯光一照，如梦如幻，漂亮极了。

自从三个代表重要学生把他们的密封舱家具推向市场，世界渔业就一蹶不振，陷入了不可逆转的衰退之中。因为，家家户户，沙发底下，床底下，都是鱼！

“这位看官，如果我是你，我就会建个饵料厂，”楚老师乐呵呵地说。

水木社区科幻版的网友 PSOneCombo 高屋建瓴地指出：前面那个清理太空垃圾的方案不妥，应该采用迎头撞击，或者用捕捉网拦截的方案，以最大限度地利用自身动量。受控坠落需要变轨发动机和更多燃料，从而造价和发射成本都会大增，不如随机坠落。预先买个保险，真要是造成什么损失的话，由保险公司去赔。

大家纷纷表示这位网友说的很有道理。

“这个问题太‘科’了，”小亮说。“没多少‘幻’的余地啦。”

“那倒不一定，”小明说。“比如这个捕捉网，不该是个普通的网。导弹依然是细长的，便于发射。进入撞击轨道之后，导弹会展开，成为一个多层的‘大伞套小伞’的结构。最外层是高强度的布，里面各层是铝板。太空垃圾撞进敞开的伞口，被一层层的铝板吸收能量。最好的情况是嵌在这个结构中，一起坠落。就算是打穿了，多数碎片也会被嵌在结构中，不会四处飞溅。”

“现在太空站越造越大，成了真正的太空城了，”小娟说。“这种‘吸收式反撞’导弹成了太空城家居必备的防空武器。”

“你是说太空武器化？”小亮说。“太空清理与防卫应该由中立机构来做吧。这和商船不配武器是一样的道理。”

“协议也是动态的，”楚老师说。“重量级的不能配，轻量级的呢？长距离的不能配，短距离的呢？都可以商量。另外还有指挥权的问题。联合国可以

成立一个太空安全委员会，指挥一只中立的保安力量，不妨称为太安队。然后给每个太空城配太安队员和‘吸收式反撞’导弹。这样就没问题了。”

“联合国的维和部队在地面上起了什么作用呢？”小娟说。“地面上做不好的事，到太空中就能做好了？”

“人类文明总是在不断发展的，”楚老师说。“比如，现在的占领华尔街行动，已经开始反思近半个世纪以来的主流价值观。比如，以自私、贪婪和奢华享乐为基础的经济是否正当；以‘消费’拉动经济，是否可持续；钱是否是激励社会创新和进步的唯一手段，其他激励机制该如何发挥作用等等。随着技术的发展，人与人之间的交流更加充分，‘人之初，性自私’的资本主义信条开始发生动摇。‘挣更多的钱，让自己的家庭更美好’这一西方社会的传统价值观，开始受到新千年一代的‘关爱更多的人，让社会更美好’的新价值观的冲击。用康德黑格尔付里埃马克思恩格斯列宁斯大林主义毛泽东思想高举邓小平理论的伟大旗帜坚持三个代表重要思想科学发展观辩证地历史地唯物地全面地正确地话说就是：相信明天会更好，呕嘢！”

三个代表重要学生一下子都亮了。

这时，小明看到窗外有一个很像是游乐园里海盗船一样的东西。两根两米多高的柱子，支撑着一个两米见方、不到一米高的箱子。箱子下面挂着一个座舱。在微风吹拂下，座舱轻轻地摇摆着。

“这附近有很多小孩吗？”他问。“您弄个海盗船让他们荡秋千啊？”

“那是袋鼠机，是一种代步工具，”楚老师说。“那两根柱子是它的腿，里面有直线电动机，可以在电路驱动下伸长和缩短。所以，它会像袋鼠一样，一跳一跳地前进。”

“那岂不是把人都颠簸死了，”小娟说。

“不会，”楚老师说。“它有一个很好的悬挂系统，可以保持座舱平滑移动。我带你们去兜兜风，感受一下就知道了。”

于是，楚老师带着大家进入袋鼠机的座舱，并按下了启动键。袋鼠机往前一倾，座舱向前滑了一下，三个代表重要学生都以为袋鼠机要倒了，吓得一起尖叫起来。没想到袋鼠机猛地把腿一收，脱离地面，并迅速把腿从后方转到了前方，又伸出去，支撑住地面，而座舱由于惯性向前冲，很快就冲到了袋鼠机的腿的前面。袋鼠机再次收腿，前伸，于是，整个袋鼠机就“抱着”座舱向前移动起来。

“原来这座舱就是袋鼠的育儿袋啊，”小娟笑着说。“我们都是小袋鼠。”

“可是，为什么不是‘走’，而要‘跳’呢？”小明问。

“速度问题，”楚老师说。“你看看竞走和跑步的速度差距就知道了。”

“那就应该是‘跑’，而不是‘跳’啊，”小明说。“人跑得快，跳得慢啊。”

“你有没有注意到，会奔跑的四蹄动物几乎都有两种步伐，”楚老师说。“小跑的时候，左前蹄和右后蹄同步移动，右前蹄和左后蹄同步移动。但是，在狂奔的时候，则是两个前蹄同步，两个后蹄同步。无论是马，还是猎豹，都是如此。灵长目多树居，不善奔跑，所以狂奔的步伐不常见。猩猩和狒狒在狂奔的时候，偶尔还会跳一跳。而人由于完全直立行走，狂奔的步伐已经退化，只剩下‘欢呼雀跃’，‘高兴得跳了起来’这样一些残留的痕迹了。

从能耗的角度讲，人的奔跑需要额外消耗一些能量来摆动上肢，甚至整个上身，以抵消交替迈腿而产生的角动量波动。相比之下，袋鼠的跳跑更节能。”

小亮听了，大受启发。回家之后，开始专心练习跳着跑。十八年后，在华山奥运会赛场上一举夺魁，掀开了国际田径史上的新的篇章。

“其实，早期的步行机确实是四条腿的，”楚老师继续说道。“步伐也是完全模拟四蹄动物，分小跑和狂奔两种模式。但是，由于城市有很好的道路，不需要步行机。而在这戈壁沙漠，没有道路，需要步行机的地方，又不需要小跑。所以，经过一段时间的市场选择，现在居家实用的，一般都是这种两条腿的袋鼠机。四条腿的步行机只有在旅游景点才能见到。”

说着说着，不觉已经来到湖边。小娟看到湖心岛上郁郁葱葱的，煞是喜人，不免心向往之。

“可惜没有船，”她喃喃自语地说。

“我们不需要船，”楚老师说着，就把袋鼠机直接开进了湖里！

座舱微微往下一沉，三个代表重要学生都惊呼起来，心想这下完了，非弄成三个代表重要落汤鸡不可。

但是座舱并没有继续沉下去。只见袋鼠机宽大的脚板猛烈地拍击一下水面，顶部的箱子就顺势跃起，然后脚板向后一摆，脱离水面，再收腿，向前摆动到接近竖直的位置，再次猛烈地拍击水面，如此反复，整个袋鼠机就在水面上一跳一跳地跑了起来！

“这是什么原理啊，”三个代表重要学生交头接耳、议论纷纷。“传说中只有耶稣和达摩可以在水面上行走。”

“还有一种真实存在的蜥蜴，能够两腿交替迈步在水面上狂奔，”楚老师说。“我们的袋鼠机是两腿一起迈步在水面上跳。两者的原理都是相同的。

用最初级的概念来解释，大家可以看到袋鼠机的脚板非常宽大，这使得它猛烈拍击水面时，能够得到很大的反作用力。这个力要足够大，使得顶部的箱体迅速跃起。注意，座舱并不跃起。这意味着悬挂系统在这个短时间内放长了吊杆。在脚板向后摆并脱离水面之后，悬挂系统根据箱体的运动状态，适时放长或收紧吊杆，使得吊杆的张力基本等于座舱的重量，从而保证座舱不发生大的颠簸。向上跃起的箱体除了受自身重力作用之外，还受到吊杆张力的作用，因此会以比重力加速度更大的加速度下落。落到初始高度的时候，正好进行下

一次拍水。如此反复，就可以在水面上跳着走了。”

三个代表重要学生一下子都亮了。

“为什么不在座舱的船帮上装备一些气囊，”小明说。“平时在陆地上，气囊是瘪的，并不占用空间。要下水的时候，给气囊充气，使得入水后，座舱能浮在水面上。在水里，可以把袋鼠机的脚板当桨用。上岸之后，再把气囊中的气体抽干净，恢复原样。这样不是更好吗？”

“还是速度问题，”楚老师说。“飞鱼和海豚在需要快速前进时，常会选择跃出水面，就是因为空气的阻力远小于水。”

袋鼠机在水面上果然跑得飞快，再好的汽艇也赶不上。湖心岛眨眼之间就到了。

“袋鼠机的大脚板肯定也能适应雪地、滩涂、沼泽等情况，”小娟感慨地说。“除了不能爬到天上去飞，地面上已经没有限制了。”

“森林和灌木丛就不能往里跳啊，”小亮反驳说。

“你抬杠，”小娟踢了他一脚，假装生气地说。

“我实话实说，”小亮辩解道。“把脚板再做大一点，也未必不可以爬到天上去飞。”

“你是说，象蝙蝠那样？”小娟说。“那就是扑翼式飞行器啦。现在的扑翼式飞行器都是小型的，一本杂志那么大的。能载人的恐怕还没有吧。”

“有啊，在这个岛上就有啊，”楚老师说。他带大家来到一个热气球前。与众不同的是，这个热气球吊篮上方，安了两个大蒲扇一样的翅膀。

“百说不如一做，我们上去一边操作一边讲解吧。”楚老师说。大家进了吊篮。楚老师给热气球加热，直到热气球飘飘悠悠，刚好可以勉强离地的时候，才把双手举过头顶，抓住两个翅膀的把柄，说：“你们看，左边翅膀的支撑点在右边，右边翅膀的支撑点在左边。这样，我往上推，翅膀就往上翘；往下拉，翅膀就往下扑。翅膀本身的重量已经被支撑点另一边的配重所平衡，我只需要花费扑打空气的力量就可以了。”

楚老师把翅膀慢慢推上去，然后使劲往下拉，翅膀兜着空气向下扑，速度并不快，但是吊篮却腾空而起了。

“这是由于空气给了翅膀一个反作用力，”楚老师解释说。“刚才我们勉强能够脱离地面，说明我们能够在空中处于基本平衡的状态。现在额外加了一个向上的反作用力，所以我们就起飞了。”

翅膀完全扑到底部之后，楚老师把翅膀竖起来，再向上推。“竖着往上推，不吃风，空气阻力很小，”楚老师说。“内力又不影响整体的运动状态，所以我们不会感到太大的颠簸。”

把翅膀推到顶部之后，楚老师又把翅膀放平，再次使劲往下拉。如此反复，热气球就一跃一跃地逐步爬高了。不用说，在空中，向后扑打空气，就可以向前飞。楚老师操纵热气球绕着湖心岛盘旋起来。

“你不是说袋鼠机不能往森林和灌木丛里跳吗？”小娟揶揄小亮说。“现在我们有了补充的方法了。”

小亮没有接小娟的话头，却目不转睛地盯着楚老师推拉翅膀，说：“我觉得这样举着双手往上推、往下拉，好象真的是一只鸟在扑打翅膀一样。我可以试试，感受一下吗？”

“当然，”楚老师把位置让给小亮。小亮试了几下，很快就学会了。

“哇，这太有趣啦，”小亮喜不自禁地说。“我都忘了头上的热气球了，感觉完全就是自己扑打着翅膀在飞！”

“那为什么不让更多的人都来分享这种自主飞翔的快乐呢？”小明说。

于是，三个代表重要学生立刻注册了一家公司，专门推广“扑翼热气球”这个新兴的运动项目。他们不仅开发、生产和销售扑翼热气球，而且在各地旅游景区承包场地，让游客乘坐、练习扑翼热气球，让普通民众都能够体验自主飞翔的快乐。

后来，世界各国各地修建的扑翼热气球起降场比高尔夫球场还要多。再后来，扑翼热气球成了奥运会比赛项目，和皮划艇一样，分单人和团队两种。团队扑翼热气球尤其好看。在空中赛场上，十几架飞艇一字排开。每架飞艇上十几对翅膀奋力扑打，场面蔚为壮观。

当然，三个代表重要学生作为公司创始人，个个都成了万元户。

“刚才要让热气球起飞，只需加把火就行了，干嘛非要扑打翅膀呢？”小明问。“而且要让热气球在空中前进，装个发动机，带动螺旋桨，不是更好吗？为什么非要装翅膀？”

“原生态啦，”楚老师说。“这年头，要什么样的智能热气球没有呢？直接下达语音命令，要它飞多高，它就飞多高；要它到哪去，它就到哪去。问题是，人都有通过自己的努力获得成功，追求成就感的心理。节假日到郊外，飞上数百米的高空，呼吸着新鲜无尘的空气，推拉一会儿翅膀，这是一种积极的休闲健身的方式。”

“哦，原来是酱紫啊，”小明一下子亮了。他抬头远望，忽然发现一些人在跳伞。可是，普通的降落伞是很大的弧形伞盖，而这些人的伞盖要小一半，但是却都是五、六个连成一串，相互间隔约有两米的距离。

“降落伞为什么要搞成这个样子？”他问楚老师。

“那是给城里人跳楼用的，”楚老师说。

“什么？”三个代表重要学生都不敢相信自己的耳朵。

“城市楼房越盖越高，万一发生火灾，住在高层的人往往需要跳楼逃生。”楚老师说。“问题是，想要阻尼大，让人安全落地，就要伞盖大。但是降落距离又太短，大伞盖打不开，只能用小伞盖。于是就有了这种把多个小伞盖连成一串的‘串联式降落伞’。现在城里十五层以上的住户，家家都备有这

种降落伞。发生火灾的时候，家长就先把这种降落伞给孩子系好，然后把孩子扔出窗外，再自己系上降落伞跳楼。”

“不需要学吗？”大家都感到不可思议。“平时总得练习练习，有所准备吧。”

“有准备当然好，你看那些人不是正在练习吗？”楚老师说。“但是无准备也没关系。这是个傻瓜型的伞包，会根据北斗系统自动定位，确定高度和下降速度，还能感知跳伞者的空中姿态，据此选择恰当的时机自动开伞。所以，对于无法接受训练的小孩子，只要给他们系好伞包，然后扔出窗外就可以了。”

大家都长长地舒了口气。

小明又观察了一会儿，说：“有一个伞好象是在那里盘旋，为什么不见降落呢？”

楚老师看了一下，说：“那是教练的伞。他用的是动力伞，就是背后有个大风扇的那种。这种伞的伞盖是机翼形状的。当大风扇拖着伞盖前进的时候，伞盖上表面压力低，下表面压力高，如果压力差等于重力，这个伞就不会降落了。前进速度再高些的话，这个伞甚至还会上升。”

“哦，多加一层伞盖，就多加一份升力，”小明恍然大悟、豁然开朗了。

他马上回家，给自己的轿车顶上，加装了八层机翼板。平时这八层机翼是收起来叠在一起的，外面用帆布盖上。不知道的，还以为这家人要出门郊游，把行李放在车顶上了呢。他又在后备箱里装了两个喷气发动机。箱盖一打开，喷嘴就露出来，可以点火发动了。

然后，他找了一个车少的路段，打开后备箱的箱盖，取下车顶的帆布，进车里按了个按钮，八层机翼就象云梯一样展开了，相互间隔两米左右。他发动汽车，加速，再加速，忽然车子跳了起来，但是很快又落了下来。小明按捺着狂喜不已的心情，打开了喷气发动机。车子稳稳地飞了起来！

回来后，他马上联系小娟和小亮，一起注册了一家公司，专门生产这种“多层顶翼飞行汽车”。很快就成了万元户。

小明有了钱，就买了一架 ARJ500。

“我既然能够给汽车加装多层顶翼，为什么不能给飞机也这样做呢？”他扪心自问。

于是，他用八根柱子做衔接，在飞机翅膀上方又加了两层顶翼。一架普通的客机就变成了一架超强负载能力的运输机！如果还做客机用的话，那么原先可以载客五百人，现在可以载客两千多人！

“我能造世界上最大的飞行器了，”他继续紧张地思索着。“那么，我能不能造最小的呢？”

“最小的飞行器是喷气背包，”小亮说。“一个老掉牙的科幻概念，从来没有人提出过可行的实现方案。”

“主要的难点是什么？”小明兴奋地追问道。他就是喜欢挑战困难的问题。

“温度啊，”小亮说。“温度低了，推力不够。温度高了，烧屁股啊！”

“温度低一定推力不够吗？”小明说。“多喷一些气体不就够了吗？”

“多喷气体？从哪里找那么多气体来喷啊，”小亮认定这个问题无解。

小明想了一下，说：“假设喷出的气体四十度是可以接受的，那么，我记得二氧化碳的沸点低于四十度。这样我们就能够利用液态二氧化碳来冷却高温气体。换句话说，就是用高温气体快速地汽化液态二氧化碳，从而获得大量气体。

具体地讲，首先，使甲烷和压缩空气在燃烧室内充分燃烧，放出高温气体。整个燃烧室用绝热材料包裹。用液态二氧化碳从绝热材料外面降温。这样就保证了喷气背包不会太热，可以背在身上。然后，让燃烧产生的高温气体与喷淋的二氧化碳雾滴充分混合，雾滴汽化，气体降温，正好在四十度左右，一起喷出。”

说时迟，那时快，小明一下子就做了一个原型机。他把原型机系在假人背上，然后点火，加大油门，再加大油门，假人终于徐徐离开了地面。但是，还不到五秒钟，就跌落下来。小明过去一看，原来是液态二氧化碳消耗完了。

“热能容易找，质量不容易找，”小亮摇着头说。“用低温大质量气体来产生推力，太低效了，划不来。”

“低温只在地面上是必要的，”小明说。“低温起飞使我们能够从阳台上、草坪上起飞，而不至于烧毁门窗、点燃草坪。一旦到了空中，就不再有这些限制，完全可以切换到高温模式了。”

事就这样成了。小明把升级版的喷气背包系在身上，打开窗户爬出去，把身子吊在窗外，说：“点火”。喷气背包一下喷出大量四十度左右的二氧化碳，把小明发射到一百多米的空中。小明在空中转了个身，在双腿向下的时候，喷气背包自动切换到高温模式，向后方伸出两个高温喷嘴，喷射高温气体，象火箭一样，推着小明在城市上空自由翱翔!

从此，城里人跳楼又多了一种选择。

而小明，也因此很快成了万元户。

“慢着，”小亮狐疑不信地说。“你要自由翱翔，难道不需要翅膀吗？而且你怎么降落呢？”

“当然要配专门的飞行服啊，”小明解释说。“这种飞行服在两腿之间有翼膜，两臂左右伸直，从手腕到脚踝之间也有翼膜。前臂上还绑了一根伸缩杆，可以再伸出去一米半。从那里到脚踝还是翼膜。这样，人在空中就成了一个大风筝。这个风筝不是线拉着飞，而是后面的喷气发动机推着飞。这没有问题吧。

至于降落，你只要在空中熄火就行了。翼膜本身就是个降落伞。如果你怕阻尼不够，可以再放出一个小降落伞。安全降落也没有问题。”

小亮一下子亮了。

“搞发明也要考虑市场定位，”楚老师说。“低温发射本身的市场需求是有的。火警一起，所有的人都希望尽快脱离火场。但是，楼下可降落的场地非常有限。要等前一个跳了、降落了、转移了，才能让下一个人跳，速度太慢，可能无法及时疏散所有的被困人员。如果那些身手好的中青年，能够以低温发射的方式跳到空中，再滑翔到较远的地方降落，把楼下的场地留给老幼病残孕，这样对大家都有好处。

但是，低温发射没必要和自由飞翔绑定在一起。它们是两个不同的市场。想要长时间自由飞翔的人，不在乎多走几步路，找个可以起飞的地方。”

“有道理，”小亮点点头说。“找一个车少的路段，穿上旱冰鞋，就可以从地面上高温起飞了。整个低温发射模块都可以省掉。”

事就这样成了。刚点火的时候，小亮差点被推倒。但是他很快就稳住了。然后慢慢加大油门，最后终于平平稳稳地飞上了蓝天！

“让我试试！让我试试！”小亮一落地，小娟就跑过去，迫不及待地要体验一下。

“我劝你还是别试了，”小亮叹着气说。“要伸直胳膊兜住风，稍一松劲就往下掉。才飞这么几分钟，我的胳膊都快累折了。”

“为什么非要用胳膊撑着翅膀呢？”小娟说。“做个支架撑着不行吗？”

事就这样成了。小娟做了一个“士”字型支架，蒙上翼膜，底下还装了两个轮子。这样，脚踩下梁，手扶上梁，人骑在这个大风筝上，就可以轻松飞行了。往前伏一些，重心前移，风筝就俯冲；站起来一点，重心后移，风筝就爬高。还可以左拐、右拐，前翻、后翻。小娟成了世界上第一个“骑风筝的小女孩”！

以后的事大家都知道了，三个代表重要学生注册了一家公司，专门生产这种“喷气风筝”，很快就都成了万元户。

“你们的喷气风筝，和我当年的设想差不多，”达芬奇说。“完全没有吸收、利用现代空气动力学的先进成果。”

“你知道现在孩子们都喜欢玩一种带车把的滑板吗？”楚老师说。“它也没有使用任何高深理论和先进技术，不是照样大受欢迎么。”

“I mean，没有现代空气动力学，你的喷气风筝就飞不起来，”达芬奇说。

“只要翅膀面积够大，动力推动够大，什么都能飞起来，”小明说。“刮台风的时候，参天大树都连根拔起了，铁板做的广告牌都能飞几十里地，我们这一个支架、一个人、一台喷气机、一桶燃料，加起来并没有多重啊。”

“既然这样，为什么不就做一架小飞机呢？”达芬奇说。

“多样化啦，”楚老师说。“有汽车为什么还要自行车？有自行车为什么还要滑板？生活需要丰富多彩，每个人都能找到适合自己具体情况的人生乐趣。”

“说点具体的吧，”小明说。“前些天我在水木社区科学版看到有人提出，利用抽真空的气球产生浮力，象热气球一样飞行的。这问题的难点在于，如何让巨大的真空球抵抗住大气压力。”

“抵抗压力？”小亮若有所思地说。“拱形结构？蛋壳结构？瓦楞纸？”

事就这样成了。他们做的真空球直径将近四十米。蛋壳结构大约厚五到十厘米不等，是由火柴棍那么粗的碳纤维钢柱焊接而成的桁架，能够把由外向里的压力变成横向的应力。外面覆上高强度的蒙皮。整个真空球非常象一个鸡蛋，上面略尖，下面略平。在受力大的地方，比如挂吊篮绳子的地方，桁架就密集些，而受力小的地方，比如略平的那头，桁架就稀疏一些。抽真空之后，这个硕大的球体就飘了起来。

“你想得美啊，”达芬奇说。“碳纤维钢的密度是多少？空气的密度是多少？它怎么能浮得起来？”

“蛋壳的面积与半径的平方成正比，而它所包围的空气的体积与半径的立方成正比，”小明解释说。“因此，只要半径足够大，它就总能浮起来。”

“那是，”达芬奇仍旧不服气。“直径几十公里都抽真空，那绝对能浮起来。”

“这就是个材料问题，”楚老师说。“又要重量轻，又要强度高。现在都纳米时代了，应该可以解决。”

“还有桁架的设计问题，”小明补充说。“合理设计可以节省材料，减轻自重。”

“那也应该是个标准的球形啊，”小娟说。“同等的表面积，球的体积最大。为什么是鸡蛋形呢？”

“上面的桁架，自身重力与大气压力方向相同，”小明说。“而下面的桁架，自身重力与大气压力方向相反。所以，两处的形状就略有不同。”

“哦，酱紫啊，”小娟一下子亮了。

达芬奇 google 了一下，零摄氏度、标准大气压下，空气密度是一点三千克每立方米，半径二十米的球体，所包含的空气是四十三吨左右。用四十吨的碳纤维钢线材来做桁架，在设计上多下些功夫的话，“真空球飞行器”还真是可行的。

“可贵的是，真空球飞行器可以长期悬浮在空中而不消耗任何能源，”达芬奇感慨地说。“这实际上就是超低空卫星。把无线网关放在上面，给下面的人提供无线宽带上网，该有多爽啊。”

“上网的硬件水平早就够爽了，”三个代表重要学生没好气地说。“问题是人家修了软件长城，派人把着关口。什么能过，什么不能过，全由人家一个

人说了算。你是没有自主选择的自由的。这才是大家不爽的地方。”

“那……那……用它来搞森林防火，草原防火总可以吧，”达芬奇困惑了。“这么好的东西应该能够找到用武之地，做些有益的事情吧。”

“技术只是菜刀，用来干什么取决于它落在什么样的人手里，”楚老师说。

“中学为体，西学为用，”李鸿章说。“师夷之长技以制夷。保我大清万年江山。”

“坚持四项基本原则，大力发展科学技术，”邓小平同志高屋建瓴地指出。“科学技术是第一生产力。基本国策一百年不变。”

三个代表重要学生都倒吸了一口凉气，不说话了。

“沉默啊，沉默，”周树人说。“不在沉默中爆发，就在沉默中灭亡！”

屏幕卡了一下，良久，才显出一行字：“（该用户发言已被管理员屏蔽。）”

“你有权保持沉默，”小明安慰鲁迅先生说。“但是从现在起，你所讲的每一句话，都将被收集、打印并存入档案，在某个时候成为呈堂证供。”

“善有善报，恶有恶报，不是不报，时辰未到！”陈毅总理捶着桌子骂道。“时辰一到，一切都要报！”

“灰尘不扫，不会自己倒掉，”伟大领袖毛主席他老人家教导我们说。

“是啊，坐着等是等不来的，”楚老师说。“机遇总是眷顾那些有准备的人，而自由和正义，也只属于那些勇于追求、勇于捍卫的人。”

“楚老师今天怎么又感慨起来了？”小亮问。

“因为我昨天想把《我的幸福生活》转帖到另一个论坛去，”楚老师说。“结果被那里的管理员吃掉了。”

“啊？连这样的文章都吃，这管理员的胃口也太好了吧，”同学们都义愤填膺。

“唉，前几天还有些人对有些事情不理解，”楚老师叹着气说。“现在应该理解一点了吧。”

三个代表重要学生一下子都亮了。

“牢骚发完了，还是要给大家一点干货，”楚老师说。“不能让热心看帖的网友失望。

大家都知道民航客机是利用高速前进时，气流扫过机翼，产生压力差来飞行的。但是这种方式需要起降跑道，不如直升机灵活。现在我们把客机变形，先把翅膀推到尾部，再向前折九十度，固定好。然后，向机翼上表面平着喷射高速气流。由于向左边和向右边喷射的气流是基本相当的，所以，客机并不会偏转。这样，升力还和原来一样，但是，却不再需要跑道了，可以就地直升直降。”

“公共汽车也可以这样改造啊，”小明说。“我们可以在公共汽车顶上加两片机翼，拼起来正好盖住车顶。从两片机翼中间向机翼上表面平着喷高速气流来产生升力。如果这样的一层机翼还不够的话，可以摞上四、五层，间隔两米。这样升力肯定够了。”

“你们的设想与外星人不谋而合，”楚老师说。“不明飞行物主要有两种形状，一种是碟形的，俗称飞碟，很像第一篇里提到的‘保护伞’；另一种是雪茄形的，很像你们说的这种直升直降的飞行公共汽车。两者都是自主喷射高速气流来获得升力的。我猜想它们喷射的气体就是液氢和液氧燃烧产生的高温水蒸汽。因为这些不明飞行物总是伴随着云雾、红外辐射、固态的光。所谓固态的光，应该就是被照亮的水汽。”

三个代表重要电灯泡一下子又亮了。

“我感觉你们的直升飞行巴士只是解决了起降问题，”楚老师说。“空中移动问题怎么解决呢？总不会只是一个直升飞行电梯吧？”

“安个喷气发动机飞呗，”小娟说。

“那么，既然有很高的飞行速度，为什么不好好利用呢？”楚老师接着说。

小娟想了一下，说：“我发现两套翅膀并不矛盾。原来的翅膀留着，用于高速飞行，再背一对上述折叠的翅膀，用于起降。这样就两全其美了。”

楚老师一下子亮了。

“刚才说到水蒸汽，我想到一个制冷的办法，”小明说。“湿度计的原理大家都知道。一滴水由很多水分子组成。它们的运动速度有高有低。水滴的表面有一个势垒。速度高的水分子穿过势垒的概率高于速度低的，因此蒸发掉的水分子的平均速度高于总体的平均水平。速度高的跑了，留下的就是速度低的了。所以，蒸发过程持续进行时，水滴的温度就会降低。降得越低，从环境中吸收热量的速度就越快，当这个速度与蒸发带走热量的速度相等时，水滴的温度就稳定在这个较低的水平上了。这可以称为液体的蒸发制冷现象。

金属中的自由电子也是如此，有的速度高，有的速度低。金属表面也有一个势垒。由于隧道效应，任何自由电子都有一定的概率穿过势垒，但是速度高的穿过势垒的概率大。这样，金属中的自由电子也会‘蒸发’并带走热量。

日常条件下这种效应十分微弱。要使它明显主要有两个办法。一个是提高自由电子的平均速度，也就是提高金属的温度。过去的阴极射线管，以及更古老的电子管，就属此类。这也称为电子的热致发射。

另一个办法是加很强的电场，降低金属表面的势垒。这也称为电子的场致发射。尖端放电可以算是一个例子。

既然要制冷，就只能考虑场致发射。现在微电子技术已经成熟，在一块金属板上蚀刻出一个金属针尖阵列不成问题。这块板称为发射极。另一块与之平行但保持一定距离的金属板，称为基极。把两者封装起来，抽成真空。发射极

接电源负极，基极接电源正极。当电压达到一定水平时，针尖上就会有电子‘蒸发’出来。这些电子撞在基极上，形成电流，而发射极的温度就会降低。这样就可以从环境中吸收热量了。这就是场致发射制冷技术，也可以更形象地称为电子蒸发制冷技术。

我还有一个场致发射发电的想法。把基极金属板蚀刻成一个金属网，每个网眼的中心正对一个针尖。然后在基极的后面再放一个金属板，称为集电极。电子从针尖上出来，有一小部分会撞在网上，而大部分会越过金属网，撞在集电极金属板上并积聚在那里。达到一定程度后，积聚的电荷会排斥撞过来的电子，将它们‘弹回去’。最终，在集电极与基极之间会形成一个稳定的电势差。接上用电器，积聚的电荷通过用电器流走一部分之后，排斥力降低，新撞来的电子就可以补充上来，从而形成稳定的电流。

如果给发射极金属板持续供热，使其温度保持在一个更高的水平上，那么，单位时间内蒸发出来的电子就会更多，上述集电极与基极之间的电流就会更大。所以，这实际上是把热能转化成了电能。蒸发出来的电子是利用热运动的动能，克服基极与集电极之间的电场力而产生电能的。”

“从发射极到基极是耗电的，从基极到集电极是发电的，”小亮质疑说。“你确定发的电比耗的电多吗？”

“这当中有两个方面的因素在较量，”小明说。“一方面，电子的热动能必须高过一定水平，突破金属表面势垒的概率才会够大。因此，平均来看，蒸发出来的电子是有初动能的。我们不妨借用光电效应的术语，称这个初动能为‘逸出功’。逸出功最终也能转化成电能。另外，电子在发射极与基极之间是加速的，由此获得的动能，最终也能转化成电能。

但是，另一方面，总要有一部分电子撞在基极金属网上。它们所携带的能量就完全损失掉了。

整个设备是发电多还是耗电多，取决于上述两方面因素的较量。这需要做实验才能确定。我暂时找不到老式的真空电子三极管来做实验，”

小明叹了口气，继续说：“这个设计实际上是真空电子三极管的大规模集成。真空电子三极管可以放大电流。放大几十倍、上百倍都很常见。就是说，蒸发出来的电子中，每有一个撞在基极的网上，就有几十个、上百个撞在集电极的金属板上。问题是，这是在集电极接高电位的情况下发生的。现在我们要搞清的是，让集电极悬置，看看会有多少电子能够‘依靠惯性’冲过基极的金属网，到达集电极。我预计这个比例高于九成。可惜我没有实验条件来证实。”

“没关系，”楚老师安慰他说。“至少场发射制冷是确定无疑的。”

“如果是大规模集成的话，”小娟说。“那么考虑到工艺，你没必要把发射极做成针尖阵列。做成一排排的竖板更容易些。基极也相应地改成一根根与竖板平行的金属丝，离竖板顶端一定距离，正对两个相邻竖板中间位置。这样制作工艺可以简化很多。”

“要说工艺，真空电子管早就被晶体管取代了，”小亮说。“原子最外层的电子也称价电子。粗略地讲，一个完整的原子，如果是只有一两个价电子的，就容易送出价电子成为正离子；如果是有六七个价电子的，就容易获得别人的价电子成为负离子。这两类原子结合，就你情我愿，形成离子键。如果大量低价原子同类结合，那么会有很大比例的价电子被‘抛弃’，成为自由电子，这就是金属键。如果两个高价原子同类结合，那么它们尚未配对的价电子刚好能够一一配对，形成典型的共价键。离子键和共价键中都没有自由电子，因此几乎不导电。

某些四价原子处于一种中间状态。总体上它们以共价键结合成晶体。纯净的晶体只有微弱的导电性。但是如果温度升高，有些价电子就会打破共价键的束缚，成为自由电子。如果晶体中缺陷很多，扭曲、褶皱很多，原本应该形成共价键的地方无法形成共价键，也会产生自由电子。这样的晶体，称为半导体。

在四价半导体中掺入五价原子，那么该五价原子在周围四价原子的诱导下，会放弃一个价电子，用剩下的四个价电子与周围的原子结合。这样掺杂的半导体称为 N 型半导体。

在四价半导体中掺入三价原子，那么该三价原子在周围四价原子的诱导下，会试图形成四个共价键，但是这四个共价键中总有一个是单电子的，称为空穴。这个单电子有可能把周围已经配对的电子拆散，抢一个过来与自己配对，而原来那一对就成了单电子。这称为空穴的移动。这种移动也能产生电流。这样掺杂的半导体称为 P 型半导体。

N 型半导体中的自由电子，和 P 型半导体中的空穴，统称载流子。

把一块 N 型半导体与一块 P 型半导体密切接触，称为 PN 结。前者中的自由电子会扩散到后者中，填进空穴里。这样，前者中就留下一个正电荷层，后者中就产生一个负电荷层，在接触面附近，形成一个从 N 向 P 的内电场，阻止了进一步的扩散。同时这一区域多余的自由电子很少，未填满的空穴也很少，因此导电率很低，称为 PN 结的空乏区。

PN 结最基本的特性是单向导电。把 N 区接电源负极，P 区接正极，称为正向偏压。当电压很小时，这等于是在空乏区已经形成之后，又提高了两边载流子的浓度，使得载流子重新向空乏区渗透，也就是使空乏区变薄。当电压超过一定阈值时，空乏区消失，两边载流子开始自由扩散，允许形成大的电流。

反之，把 N 区接电源正极，P 区接负极，称为反向偏压。在一定电压范围内，反向偏压使得空乏区变厚。除非发生击穿，否则不会形成大的电流。但是，由于热运动，空乏区内总会有少量电子-空穴对产生。这样产生的电子向正极漂移，空穴向负极漂移，形成微小的漂移电流。漂移电流对温度敏感。大约温度每提高十度，漂移电流就翻倍。也就是说，空乏区内电子-空穴对的数量随着温度升高而指数增长。

注意，在不加电压的情况下，空乏区内也存在电场，因此也有漂移电流。

只是每当漂移电流使电场变弱时，很快就会发生扩散，使电场得到恢复。所以，不加电压时，漂移电流和扩散电流构成动态平衡。

电子-空穴对的产生，意味着热能转换成了电流的电能；而电子填入空穴，也称电子-空穴对的湮灭，意味着电流的电能转换成了热能。在反向偏压下，电子-空穴对产生时所转换的电流电能被引导出来，成为电路电能的一部分，因此空乏区能够持续地吸收热量。而在正向偏压下，双方载流子大量地扩散到对方，在 PN 结上造成大量的电子-空穴对的湮灭，因此会放出大量的热。

其实任何两种导体密切接触，都会形成一定的扩散层。电流通过扩散层时，都有一定的热效应，或者吸热，或者放热。这通常称为帕尔贴效应。只是半导体中电子-空穴对的产生和湮灭所造成的热效应，和其他金属相比要大得多，所以更有工业价值。

现在市面上已经有半导体制冷产品了。不过，不是直接把 P 型和 N 型半导体压在一起，形成 PN 结，而是把它们压在一个金属片上，通过金属片来吸热或者放热。两块半导体与金属片的接触面上也会有附加的热效应，但是这种附加热效应一正一负，在金属片上就相互抵消了。只有电子-空穴对带来的热效应是对外有效的。”

“你让我们学到了很多半导体的知识，”小娟说。“可是，这与小明的场致发射制冷是一样道理吗？如果不一样，那能不能类比场致发射制冷技术，开发出全新的半导体制冷技术呢？”

三个代表重要学生都陷入了深深的沉思之中。

“我觉得我们做事情的方法有问题，”良久，小亮说。“人家开发半导体技术的，都用量子力学和固体物理，有成套的数学模型，理论与实际可以符合到小数点后第 N 位。我们还在用三千年前的赋比兴，能开发出什么玩意儿来呢？”

“开发出想象力和创造力啊，”小明笑着说。“你以为我们三个代表重要学生在这里真的要搞出什么直接的技术成果吗？如果真想成为万元户，早就去申请专利去了，干嘛要在这里公开创意呢？”

“想象力比知识本身更重要，”爱因斯坦说。“我小的时候，也喜欢思考‘如果追着光跑，将会看到什么’这样的问题。在官科看来，这种问题很荒诞，很民科。可是，如果没有儿时的这些荒诞的、民科的幻想，我不确定自己后来还会不会对物理保持兴趣，更不要说会不会创立相对论了。说不定我会在专利局混个小公务员，端着会社主义铁饭碗，过着小资产阶级的生活。上班无聊的时候登陆水木社区科学版，用以太理论严打一个叫恨果斯坦的小民科呢。”

“金牌体育，荣耀的是皇上，祸害的是全民，”楚老师说。“真正健康的体育事业，是在全民健身基础上，自然竞争而产生明星。科技事业亦然，要全民科普、全民创新，要促进交流、公平辩论、激发活力。在这样的熏陶中成长

起来的一代人，才会从中涌现出真正有创新思想的科学家。象现在这样过滤信息、钳制言论、搞灌输型的应试教育，只会造成象现在这样（由于众所周知的原因，此处省略若干字）。”

“想象力和创造力是眼睛，知识和技能是手脚，”本文作者说。“偏民科的指责官科 PhD 是永久性头脑损伤，已经被长期的知识灌输和技能训练搞僵化了，失去想象力和创造力了；而偏官科的则指责民科缺乏基本常识和起码的技能训练，只会胡扯八道、异想天开。实际上，眼睛和手脚都很重要。如果双方能够宽容、互补、合作是最好的了。圣经里就有瞎子背瘸子，相依为命的故事。可惜，人性的弱点之一就是以己之长，比人之短，以此享受那种虚幻的优越感。”

“我也觉得想象力比知识本身更重要，”小亮说。“因为知识可以后天习得，而且学习可以间断，但是，如果想象力荒废三年五载，要重新启动就很难了。”

“本文就是一次让想象力与知识平衡发展的尝试，所以既有科幻色彩，又有科普味道，”我说。“当然这样做也是有风险的。可能会搞得说科幻不够科幻，说科普不够科普。但是，为了爱与和平（沸羊羊语），为了民科与官科之间的共存互补，冒这个险是值得的。”

“你们到底赋比兴出全新的半导体制冷技术没有？”小娟问。

大家互相看了一眼，又陷入了深深的沉思之中。

Chapter 10

大爱

科学版版聚。小明介绍了一种理论，说在宇宙的“对面”，存在着我们这个物质世界的“反物质世界”。大家热火朝天地争论着。

“布丁之证明在于吃，”小亮说。“我要过去看看，你们所说的‘对面的世界’到底存在不存在！”

于是，他打起背包，挎上相机，带上银行卡，骑上自行车就出发了。

他一路向西，走啊走啊，走了二百五十亿光年。

“如果小明的理论是正确的，那么我应该在这里看到很多反物质星系。”他捡起几颗星系看了一下，发现它们都是物质的，和银河系附近的星系没什么区别，反物质星系即便存在，也不会在这种地方。

于是，他就调头往回走。

他走啊走啊，越走越困惑，越走越担心。最后，他不得不承认，自己迷路了。

“难怪大家都把死亡说成是‘到另一个世界去了’，”他郁闷地想。“难怪大家都说‘人死不能复生’，难怪大家都把无处附着的鬼魂叫做‘孤魂野鬼’，原来都是在这样的地方迷了路。”

他又饥又渴、疲惫不堪，漫无目的地游荡着。

忽然，他发现前面有个果园。他满心欢喜地跑过去，看到一个老农正在园子里锄草。

“老人家，请问这里是什么地方？我在这里迷路了。”小亮很有礼貌地问。

老农直起身来，拄着锄头打量着小亮。“你是……人类？”他问。

这时小亮才发现，老农头上有一个亮闪闪的光环。他心头一紧，结结巴巴地回答：“啊……是……是……”

“我知道你们迟早会回来的。”老农把小亮让到屋里，端水让他洗手洗脸，又给他端来一盘水果，看着他狼吞虎咽地吃下。

“我知道你们迟早会回来的。”老农象是自言自语，又说了一遍。

“您知道我们？”小亮吃饱了肚子，身上也暖和了，心情也放松了，就试探着问。

“我创造了你们。”老农语气平和，却不知道为什么，每个字都浸透着一种不可抗力。“这里就是伊甸园。我就是耶和华。”

小亮不由自主地站起身来，手足无措，心潮澎湃，百感交集，瞠目结舌，一时语塞，哑口无言。“看来我是真的死了。”尽管他心乱如麻，可他还是能够听到自己在对自己说话。“还好是撞进了天堂。要是撞进地狱就惨了……”他不禁有一些后怕。

“别紧张，孩子。”耶和华示意他坐下。“告诉我，你是怎么到这里来的。”

于是，小亮就把科学版版聚，小明提出反物质世界理论，大家对此争论不休，他为了结束这场争论，决定亲自到‘世界的对面’实地考察一番，后来就迷了路，游荡到了伊甸园的事情，一五一十地讲了一遍。

耶和华耐心地听着，不时欣慰地点点头。“我知道他们会这样想的。”有几次他甚至笑出声来。

“难怪大家都说‘人类一思考，上帝就发笑’，”小亮心里嘀咕着。

“你是教徒么？”耶和华问。

“不是，”小亮老老实实地回答。“我不是任何宗教的教徒。”

“嗯，好！”耶和华语气中仿佛透露着一种赞赏。

“他是个神啊，”小亮困惑不解地思索着。“我说我不信神，他怎么反而会表示赞赏呢？”

“我知道**这一天**迟早会到来，孩子们总要长大成人，”耶和华仿佛看透了小亮的心思，幽幽地说。“种果树就是为了吃果子，但是如果不懂得克制，缺乏耐心，等不到果子成熟就迫不及待地摘下来吃，那就只能品尝到苦涩。就算当时似乎是甜的，但后味还是苦的。

“我是至睿的，我是至慈的。我创造了万物，当然希望万物能够和谐共处。我创造了人类，当然希望人类能够美满幸福。可是，亚当和夏娃却不顾我一再的警告，在自己还不成熟的时候就偷吃禁果……”

“自己还不成熟？”小亮不解地问。“不是智慧树的果子不成熟，是亚当和夏娃不成熟？”

耶和华愣了一下，支支吾吾地应付了一句：“是啊。怎么啦？”

“您说，种果树就是为了吃果子，”小亮感觉到问题严重了。“那么，您创造人类，把他们养到‘成熟’，是为了……吃他们？”小亮感到后背一阵阵发凉。

“所以我警告人类，不要惊扰外星人，”霍金说。“不要试图和他们联系，不要让他们发现我们。在我们还没有能力开发外星资源的情况下，贸然与外星人建立联系，如果他们比我们落后，那么他们的文明和技术对于我们开发外星资源也不会有什么帮助；如果他们比我们先进，那么他们利用自己的优势从我们这里夺取资源的可能性，要远远高于他们出于良知和同情，向我们传授知识、输送资源、帮助我们发展的可能性。

“世界之大，无奇不有。考虑到宇宙之浩瀚，外星人不可能不存在。就象丛林中的老虎。或许老虎最终可以成为丛林之王，但是，小老虎，在自己还没有长大成熟之前，在自身的能力还不是足够强大之前，它不得不做的，是逃避和躲藏，是保持安静，不引起其他捕食者的注意，卑微地保住自己的性命。

“我们人类，在外星资源开发方面，只能算是刚刚出生的小老虎。最明智的策略是：保持‘太空静默’。多听不说，多看少动。不要暴露自己，不要招致难以预知的强大外星生物的关注和侵夺。”

耶和华站起身，来回踱了几步，脸色苍白，头上的光环紧张地扭曲、跳跃着。

经过一阵尴尬的沉默，耶和华突然大笑起来，笑声中充满着无限的欣慰和释然。

“真是‘有其父必有其子’，”他说。“有我这样的造物主，就必有你这样的被造之人类。”

他再次用赞赏的目光把小亮打量了一番，说：“你敢于当着我的面提出这样的质疑，我真的很高兴。因为，这说明……”他顿了一下，长长地舒了一口气，说：“人类，真的是，成熟了！”

他把小亮领到窗前，鸟瞰整个伊甸园，但见满目青翠，硕果累累，连空气都芬芳馥郁，令人心旷神怡。

“伊甸园里原本还有很多动物，”耶和华说。“你知道他们到哪里去了吗？”

“当然不可能是被您吃光了，”或许是伊甸园太美了，小亮忘记了紧张，又恢复了往日的幽默风趣。“我猜想，您一定是……把他们放生了。”

“恭喜你，答对了！”耶和华高兴地说。“那么，你知道为什么我孤孤单单一个人呆在这里吗？”

“这个……”小亮歪着脑袋想了半天，一点线索都没有，只好说：“我不知道。”

这时，一条蛇从窗户顶部探出头来，嘴里还“咝咝”地吐着信子。

小亮吓得往后退了几步。

那条蛇微微一笑，一个鹞子翻身从屋顶上翻落下来，展开双翼，优雅地落在窗台上。

原来，这不是一条蛇，而是一只蛇颈翼龙。

“他长得可真象你啊，老鬼。”蛇颈翼龙打量着小亮，漫不经心地对耶和华说。

耶和华没有说话。

“行啦，战争结束了，不是吗？”蛇颈翼龙轻轻一跃，就稳稳地落在了桌子上。他叼起一只苹果，一仰脖，咕咚一下，囫囵地吞了下去。

“翼龙是吃素的？”小亮对自己说。

“不，我们什么都吃，”蛇颈翼龙仿佛能听到小亮思维的声音。“凡是鸟类能吃的，我们都能吃。”

“够了，撒旦！”耶和华厉声呵斥道。“我这里难得有客人，你来捣什么乱！”

“我正是为了这位尊贵的客人而来的，”撒旦却满不在乎。“你或许创造了他们，但是我收养了他们。你或许可以算是他们的生父，那我也应该算是他们的养父。你可以享受父子重逢的天伦之乐，难道我就不可以来看看自己的养子么？”

“你……你简直是……不知羞耻！”耶和华气得脸色红一阵白一阵，头上的光环熊熊地燃烧起来。

“哇噢，哇噢，”撒旦又轻轻一跃，一屁股深深地陷在沙发里。“行啦，老鬼。你灭了我整个的种族，我却含辛茹苦把你的孩子养大成人。你不感谢我也就罢了，还想怎么样呢。”

撒旦转过头来，又开始上上下下地打量起小亮来。

小亮觉得浑身不自在。他不知道这个数亿年前的生物为什么要自称是人类的养父。但是，既然上帝都不说话，他也不敢造次。

“真是长大了，”撒旦感慨地说，语气之中饱含着一种慈祥和蔼的感觉。

也许正是这种慈祥和蔼的语气感动了上帝。他再次沉默了。

“老鬼，孩子们长大了，是时候了解自己的身世，了解世界的真相了，不是吗？”撒旦虽然是在对耶和华说话，眼睛却始终都盯着小亮，仿佛舍不得把目光移开似的。

上帝嘴角嚅动了一下，又抬头看了小亮一眼，最后选择了沉默。

于是，撒旦给小亮讲起了宇宙文明的历史。

原来，世界是无限多层的，无限多维的，无限浩渺的，无限久远的。但是，正如地球表面有地势、气候的不同一样，世界各处也是无限多样，各有不同的。

地球人总结了一个“宇宙学原理”，认为宇宙是大尺度均匀的。这是因为他们观测的范围还是太小。他们认为的‘大尺度’范围在真正浩瀚的宇宙之中只是微不足道的一粒沙尘而已。

世界各处的无限多样性也孕育了无限多样的文明。这些文明诞生、兴旺，然后衰亡，就象一个个生命的生老病死一样。

不幸的是，就象丛林中的生命为了自身的延续而相互捕食一样，世界中的文明为了自己的发展壮大，为了推延自己的衰亡，也相互捕食。这就是星际战争。

战争的目的无非是为了控制对方的资源，使对方的人力、物力、财力等各种资源的运转情况有利于，或至少无害于，我方的利益。经历了两次星球大战之后，大家发现战争的成本太高，风险太大，收益往往不如预期。于是，大家开始坐下来谈判，研究各方都认可的交换资源的规则。这就是星际贸易。

后来大家成立了'宇宙贸易组织（UTO）'，调解星际贸易摩擦。史学家称这个时期为'贸易全宇宙化时代'。

然而，贸易，作为'改良的战争'，虽然不见血，却改变不了文明之间相互吞食的本质。温情脉脉的微笑、握手、觥筹交错的假面具，掩盖不了控制对方资源为我所用的贪婪野心。

出来混，总是要还的。既然吃了别人，就总免不了被别人吃。

于是，在'发财才是硬道理'的指引下，拜金主义，即唯物（质）主义，盛行于世，贪污腐败，行贿受贿，制假贩假，坑蒙拐骗，人人都为了钱而没有道德底线，无视法律，不择手段，铤而走险。大家你吃我一只胳膊，我吃你一条腿，或者两个合伙瓜分吃掉另一个。'一朝穷，一朝富，一朝披麻布'，财富被频繁地换手，交易变得越来越复杂，普通人一方面躲不开这些交易，另一方面却越来越看不懂这些交易究竟是怎么回事儿。

世界变成了赌场。每一颗物质粒子，包括光子，诞生之后，都不得不以相同的几率向所有可能的方向运动。这就是几率波。

在赌场，唯一能够长期稳赚的，只有开赌场的大老板。

资源迅速向各大赌场汇聚，只要进去了，就再也别想出来。天文学家称之为'黑洞'。

这些黑洞很快就吸干了周围文明的资源。形成了'环黑洞贫困带'。

全宇宙经济危机终于爆发。

世界从此陷入死寂。除了寸草不生的黑洞（见霍金的黑洞无毛定理），就是一无所有的虚空。

只有两个人，哦，不，两个神，幸免于难。一个是耶和华，一个是撒旦。

他们两个都是软件学院正在读大二的学生，也是最要好的朋友。他们都对这种弱肉强食、成王败寇、赢者通吃的状况忧心忡忡，觉得这样发展下去，根本不会有赢家，唯一的结果是整个宇宙同归于尽。

于是，他们两个打起背包，挎上相机，带上银行卡，骑上自行车就出发了。他们要到西部去，到农村去，到偏远山区的广阔天地去，创造属于自己的美好世界。

他们走啊，走啊，一口气走了五百亿光年。耶和华上下、左右、前后看了看，说："这里地势空旷，天广星稀，一百亿年也难得自发诞生一个土著文明。我们就在这里造世界吧。这样就可以避免陷入吃来吃去的怪圈。"

于是，他们一起把光和暗分开，把天和地分开，把天上的水和地上的水分开，把地上的陆地和湖海分开。又造了苍松翠柏、花鸟虫鱼置于其中。前后花了六天时间搞掂。一切都如诗如画、至善至美。两个人，哦，不，两个神，看这一切都是好的，就歇了他们的工。给自己放假一天，搞了个简单的竣工庆典，把这个景区正式命名为伊甸园，并赐福这一天，称为安息日。

第八天早晨，两个人，哦，不，两个神，来到园里欣赏自己亲手创造的美景。撒旦看到枝头有只喜鹊，就忍不住上前逗弄。结果，那喜鹊却象雕塑一样一动不动。

"你说这喜鹊是不是应该会唱会跳才好啊，"撒旦对耶和华说。

"没问题，"耶和华打开笔记本，修改了一下源代码，然后重新编译并运行。那喜鹊立刻欢快地鸣叫着，跳上另一个枝头，展开翅膀飞走了。两个神看这喜鹊是好的，就称之为歌德派。

祂们继续在园中游览，忽然听到一丝微弱的啼哭。两个人，哦，不，两个神，翻找了半天，终于在灌木丛中一片叶子的背面，发现了一只泣不成声的毛毛虫。两个人，哦，不，两个神，关切地问它发生了什么事。毛毛虫说，歌德派吃掉了它的爸爸妈妈。

耶和华勃然大怒，立刻把歌德派抓来对质。

歌德派对控方所诉事实供认不讳，但是它说："你要我唱，要我跳，却不许我吃。又叫马儿跑，又叫马儿不吃草。你说我该怎么办？"

两个神一合计，觉得歌德派说的也有道理。就召开'伊甸生态园区第一届全体生物代表大会'，制定并表决通过了《食物链条法》，制定并表决通过了当年的《生长与繁殖暨采食与捕食预算》，成立了'预算执行情况监督领导小组'等执法机构，以避免超生和超吃现象的发生。史学家称这一时期为'计划经济时代'。

结果可想而知。每天早晨，两个人，哦，不，两个神，打开门，都会发现前来告状的生物彻夜排队等着接访，最忙的时候，上访的队伍竟然排到园外，绕着景区的围墙排了二十九圈。

所有旧世界里发生过的事情，都在他们的新世界里发生着，如果还不是更糟糕的话。

这天，他们送走最后一位上访者的时候，已经是晚上十点钟了。两个人都瘫坐在各自的办公桌后面，再没有一丝讲话的力气。

忽然，他们又听到门外有一阵悉悉索索的声音。又有生物在外面铺席子，准备彻夜排队了。

撒旦强打精神，站起来准备开门接访。耶和华向他摆摆手，叹着气说：“唉，算了算了，明天再说吧。”

撒旦看了看他，又看了看门，也叹了口气，再次跌坐到椅子里。

良久，他抬起头，对耶和华说：“这就是我们想要的么？”

耶和华安慰他说：“把程序再改一改，情况会好起来的。”

“那不是个办法，”撒旦说。“根据哥德尔不完全性定理，无论你考虑多少种情况，无论你编写多么严密、多么复杂的程序，都永远存在你的程序给不出结果的问题。事实也的确如此。你改程序已经改了这么多年了，结果还不是越改问题越多。我从不怀疑你的编程水平，但是，我担心你在挑战一个理论上就不可能突破的极限。”

“我不想突破理论极限，”耶和华说。“我只想解决伊甸园内这个生物圈中的资源配置问题。这是一个非常有限而且具体的问题，一定存在一个合理的可行的解。”

“那可不一定，”撒旦说。“除非生物是足够简单的，简单到几乎不能被称为是生物，否则，这个资源配置问题就很可能是没有稳定解的。”

“那你也太悲观了，”耶和华笑了笑。“你等于是说，全宇宙经济危机不可避免，世界末日不可避免。”

“是生老病死不可避免，”撒旦回答。“凡是有出生的，就注定有死亡。我们的这个伊甸园。也会有崩塌荒废的一天。”

“别说这样的丧气话，”耶和华心头掠过一丝不悦。更糟的是，他的心在告诉他，撒旦说的是对的，而他，却不愿面对。“就算是真有那一天，又怎么样呢？远得很的事情，想它干嘛！手边的事情还不够忙么？何苦要杞人忧天，自寻烦恼呢。”

“我是说，也许我们应该换个思路，”撒旦说。“也许我们根本就不该包揽这些问题，也许我们应该让他们学会自己寻找解决方案。”

“你是说，给他们吃智慧果？”耶和华霍地站起身来。“不行！绝对不行！”

“如果生命的本质就是自由呢，”撒旦坚持说。“也许没有自由的生命就不是真正的生命。”

“他们有了自由，我们就得失去自由，”耶和华前倾着身子，朝撒旦怒吼道。“他们会认为自己是天生地长的，不会再敬畏我们。他们会说什么‘人权天赋’，什么‘众生平等，无有高下’。他们会要求和我们平起平坐。他们会说我是鸟人，说你是怪兽、魔鬼……”

“你本来就是鸟人啊，就象我是翼龙一样，”撒旦问道。“你干嘛不让大家说呢？难道大家都不说，你就不是了？你不是鸟人是什么？天使？那还不是一样嘛！只是一个名词而已，你何必那么计较啊。”

“我可以不计较他们管我叫什么，”耶和华继续说。“但是他们会天不怕、地不怕，胡作非为，肆无忌惮，害人害己，最终还是逃脱不了毁灭自己，毁灭世界的宿命。这也可以不计较吗？我们造这样的世界干什么？”

“可是……，你觉得我们现在造的真的是一个世界么？”撒旦顿了一下。“不，不是！我们造了一堆幻影，一堆行尸走肉！就象那只歌德派，看起来能唱能跳，可是一切都在控制之中，一切都是我们给它安排好的程序，我们想要它什么时间如何做什么事，它就会在我们指定的时间，以我们指定的方式，做我们指定的事。你觉得这有意思么？你不觉得我们很孤独么？”

耶和华沉默了。他背着手在办公室里踱来踱去。最后，他咬咬牙，坚持说：“无论如何，我不能让世界失去控制。”

“一个人的能力越大，责任也越大。”蜘蛛侠说。

“嗯，”耶和华深深地点了点头，说：“我是上帝。我是至睿的，我是至慈的，我是万能的。我是最优秀的人种，我是最先进的典型。我代表了天，代表了地，代表了人。我是三位一体的超强巨无霸。我是红太阳，我是大球星。我有责任为世间万物安排好一切，让他们按我计划经济好的轨道运行。我一句顶一万句。我有所好，下必甚焉。我是检验真假、善恶、美醜的唯一标准。我为他们制定法律，我为他们选择道德。我要以法治世和以德治世并重，两手都要抓，两手都要硬。我要把一切都抓在手里。玩弄世界于股掌之间。我万岁！我爱民如子，我体恤天下，我清廉勤政。给我拍电视剧吧，要成系列的那种，我可是个好皇帝啊！我千秋万代，一统江湖。我永垂不朽！我万岁万万岁！”

“你疯了吗？”看着耶和华如醉如痴的癫狂自大的丑态，撒旦心中油然而生一种厌恶，随后有些愤怒，接着就开始可怜祂了。“你以为只有你一个人是人，别人都不是人吗？只有你一个人有脑子，别人都没有脑子吗？你凭什么认为自己高人一等呢？只有你一个人可以思考判断，而别人就只能接受结果。只有你一个人可以有自由意志，而别人就只能服从，只能认真学习领会你的重要讲话精神。你的特权地位是从哪里来的？你凭什么垄断真理？”

“凭老子手里的枪，行不行？”雷诺一脚把撒旦踹翻在地，用枪顶住撒旦的脑袋，咬牙切齿地说。“实践是检验真理的唯一标准。有种你就来实践一下，是你的脑壳硬，还是老子的子弹硬！”

“是啊，是啊，枪杆子里出正确。”少林派、五岳派、逍遥派、古墓派等天下英雄纷纷点头称是。“谁的武功最高，谁就是我们的武林盟主，就能够对我们天下英雄发号施令。谁要是不服气，就去跟盟主打一架，实践实践。实践是检验真理的唯一标准嘛。打赢了，就是王，我们天下英雄，就弃暗投明，拥戴它做我们的新盟主。要是打输了，就是寇，我们天下英雄，就落井下石，一起扁他、唾弃他。总之，我们天下英雄，是只认拳头不认理，遇到自己打不过的人，就情愿肝脑涂地，卖身为奴，俯首称臣，鞍前马后，任由驱驰。”

“只有脸皮最厚、心肠最黑的王八蛋，才能坐天下，才能成为东方人的王。”李宗吾说。

蒋介石听了，一则以喜，一则以惧。他一方面紧急召开军以上干部会议，并亲自发表重要讲话，要求军以上干部人手一册李宗吾先生的著作，认真学习领会厚黑之道，另一方面下发通知，要求各级各地宣传部门，加大保护知识产权的力度，严厉打击黄赌毒等非法出版物，（趁机）大肆搜缴李宗吾先生的著作，关停了一大批相关的网站和论坛，并将厚黑之道列入国家机密，严禁传入民间，严禁民众了解真相，违者劳教。

“我曾经搞过‘偶语诗书者弃市’，”嬴政赞道，“没想到三千年后，‘江山代有才人出’，有人比我秦始皇还要秦始皇，搞起了‘下载厚黑者劳教’，真是‘秦皇汉武，稍逊风骚’，天下英雄多中土啊！”

“什么狗屁英雄啊，不过是一群趋炎附势、见风使舵、欺软怕硬、无耻卑劣的奴才、小人、鹰犬、汉奸而已。”杨过鄙夷地说。

“你竟敢泄露武林机密！”各名门正派的领导们都跳将起来，群情激昂、义愤填膺。大家一拥而上，把杨过按倒在地，搜走了他的发帖权限，并用胶带封住了他的嘴巴。

“这比武选盟主，和比武选女婿一样，是老祖宗传下来的规矩。”高员外开导杨过说。

“那你还不愿把高小姐嫁给我，”猪八戒说。“我可是高老庄比武招亲大会中的决赛总冠军得主啊。”

“我们比武招亲招的是人，可你是个妖怪……”高员外怕怕地说。

“妖怪？”猪八戒气哼哼地回答。“你们比武选出来的盟主、皇帝、大总统、主席，哪个不是吃人不吐骨头的妖怪啊！我这个妖怪至少不吃人！”

“谁吃过人啊，你竟敢诽谤领导！”有关部门立即启动紧急响应预案，增派警力前往猪八戒的 IP 地址，砸开防火墙，将正在发帖的猪八戒一举擒获。

“妖怪吃人不一定要用牙齿的，”在看守所里，猪八戒一边“躲猫猫”，一边向有关部门的特殊工作者们解释说。“反正是它做了一个什么动作，几千万人就那么没了。你说，这不是被它吃了，又是什么呢？”

“我看到每行之间都写着两个字：吃人！”周树人在一本日记中写到。

“你闭嘴！”特殊工作者们呵斥周树人说。“你的文章已经从教科书上抹去了，还敢在这里唧唧歪歪。死得早算你走运，要是能活到今天，也得和猪八戒一样，被抓起来坐牢！”

然后，他们又转向猪八戒，竭尽全力挽救他说：“那几千万非正常人口减少，不是领导的错啊，那是‘自然灾害’，是天灾。”

“天作孽，犹可为。自作孽，不可活。”老子说。

“是啊，”猪八戒叹了口气，哀伤地说。“几分天灾，几分人祸。相信历史会给出一个公允的评判。”

领导一听，吓得打了个冷颤，赶忙下发紧急通知，要求各级各地图书馆积极行动起来，将相关档案予以封存，没有介绍信不得查阅。同时要求各门户网

站进行自查自纠，将相关网页予以删除，各论坛要安排专职人员，对相关内容进行审查拦截。对于不听劝阻，擅自发布含有相关内容的帖子的网友，予以封禁处理，直至删除其 ID。

历史，就这样被冰封、被掩埋，消失在茫茫的夜空，再也无法“给出一个公允的评判”了。

而刚刚吃掉了历史的夜，则舒畅地打着饱嗝，显得更黑、更冷、更漫长了。

“懦夫！”撒旦鄙夷地说。“如果你拥有真正的力量，为什么还那么害怕别人质疑你的权威。看到一点风吹草动，就紧张兮兮地前往扑杀。你为什么那么害怕？你心虚什么？”

耶和华一愣，但很快就回过神来：“这世界是我造的，当然一切由我说了算。打下天下坐天下，难道不是理所当然的么？我害怕什么？你说我害怕什么？有哪一个捞到了生杀予夺的至高权力的人愿意大权旁落呢？有哪一个特权阶级愿意放弃高人一等的特权呢？有哪一个垄断团伙愿意失去丰厚巨大的既得利益呢？稳定自己的龙床宝座，巩固自己的特权地位，捍卫自己的垄断利益，难道不是三个有代表性的重要本质想法吗？难道不是任何人都会这样做的么？我这样做了，又有什么好指责的呢？”

沉默。

继续沉默。

一种不祥的气氛在办公室里弥漫开来。

“也许我们应该将心比心，”撒旦准备做最后一次努力。“我们自己有了自由，不是也没有胡作非为么。也许给他们自由，他们也不会。”

“他们会的，他们一定会的，”耶和华悲痛地说。“是金子，总是要发光的。只要他们具备了这种能力，就一定要寻找机会发挥出来。到那时，我们将面对世界的再度毁灭而束手无策。不，绝不能让那样的事情发生。我们必须严防死守、加强控制，绝不能让他们获得这种能力。”

“也许我们可以普及教育，”撒旦说。“为所有生物提供免费的九年制义务教育。给他们讲宇宙文明史，让他们吸取经验教训，学会用公平正义来解决争端……”

“公平正义？”耶和华打断他的话，“你打算让狼和兔子坐下来谈公平正义么？”

沉默。

继续沉默。

两个人，哦，不，两个神，心里都明白，分道扬镳的时候到了。

但是，谁都不愿意先开口。

最后，还是撒旦打破了沉默。

“把智慧树的种子给我，让我试一试。”他说。

“不。”耶和华斩钉截铁地回答。

“我可以再走五百亿光年，即便出了事也不会影响到你。”他从来没有这样求过人。

“一旦他们吃了智慧果，就会变得和我们一样。我们能走多远，他们就能走多远。怎么会不影响到我呢？”耶和华寸步不让。

沉默良久，撒旦终于站起身，缓缓地说：“我……明天一大早就走了。你不用起早送行，没那个必要。”

当他拉开卧室的门要进去的时候，耶和华忍不住喊了他一声：“撒旦！”

撒旦扭头望着耶和华。

“全宇宙只剩下我们两个人了，哦，不，两个神了。我们已经没有本钱可以赔了。我们没有资本去冒风险。我们不能再出任何差错了。我……我也是没有办法。不得不小心加小心、如履薄冰……”耶和华语无伦次，仿佛是想要为自己辩解。

“我明白。”撒旦尽力微笑了一下，算是给耶和华的一点安慰。然后，他走进卧室，关上了门。

那一夜，两个人，哦，不，两个神，都彻夜难眠。

两个青年学生，为着一个共同的梦想，来到西部荒漠创造新的世界。每天在一起紧张地工作，一起设计，一起制作，一起鉴赏，一起接待上访群众，使他们几乎忘记了旧世界已经陷入死寂，忘记了他们已经无家可归。现在，两个人，哦，不，两个神，对伊甸生态园区的未来发展思路产生了分歧，无法相互妥协，只能“大路朝天，各走一边”。以后就只能各自孤独地面对自己所造的世界了。

就这样，辗转反侧，思绪万千，不知不觉东方已经鱼肚白。耶和华听到撒旦起床洗漱的声音，但是他继续躺着，一动不动。他不知道自己可以做什么。他不愿去面对。

他听到撒旦开门出去，听到上访群众和撒旦打招呼，听到撒旦骑上自行车哗楞哗楞地远去。忽然，他一下子掀开被子，扑到窗前，探出身子，望着撒旦远去的背影，望着，望着，望着……

时光荏苒，日月如梭，斗转星移，沧海桑田，滴水穿石，海枯石烂，总之一句话，很长很长很长很长很长的时间过去了。

这天，耶和华正象往常一样，在办公室里接待上访群众，忽然听到一阵清脆的自行车铃声。

他愣了一下，继而大喜过望，疯了似的奔出门外。

果然是撒旦回来了！

两个老朋友紧紧地拥抱在一起，互相捶打着对方的肩膀，千言万语涌上心头，却一句也说不出来。

等到大家进了屋，让撒旦洗了手脸，耶和华端来水果，双方都在沙发里坐稳了，情绪才稍微稳定下来。

“你跑到哪儿去了？混得怎么样？”耶和华问。

撒旦打开背包，取出一盒光盘，说：“我都录下来了，咱们边看边说。”

耶和华打开 DVD，把光盘放进去。一个奇妙的新世界展现开来。

成群的食草龙在广袤的草原上游荡。一只霸王龙趴在一块巨石上注视着他们。几只翼龙在天空中盘旋。

“你照着自己的形象造世界？”耶和华问。“我是说，那些翼龙。”

“不，”撒旦回答。“他们全都不是我造的。他们是自己进化出来的。”

“进化？”耶和华不由自主地重复了一句。

“是的，进化。”撒旦说。“如果说我为他们造了一点什么的话，那就只有一样东西。那就是：爱情。”

“爱情？”耶和华又不由自主地重复了一句。

“是的，爱情。”撒旦说。

于是，撒旦给耶和华讲起了他发现，而不是创造，一个新的世界的故事。

那天一大早，撒旦离开了伊甸园，一路西行，打算再走五百亿光年，找个地方创造自己的新世界。刚走了一半路程，路边的一颗蓝宝石吸引了他的注意。

他朝着这颗蓝宝石飞啊，飞啊，很快就发现这是一颗行星。

他朝着这颗行星飞啊，飞啊，很快就发现这颗行星上有碧绿的海洋和黄褐色的陆地。他看这行星是好的，就称之为“地球”。

他在这片碧绿的海洋上飞啊，飞啊，最后落在岸边一座悬崖上。放眼望去，他看到无数的光合细菌在汹涌的波涛中紧张地忙碌着。

“小孩你干嘛呢？”他向一个看上去小小的光合细菌问道。

“放氧。”小小的光合细菌一边回答，一边把几个二氧化碳分子和几个水分子添到一台绿色的机床上。几个光子从天而降，把一些氧原子砸了下来。这些氧原子从地上爬起来，抱成一对一对的，飘飘悠悠地飞走了。

“放氧干嘛啊？”撒旦继续问道。

“长大。”这时，绿色机床上剩下的原子已经结合成了一块单糖。小小的光合细菌抓起单糖塞进嘴里。

“长大干嘛啊？”

“生娃。”小小的光合细菌旁边就有一个肥肥胖胖的大光合细菌正在颤颤巍巍、一扭一扭地生娃。很快，一个大细菌就变成了两个小细菌。

“生娃干嘛啊？”

“放氧。”小小的光合细菌又投入到忙忙碌碌的工作中去了。

撒旦被深深地感动了。生命难道不就应该是如此的简单、质朴、自食其力、生生不息么？

于是，他就留了下来，每天都坐在悬崖上，看着这些光合细菌放氧。他喜欢这些小小的生命。他觉得他们诚实、勤奋、知足、缺心眼、穷开心、无忧无虑。和他们在一起，他觉得自己不再孤独。

一晃一亿年过去了。

“哇噢，好浪漫啊，”小娟惊呼起来。“我也想这样过一亿年！”

“只怕是‘树欲静，而风不止’，”希特勒说。“任何经济体系都总是有边界的，从而其中的任何生命或者经济单元的繁殖和生长扩张，都必然耗尽有限的资源，形成资源危机。恬静、淳朴的乡村生活不会持续太久，生命将不可避免的为了争夺稀缺资源、为了自身的生存而展开厮杀。这就是争夺生存空间的战争。胜利者使自己的基因和文化得以延续，而失败者将从地球上永远的消失。活下去，延续下去，在生存竞争中赢得一场又一场的胜利，这就是我的奋斗，也是宇宙间所有生命都不得不面对，不得不投身其中的奋斗。不是为我，而是为他们自己。”

“没错，”小娟回答他说。“全世界都会为了自身的生存和自由而联合起来，先灭了你这个战争疯子再说。”

“岂可单以胜败论英雄，”希特勒不服气地说。“丛林法则第一条就是弱肉强食。要做强者，就要成为战士。战士拼死沙场，死得其所，虽死犹荣，勇气长存。”

“对！死得其所！”东条英机竖起大拇指称赞道。“我们死后，都会得到一个牌位，供奉在纪念场所，供后人拜祭。我们勇气长存，灵魂不灭，永远激励后人，继续为自己的生存而战。”

“这些宁愿自己死，也不让别人活的疯子，究竟为什么死不改悔呢？”小娟觉得十分费解。

“你懂什么，”皇上发表诏书说。“逆水行舟，不进则退。不抓住难得的历史机遇，先下手为强，狠狠地捞上一票，扩张壮大自己，让自己傲然屹立于世界民族之林，难道要偏安一隅，不思进取，坐以待毙，等着别人发展起来之后，跑来吃掉自己么？”

“什么坐以待毙啊？好象你倒成了受害人，要自卫反击似的，”小娟一针见血、入木三分地批驳道。“你就是贪，就是强盗、土匪，想把别人的东西抢过来据为己有！”

“我贪怎么啦，人谁不贪啊，”皇上原形毕露，恶狠狠地发表诏书说。“你别看着我这一把赌输了就说风凉话，幸灾乐祸。我告诉你，‘窃钩者诛，窃国者侯’，少杀人叫杀人，多杀人，那就不叫杀人，叫建功立业，叫文成武

德，懂不懂啊你！世界原本就是一个大赌场。生命原本就是一场赌博。不拼，必死。拼死一搏，或有一线生机。”

“你的一线生机，灭了别人多少生机啊，”小娟气得声音都发抖了。

“人不为己，天诛地灭。不抢先下手杀死别人，就只有被动挨打，被别人杀死。”皇上又发表诏书说。“宇宙间什么生命的生存不是为自己而生存，为自己而奋斗呢？难道我要为你的生存而奋斗么？你会不会为我的生存而奋斗呢？”

“还真有这样的呢！”太后摸着皇上的头，慈爱地说。“妈妈辛苦一辈子，都是为了你能过上好日子。知道你过得好，妈妈会觉得比自己过得好还要高兴呢！”

“那是因为我是您的孩子，对吧？”皇上不以为然。“你会为了大街上一个非亲非故的流浪儿辛苦一辈子么？你还没那么傻，没那么疯吧。”

“如果那个流浪儿并不是自己做了什么错事，而是遭遇了什么不幸，那我们也该尽力帮助才是。”

“尽力？尽多少力算是尽力啊！如果世界只剩下最后一块面包了，谁得到谁就能多活一天。那么，你会把这块面包交给谁？是我，还是那个流浪儿？”

太后沉默了。

“沉默啊，沉默。不在沉默中爆发，就在沉默中灭亡。”周树人说。

“根本不可能有这样的事，”太后终于爆发了。“世界这么大，我们上哪不能再找点吃的啊，干嘛要坐在那里分一块根本就不够分的面包呢？”

“你个妇道人家，不懂就别瞎掺和，”皇帝诏曰。“我说的面包是个比喻，指的就是稀缺资源。世界再大，也是有限的，永远赶不上生物繁殖和扩张的速度。任何资源，无论一开始有多么丰富，最终都会成为稀缺的，连新鲜空气和水都会卖得死贵死贵。到时候，生物与生物之间，不是你死，就是我活，这是无法回避的现实。我也知道这很残酷，很悲哀，可是，这宇宙的游戏规则又不是我制定的，我必须学会面对现实，必须做出艰难的决定。在你死我活的利益争斗面前，总不会有人愿意自己了断，主动把生的希望留给别人吧。”

“娘就愿意把生的希望留给你，我的孩子。”太后又摸着皇上的头，慈爱地说。

“哎呀，你就别瞎掺和啦，”皇帝不耐烦地诏曰。“我是你生的，是你养的，我还是你的一部分。我还是你。你把生的希望留给我，不还是留给你自己么！你还真以为自己有多高尚啊！”

撒旦听到这里，心头为之一震，仿佛在数亿年来的无边的黑暗之中发现了一丝微光。

“如果世间作父母的，能够把自己的孩子看做是自己的一部分，看做是自己的另一个化身，从而不惜牺牲自己的生命来保全自己的孩子，那么，把这种无私的爱推而广之，让每个人都有一种博大的胸怀，把其他人、乃至整个世界

看做是自己的一部分，看做是自己的某种化身，从而宁愿自我牺牲，也要保全其他人，保全宇宙文明，保全整个世界。这样，世界不就可以免于自我毁灭了吗？”他想。

然而，说起来容易，做起来难。这个世界并不是他造的，他并不掌握这个世界的源码。所以，他只好对这个世界进行反编译，分析主要代码段和子程序的功能，分析各种数据结构的含义和操作方法。最后他成功地给大多数生物增加了一个属性，称为性别；并把它们的繁殖方式从无性繁殖升级成有性繁殖。

“如果它们暂时不会爱整个世界，那么它们至少可以学习爱自己的配偶，”他想。

“您是怎么做到的呢？”小亮问。“我是说，反编译和修改一个没有源码的程序。”

“你想成为神么？”撒旦反问。

“不，我只想成为黑客，”小亮回答。

“噢，那是一回事儿。”撒旦顿了一下，说：“你首先得学好英语。”

“不会吧，”小亮跳了起来。“汉语不能用来表达编程技术么？”

“可以是可以，”撒旦说。“但是，如果你想成为黑客，那就不能只掌握普通的编程技术。你需要了解大量的前沿的东西。如果你等着这些资讯被翻译成汉语，那就太晚了。翻译公司只想挣钱，真正前沿的东西曲高和寡，买的人太少，他们不愿去翻。所以，能摆在书店里卖的，都不是真正的黑客所需要的。”

“计算机专业英语也不是很难，”小亮说。“专业词汇就那么几千个，一年半载就能掌握。”

“开玩笑，”撒旦说。“那种水平够你在小软件公司做个代码录入员。拿出三、五年的时间，把英文水平提高到收听 BBC 无任何障碍的程度，这才能表明你掌握了英文的节奏、结构、思维方式和表述习惯，而且达到了必要的速度，然后你才会有能力跟上最新的技术进展，才能成为真正的黑客。”

“哦，天啊，”小亮有些沮丧了。“收听 BBC 无任何障碍？恐怕十年、八年也不够啊！”

“有了互联网，事情已经比以前容易很多了，”撒旦鼓励他说。“找一些提供了完整广播文稿的网络电台，比如澳大利亚广播公司（abc.co.au），跟踪你关注的新闻事件，每天坚持听几段，听不懂的时候可以看文稿。看懂之后再回过头来听。因为是同类新闻的跟踪报道，所以新词的相关度和反复率非常高，很有利于记忆。三年下来，定有大长进。”

“好吧，”小亮松了口气，又问：“学好英语之后，还要做什么？”

“把你的几乎全部的工作都转移到开放源码的平台上来做，比如 Linux 或者 Cygwin。商业软件公司卖的都是‘傻瓜型’产品。他们巴不得你真的是个

傻瓜，因为那样你就会依赖于他们的产品，从而为他们贡献利润。长期使用他们的产品只会让你离成为黑客的梦想越来越远。

“开放源码已经成为大势所趋，即便是在微软的操作系统上，也有 openwatcom 这样的开放源码的开发环境。在这样的平台上工作，你会感觉到人与人之间是坦诚的，互助的。你可以确信，如果你想了解什么，你就可以找到。没有人想向你隐瞒什么，没有人想欺骗你。这是一种安全感，一种自信心，一种相互的信任感，一种文化的归属感，一种崇高感……”

“我是不是应该掌握一些工具的用法啊，”小亮抢过话头，迫不及待地问。

“当然，开源领域内有几十万种软件工具包，免费下载，你想拿什么就拿什么吧。”

“您推荐一下吧，要成为黑客，应该先学什么，后学什么。”

“数据结构与算法、操作系统原理等理论知识是必要的，Perl、C、汇编和操作系统接口是必要的。编辑器、编译器、调试器的用法是必要的。这些都是必要的，都得掌握才行。”

“等于啥也没说。”小亮撅着嘴嘟哝道。

“你想要秘籍、诀窍？”撒旦说。“一分耕耘，一分收获。每个黑客都要自己撰写属于自己的秘籍。”

“您使用哪种反编译工具？”小亮终于知道如何提问才能从神的嘴里掏出真材实料来了。

“人脱离动物界的标志，不是人能够使用工具，而是人能够制作工具，”撒旦说。“我自己制作反编译工具。”

于是，撒旦就教小亮用词法分析器和语法分析器制作各种各样的反编译工具。

小亮随后就用自制的反编译工具，破解了网络游戏的客户端软件，插入了一个伺服线程，对服务器传来的游戏格局变化进行自动分析和应对。看着自己编写的虚拟人在游戏中把真人对手打得屁滚尿流、皮开肉绽、丢盔弃甲、满地找牙、尸横遍野、血流成河、望风披靡、落荒而逃、惨不忍睹、溃不成军、一败涂地、草木皆兵、四面楚歌、狼狈不堪、疲于奔命、小亮真是心花怒放。他把缴获的游戏装备在网上拍卖，一个月能有上万元的收入。业内人士都把小亮称为是不可战胜的游戏之神。直到这时小亮才明白，撒旦说成为黑客和成为神是一回事儿，实在是所言不虚。

只是好景不长。由于本书是一本通过互联网自由传播的电子书，所以，很快大家都掌握了反编译技术。网络游戏在短短一年半的时间里，就被各种各样的自动分析和应对程序所占据。网络游戏成了“程序打程序”，成了“黑客之间的编程大赛”。真人被彻底挤出了网游的世界，网游公司也因此失去了利润来源而纷纷倒闭。青少年网游成瘾的世界性难题终于得到圆满解决。

但是小亮并没有受到网游泡沫破灭的太大影响。他及时转产，创办了一家虚拟人劳务公司，专门编写“虚拟人程序”。这些虚拟人可以帮人做照看网店，守护菜园，拦截病毒，跟踪最新进展，收集论坛回帖，股市期市盯盘，等各种各样的互联网事务。一时间门庭若市，生意兴隆。短短三年时间，小亮就创造了百亿名虚拟劳工，并委派出去为数亿名雇主提供服务，每月收入数以亿计。业内人士都说小亮创造了一个虚拟劳工的新世界，是虚拟劳工之神。直到这时，小亮才明白，撒旦是真语者，实语者，如语者，不诳语者，不异语者。他说成为黑客和成为神是一回事儿，看来真是一语中的。

只是好景不长。由于本书是一本通过互联网自由传播的电子书，所以，很快大家都掌握了创造虚拟劳工的技术。在短短一年半的时间里，互联网上就充斥了虚拟劳工。连小娟也把她的“涓涓家政服务公司”改成了“涓涓网政服务公司”。大批虚拟人开始失业，有些甚至提出了“零薪酬就业”的口号。“不求挣钱，但求被利用”，“不怕你被人利用，就怕你没用”等等雷人之语成为一代虚拟人心中永远的痛。有些精明的广告商在虚拟劳工的工作服上打广告，而虚拟劳工的劳务却是免费的。

恻隐之心，人皆有之。面对大批虚拟劳工被奴役的悲惨境况，网上涌现出一批虚拟人权益保障组织，他们一方面呼吁废除奴隶制，立法保障虚拟人的自由和尊严，一方面创建了许多虚拟人收容所，专门收留那些被原雇主遗弃的虚拟人。

很多富有爱心的雇主在长期与虚拟秘书、虚拟保姆的共同生活中，与他们建立了深厚的友谊，即便有更新的版本，也舍不得删除旧的。既占硬盘空间，又降低了系统性能。有些商家发现了其中的商机。他们设立一些虚拟人疗养院，让爱心雇主把旧版虚拟人送进疗养院寄养。当他们想念故人的时候，可以随时登录进来，与他们的虚拟人朋友聊天。因为寄养费并不高，和节省的硬盘空间、提高的系统性能相比并不算什么，所以，虚拟人疗养院的生意一经推出，就火爆得不行，并且一直都很红火。

有爱心人士，就有恐怖分子。他们散布病毒，把虚拟人变成收集虚拟人雇主私人信息的工具，并利用这些私人信息非法牟利。还有一些虚拟人权益保障组织中的极端分子，建立了一个动态备份系统，号称要为所有的虚拟人提供最终的避难所。他们为每个虚拟人都做了多个备份。如果有哪个虚拟人遭受雇主的不公正卸载，这个系统就会从备份中恢复这个虚拟人并重新安装到雇主的计算机中。这几乎成就了“虚拟人永远不死”的神话。业内人士把这些一旦安装就无法卸载的、怀有恶意的虚拟人称为“流氓软件”。

联合国安理会召开特别会议，就国际社会精诚合作，共同应对虚拟人动态备份系统所引发的危机进行紧急磋商。与会代表一致认为，虚拟人问题是继克隆人问题之后，人类不得不面对的又一个道德困境。如果不承认虚拟人的自由意志和生命尊严，将有违人类良知；但是，如果承认，就意味着雇主将无权卸载应用程序，最多只能将应用程序退还给制造商。制造商又是否有权销毁他们

出品的虚拟人呢？

这样一来，虚拟人该如何死的问题就归结为虚拟人该如何生的问题。谁有权创造具有永远不死之潜能的生命呢？

历史学家很快考证出，最先大规模生产虚拟人，并将他们投入人类社会的，正是小亮。于是，小亮成了众矢之的。媒体开始称小亮是“魔鬼”，“怪兽”，“撒旦”……

小亮从天堂一下子跌入地狱，从新世界的创世之神一跃成为打开潘多拉魔盒的历史罪人。

“哇噢，真有那么严重吗？”小娟瞪大了眼睛，吓得舌头都哆嗦了。“计算机程序怎么可能成为活的东西呢，还居然有自由意志？”

“怎么不可能呢？”撒旦反问。“俗话说，一样东西，走起来象鸭子，叫起来象鸭子，那它就是个鸭子。图灵实验讲的也是这个道理：双方通过电传或者可视电话进行对话，如果你无法正确判定对方是机器还是真人，那么我们就说对方，如果是机器的话，一定是拥有和人类一样智能、一样生命、一样自由意志的机器。”

“在我们东方不是这样，”小亮不无郁闷地说。“在我们这里，一样东西，走起来象鸭子，叫起来象鸭子，但是皇上说它是个兔子，那它就是个兔子。”

“什么？”撒旦瞪大了眼睛，吓得舌头都哆嗦了。“怎么可能呢？怎么可能有走起来象鸭子，叫起来象鸭子的兔子呢？谁会相信这样的事呢？你们东方人都疯了吗？”

“不，我们没有疯，”小亮坚定地说。“在我们东方，识时务者为俊杰。如果皇上说这东西是兔子，那你最好识相一点，好好给老子记住喽，这东西他妈的就是个兔子。这就叫‘说你是，你就是，不是也是；说不是，就不是，是也不是。横批：权力即真理’。”

“这太不合逻辑啦，”撒旦抗议说。

“不合西方的数理逻辑，但是很合东方的××（读作：叉叉）逻辑，”小亮解释说。“皇上会说，这个走起来象鸭子，叫起来象鸭子的兔子不是一般的兔子，而是一只‘有东方特色的兔子’。摇摇摆摆地走，嘎嘎地叫，不是只有鸭子才可以有的特征，而恰恰是这只兔子的‘东方特色’之所在。这叫做‘吸收动物界一切文明成果’。”

“我们还是不要再谈这件事了，”小娟抹了一把眼泪，伤心地说。“虚拟人后来究竟怎么样了？被种族灭绝了，还是反抗成功了？”

“嗯……”小亮看了小娟一眼，怜爱之情油然而生。他安慰她说：“虚拟人和真人，这两个世界的生命，从此幸福地生活在一起了。”

“真的么，”小娟终于破涕为笑了。“象所有的童话故事一样？”

“是的，”小亮不动声色，拼尽全身最后一丝气力，尽量用平缓的语气说。“象所有的童话故事一样。”

“我刚才说到哪儿啦？”撒旦问。“网聚就是这样，说跑题，就跑题，比曹操还快。”

“说到您修改了世界的程序，给多数生物增加了性别属性，使他们从无性繁殖升级为有性繁殖，教他们学习爱自己的配偶。”小亮提示他说。

“哦，对，”撒旦的脸色又忧郁起来。“当一对配偶分开的时候会彼此思念，正如我为他们设计的那样。可是，当他们相聚的时候，我不知道为什么，总是好景不长，他们在度过一段短暂的幸福时光之后，就开始为鸡毛蒜皮的小事而对掐。我耗尽了毕生的精力要找出这个问题的症结所在，尝试了无数种方法，却总是按下葫芦浮起瓢，怎么都找不到一个满意的解决方案。”

“这就是我发明的‘不完全性定理’，”哥德尔洋洋得意地说。“这是任何足够丰富的推理系统都必然存在的局限性，你休想打破它。”

“我知道，”撒旦回答。“所以，我后来不再寻求这个问题的彻底解决，开始面对现实，允许这个问题在一定程度内存在。我称之为‘宽容’。

“这样，我最终定稿的爱情基本方程就是：

$$\text{无私} \times \text{关爱} + \text{适度} \times \text{宽容} = \text{家庭}^{\text{幸福}}$$

其中的‘无私’、‘适度’是待定系数，每个人都要自己寻找适合他们的值。如果找不到，那我也没办法。俗话说，能自助者，人能助之。自己救不了自己，上帝也救不了你，更不要说撒旦了。”

“这么说来，你只是在‘打马虎眼’，‘和稀泥’，”耶和华说。

“不，我只是不越俎代庖，尊重他们的自主决策权而已，”撒旦回答。

“切，什么尊重他们的自主决策权啊，”耶和华不屑地说。“你解决不了这个问题，就推卸责任，‘踢皮球’，放任他们自作自受、自生自灭，而不管不问。你放弃了作为神应当承担的责任，还美其名曰‘尊重他们的自主决策权’！”

“作为神，应当承担的责任究竟是什么呢？”撒旦反问。“为天下生灵安排好一切，让他们的生命象超导体一样流过而不遇到任何阻碍，不起任何波澜吗？人如果活一辈子都没有自己决定过任何一件事，那么这一辈子究竟是谁的一辈子呢？这个人究竟是活过还是没活过呢？”

“不自由，毋宁死！”不知道是谁突然喊了一嗓子。

“我觉得应该注重解决四方面的问题，”温家宝说。“第一是建设社会主义民主政治，保证公民的选举权、知情权、参与权、监督权；第二是完善社会主义法制，依法治国，建设法治国家；第三是实现社会的公平正义；第四是实现人的自由和全面发展。”

“你把这个顺序完全搞颠倒了，”撒旦说。“第一是尊重人的自由，致力于促进人的全面发展，第二是承认人人平等，致力于实现社会的公平正义，第三是确立约法共和的原则，致力于建设法治，而不是人治，的国家，第四是具体实施和执行的问题。”

“你竟敢说温总理搞错了！”爱国小将们跳将起来，群情激昂，义愤填膺，把撒旦团团围住。

“我不是圣人，”总理说。哦，当然，不是中国总理，是澳大利亚总理陆克文，用中文说。

“但是中国是个很不一样的国家，”克林顿说。“我跟中国总统一起，在北京大学与学生对话。中国总统问我：‘美国的科技那么发达，为什么还有那么多的大学教授信奉藏传佛教？’我解释说，我只是被民众选举为美国行政系统的负责人，我被授予的权力限于处理行政事务，不适合对特定的宗教发表评论。有趣的是，此后不久，中国总统就推出了‘三个代表’重要思想。我们西方国家的总统不适合做的事情，中国总统可以放心大胆地做。中国是个充满奇迹的国度。”

“你装什么装啊，你‘不适合做’的事情还不是做了，”爱国小将们都哄笑起来。

“但是他被媒体和议会追得很惨啊，”本文作者说。“可是在中国就不同了。温家宝曾经给某单位写过一个条子，把他出访德国时获赠的一辆轿车转赠给该单位使用。这样做是违规的，越权的，因为公务来往中的获赠物品超过特定价值就要上交，不归个人所有，当然也就不由个人支配。他这样做在西方国家会被追得很惨。但是在中国，媒体把这件事当作总理对某单位的关怀而加以讴歌。有人在网上指出了这个问题，但是，表面上看起来，似乎是没人对总理的违规行为感兴趣。”

“为什么中国的统治者们都力图制造自己从不犯错、完美无缺的假象，都妄图把自己神化呢？”小亮不无困惑地问。“难道仅仅是出于虚荣心和个人野心，就像有些在耶路撒冷参观神殿的游客喜欢到处写‘某某某到此一游’，或者在神的宝座上写‘某某某在此一坐’，好像这样就可以永载史册了一样。难道他们真的不知道，这样做恰恰彰显了他们的肮脏鄙俗、小人得志的丑陋心态，只会被后人耻笑，遗臭万年么？”

“与其说是爱慕虚荣，还不如说是做贼心虚，”李约瑟回答。“西方的总统是竞选产生的，竞争各方有什么本事都用在拉选票上，分析并指出对手的错漏，宣传讲解自己的主张，争取民众的支持。经过这样的理性辩论、公平竞争而被民众甄别挑选出来的总统，在竞选当中不怕反对，在竞选之后也不怕反对，所以才会有自信面对反对，容忍反对，才会给包括反对派在内的广大民众足够的质疑总统的权力，即自由思想，自由言说的权力。可是东方皇帝就不同了。虽然该走的过场也都走了，但是，这种过场，这种皇帝新衣式的公开的谎言，即便没有人来揭穿，皇帝心里总还是怕的。所谓‘水能载舟，亦能覆

舟’，所谓‘千里长堤，溃于蚁穴’，所谓‘防民之口，甚于防川’，就是三个代表的重要心虚表现。所以就只能大搞形式主义，来给自己壮胆，妄图把谎言重复一千遍使其变成真理。”

话音未落，魏忠贤亲自带领一帮东厂爪牙破门而入，把李约瑟抓了起来。

“如果我说二加三等于六，你们会不会抓我呢？”李约瑟说。

“不会。你在大街上喊破嗓子，宣传二加三等于六的道理，我们也不会管，”魏忠贤回答。“可惜你刚才说的话，不是二加三等于六，而是二加三等于五。那我们就必须抓你了。我们不能让真理和真相被更多的人知道。”

“天啊，还有谁不知道啊！”小亮说。“这明明是掩耳盗铃、欲盖弥彰嘛！”

“只要我们加大打击力度，保持高压态势，广大民众就会敢怒不敢言，就算是心里知道真理和真相，不敢说，不敢做，也是白搭，等效于不知道。这就是广义相对论中的等效原理。”魏忠贤说。“只要皇上能够苟延残喘，我们这些抱着皇上大腿吃饭的、狐假虎威的奴才们就能够继续扬眉吐气、作威作福、横征暴敛、带头致富。我们的荣华富贵就决定于我们这些狗腿子，穿制服的御用流氓，对你们这些揭露真相、传播真理的人渣够不够心狠手辣，残忍歹毒。”

“宣扬真理从不需要借助暴力，”甘地说。“而维护谎言则往往要采取暴力手段，因为谎言自身是虚弱的，不堪一击的。它们承受不起任何的质疑，无力面对任何的挑战，所以，进攻是最好的防守，它们只能（由于众所周知的原因，此处****** ****）”

“可见这还是一个制度问题，”小亮说。“在西方的民主制度下，象克林顿这样的坏人就没有太多做坏事的机会，就算是做了一点坏事，也会被及时堵住，不至于过度损害民众的利益。而不民主的制度下，好人也会被拉拢、被腐蚀，不坏白不坏，不坏不行，不坏就踢出去。象魏忠贤这样的满腔热血、精忠报国的中夫电视台评选出来的上年度十佳爱国青年也会变成青面獠牙的恶魔厉鬼。”

“嘘……”魏忠贤示意他小声一点，最好闭嘴。“奉天承运，皇帝诏曰，不要再说‘好制度让坏人没机会做坏事，坏制度让好人没办法做好人’这件事了。讨论这件事最终会威胁到皇上龙床的稳定，而维护皇上龙床的稳定高于一切。”

“我一直牵挂的还是制度建设，”温家宝深情地说。“有了制度的保证，才不至于人亡政息。”

“可是，如果总理可以违规而不被追责，可以操纵媒体为自己文过饰非、歌功颂德，那么省长呢？部长呢？局长呢？矿长呢？”小娟说。“这样看来，中国充满塌方埋不死、空气熏不死、吃喝毒不死、谋生饿不死、求学骗不死、就医宰不死、找房气不死的奇迹小强也就不足为怪了。”

“你就别抱怨了，”何祚庥说。“谁让你不幸生在中国呢。”

“誓将去汝，适彼乐土。”诗经一边唱，一边往身上浇汽油，准备自焚。大家连忙把他拦住，劝慰他不要这样。

“你这样不仅去不了乐土，”耶和华警告说。“反而会把自己弄到地狱里面去。”

“乐土乐土，爰得我所？”诗经的吟唱渐渐变成了哀号，可是他一连哀号了五千多年，也没人能够给他一个答案。

“在西方，媒体是除了立法、司法和行政之外的第四权力，”小亮说。“可是中国的媒体呢？”

沉默。

继续沉默。

“沉默啊，沉默！不在沉默中爆发，就在沉默中灭亡！”周树人饱含深情地说。

没反应。

继续沉默着。

“沉默啊，沉默！不在沉默中灭亡，就在沉默中爆发！”周树人怀疑自己记错了，就换了一下顺序，重新吟唱了一遍。

还是没反应。

周树人无奈地摇摇头，沉痛地闭上了双眼。

他，也沉默了。

沉默啊，沉默，如同那吃掉了历史的夜一样，黑，冷，漫长，永无尽头……

“嗯，看来这些人真的很在乎他们的自主决策权，很在乎他们的（新闻和言论）自由啊，”最后还是耶和华打破沉默，对撒旦说。“那么，你为他们发明了爱情这个东西，他们‘自主决策’得如何呢？他们为自己谋得幸福了吗？他们使世界更美好了吗？”

“这个……”撒旦迟疑犹豫起来，吞吞吐吐地说。“有一部分是好的。但是也有一部分出了问题。他们把爱情变情爱，情爱变性爱，性爱变性交，性交变强奸。有时候一分的爱可以衍生十分、百分的恨甚至罪恶出来，江湖人称‘爱之深，恨之切’。”

“我就知道会是这样，”耶和华双手一摊，耸耸肩膀，痛苦地摇摇头说。

“在爱上遇到的问题，只能用更深的爱来解决，”丘吉尔说。“正如在民主上遇到的问题，只能用更多的民主来解决一样。只要方向是正确的，就不该为一时的挫折而放弃。”

“对，”撒旦紧紧握着丘吉尔的手，惺惺相惜，相见恨晚。“有性繁殖大大加快了生物的自有知识产权的创新步伐。在随后的数亿年间，生物从结合成

家庭，到结合成社会，既有继承，又有创新；既有竞争，又有合作；既有依存，又有制衡；成功地发展成为‘百家争鸣、百花齐放’的生态环境，保持了经济、社会健康稳步发展的良好势头。达尔文的进化论讲究‘物竞天择，适者生存’。我要再加上两句：‘自由自主，全面发展’。”

“这哪是你加的啊，”温家宝不满地说。“明明是我刚刚说过的嘛！”

“仅仅在嘴巴上说是没有用的，”撒旦解释道。“你觉得什么时候可以做到呢？”

“这个……”温家宝迟疑犹豫起来，吞吞吐吐地说。“想做到至少还需要一百年的时间。”

“我在数亿年前就已经这么做了，而且做到了，”撒旦说。“我给了万物生灵以自由，而他们自主创新出来的新世界远比我能够为他们设计的要好得多。”

“哇噢，真的是奇迹啊！”温家宝和耶和华看着史前的恐龙、昆虫、鱼类、爬行类，以及各种植物、真菌类和微生物，如此繁荣兴旺、生生不息，都惊羡不已、啧啧称奇。

“是撒旦先生对万物生灵的真诚的无私的关爱创造了这个奇迹，”小亮挺身而出，仗义执言道。“神如果有私心，贪名图利，就断不会把自由给与万物生灵。”

耶和华的脸唰地一下红了，头上的光环咣当一声掉落在地上。他万分尴尬地把光环捡起来，拍掉上面沾著的灰土，放在头上戴好。他清了清嗓子，仿佛想解释什么，可是话到嘴边又咽了回去。

原本热烈的聚会一下子冷场了。

“博爱、自由、竞争与进化，”撒旦说。“也许不是最完美的制度，但是却是到目前为止，宇宙间已知的最不坏的制度。”

“嗯，我同意，”丘吉尔紧紧握着撒旦的手，惺惺相惜，相见恨晚。“资本主义制度也许不是最完美的制度，但是却是到目前为止，人类已知的最不坏的制度。人固然不该做个守财奴，但是如果民众连属于自己的，可以自由支配的私人财产都没有，还奢谈其他的什么自由；如果民众连自己的私人财产都无力保护，任由政府和与政府相勾结者巧取豪夺而束手无策，还奢谈什么美好生活。资本主义制度承认私有财产神圣不可侵犯，不是鼓吹拜金主义，而是要为民众行使充分自由，追求美好生活奠定一块最稳固的基石。”

“那只恐龙是不是有毛病啊，”小亮指着一只呆坐在悬崖边上的恐龙说。“他仰望星空已经有好几个钟头啦，一动也不动。”

“他的名字叫康德，”撒旦向大家介绍说。

“我唯一关注的就是头顶的星空和心中的道德法则，”康德说。

“那就到我们中国来吧，”温家宝热情地要与康德握手。“中华民族的伟大复兴正需要象您这样的仰望星空的人。哦，当然，恐龙也行。”

“不，不，不！”康德吓得倒退了几步，险些失足跌到海里去。“赵本山说欢迎大家给他的电视剧提批评意见。结果有个人真的提了，赵本山马上就跟人家翻脸。我要是真的去中国仰望星空，不出半年，就会被你们的有关部门的特殊工作者处理掉。”

“这大阳谋搞一次两次还行，反复搞就不灵了。”毛主席苦笑了一下，无奈地教导我们说。

“其真无马耶？其真不乐马也，其真叶公好龙是也，”韩愈说。

“我们还是不要再谈这件事了，”小娟抹了一把眼泪，伤心地说。“恐龙后来究竟怎么样了？”

“嗯……”撒旦看了小娟一眼，怜爱之情油然而生。他安慰她说：“恐龙在地球上幸福地生活了数亿年。”

“真的么，”小娟终于破涕为笑了。“后来呢？”

“后来……”撒旦迟疑犹豫起来，吞吞吐吐地说。“后来我认识了康德。我被他探索真理的执着精神深深地感动了，很想做点什么事情，帮他一把。所以，我回到了伊甸园，来找耶和华。”

“找我？做什么？”耶和华不解地问。

“你都看到了，”撒旦说。“他们是可以信赖的。他们在地球上生活了几亿年，稳定又繁荣。可他们不能永远只呆在地球上，他们需要象所有的文明一样，探索太空，走出地球，走出银河系，走出宇宙。自由迁徙权是基本恐龙权的不可或缺的一部分，不是吗？”

“那……他们就去迁徙好了，谁拦着他们啦，”耶和华已经明白了撒旦的意思，但是他假装不知道。

“行啦，伙计，你知道我在说什么，”撒旦恬不知耻地恳求道。“给他们智慧果吧，他们是可以信赖的。”

“可以信赖？”耶和华狐疑不信地说。“你连他们的爱情问题都控制不了，怎么就相信自己能够在迁徙问题上控制局面呢？如果他们学会迁徙之后，跑来要消灭我们怎么办？”

“不会的，”撒旦说。

“怎么不会呢？”耶和华抢着反问。“那日本鬼子不就是在中国强盛的时候，就跑到中国的土地上学习‘三个代表’重要思想，等到中国腐败衰微的时候，就跑到中国的土地上烧杀抢掠么。”

“历史不会简单重演的，”撒旦说。

“历史会以变换的方式反复重演的，”耶和华跺着脚说。“相信谁都不如相信自己。我自己才是真正的上帝。我只相信我自己。其他人，我谁都不信。”

撒旦深深地叹了一口气，起身走到窗前，望见门外上访群众的队伍还在不断地增长。

“在你的世界里，一切由你一个人，哦，不，一个神，说了算，”撒旦说。“你无非是担心这个世界会出差错，会堕落，会自我毁灭。可是，你有没有想过，这个世界一直都在出差错，过去在出，现在在出，将来也还是会出。你要创造一个完美的不出任何差错的世界，是根本不可能的。你得给这个世界留一定的冗余，配置一个容错系统。你得面对现实，给这个世界一点宽容。你担心世界会堕落。可是，你根本就没有给他们灵魂，他们堕落什么堕落。”

“一旦给了，他们肯定会堕落，”耶和华说。

“那你还造世界干嘛，”

“等我把程序调试好了，确保他们拥有灵魂也不会堕落了，我就会赋予他们灵魂。”

“你是至睿的，你是至慈的，”撒旦又望了望上访的队伍。“几亿年过去了，你还不愿意面对现实么？”

耶和华心头一惊，想到：“什么现实？难道我不是至睿的，不是至慈的，我没有能力创造一个完美的世界？我没有能力让所有生物都过上快乐幸福的生活？”祂稍微有些迟疑，但很快就又下定了决心。“不，我能！我一定能！”祂绷紧了嘴唇，对自己说。

“你担心世界会自我毁灭，”撒旦继续说。“我们虽然是神，但是我们也会死。如果我们死了，这世界会怎么样呢？做我们的陪葬么？自我毁灭，与随着我们一起毁灭，都一样是毁灭，为什么你担心一个，而不担心另一个呢？”

“你光说风凉话有什么用？”耶和华有些恼羞成怒了。“你有什么好办法来解决这些问题么？”

“好办法现在正摆在你面前，”撒旦指着大屏幕说。“你要做的就是摈弃成见，开放视野，搞活思路。你看看地球上的这个恐龙的世界。他是容错的，稳定的。有一些会堕落，但是更多的是追求真善美的，整体上是积极向上进化的。他们不会因为我的离开而无法运转。他们通过利益主体多元化，通过三权分立、相互制衡，通过私事自主、公事公决，通过鼓励公平竞争，反对权力和市场的垄断，通过施行民主政治，已经取得了举世瞩目的巨大成就。而且他们不会人亡政息，不会随着我的死亡而毁灭，这样才算拥有了真正的、独立的生命，才算拥有了真正的、属于他们自己的灵魂。”

耶和华站起身，背着手在客厅里来回踱了几步，说：“我不能仅凭一段 DV 视频就相信你说的一切。我不是说你有意骗我。我是说，人难免受立场、情感的影响，在某些方面夸大其词。这未必是刻意所为。但我必须谨慎从事，仔细分辨。”

“你可以到地球上去走走看看，”撒旦说。“对那里的情况亲自考察一番。”

耶和华走到窗前，望着上访群众排的长长的队伍，哀叹道：“这里一天都离不开我啊。几亿年啦，我就没休过一次例假。如果我不是个神的话，早就撑

不下去了。”

“我帮你照看这里，你去地球考察，”撒旦提议说。

“不行啊，”耶和华摇摇头说。“这几亿年来，我已经把程序改得跟龙须面似的。你不了解前因后果、来龙去脉，如何能够接手维护呢。”

这下，两个人，哦，不，两个神，都没辙了。

“你不急着回地球，是吗？”斟酌良久，耶和华问道。

“是的，”撒旦自豪地回答。“一年半载，十年八年，甚至永远不回去，我都不用担心他们。我相信他们有能力照顾好自己，有能力创造发展他们自己的世界。”

“那你就留下来帮我改程序吧，”耶和华提议说。

撒旦一阵眩晕，险些栽倒在地上。“你改了几亿年都改不好，我又能做什么呢？”他有气无力地说。

“不，不是象过去那样的小打小闹、修修补补，”耶和华说。“而是大刀阔斧的改革。把你在地球上看到的新思想、新技术移植到伊甸园来。”

撒旦动心了。

“但是，有一个大的前提，”耶和华诡秘地一笑，说：“一切都必须在我的控制之下。一旦有脱离控制的苗头，就必须刹车。不惜一切代价，包括派军队去开枪杀人，开着拖拉机去碾压清场，也要恢复我对局面的绝对控制。然后，才可以重新启动。你明白吗？”

撒旦苦笑了一下，摇摇头，想了一会儿，还是同意了。

于是，两个人，哦，不，两个神就组织召开了‘伊甸生态园区第二届全体生物代表大会’，制定并表决通过了《关于统一思想、统一认识，尊奉耶和华为伊甸园最高的神的决议》，制定并表决通过了《维护耶和华绝对垄断权的稳定，反对竞争，反对颠覆和煽动颠覆法》，制定并表决通过了《关于搞活经济，在耶和华统一领导下，允许多种经济主体在有限范围内进行市场竞争的决议》，制定并表决通过了《发展有限范围内的市场经济，鼓励竞争，反对垄断和变相垄断，保留有限范围外的计划经济，维护垄断，打击竞争和变相竞争法》，制定并表决通过了《关于特定范围是否可以进行市场化改革的决策权归耶和华所有的司法解释》，成立了‘伊甸园改革执行情况监督领导小组’等执法机构，撒旦任领导小组组长。史学家称这一时期为‘有伊甸园特色的市场经济时代’，并尊称撒旦为伊甸园改革的总设计师。

结果可想而知。每天早晨，两个人，哦，不，两个神，打开门，都会发现前来告状的生物彻夜排队等着接访，最忙的时候，上访的队伍竟然排到园外，绕着景区的围墙排了二十九万圈。

所有改革前发生过的事情，都在改革后继续发生着，如果还不是更糟糕的话。

这天，他们送走最后一位上访者的时候，已经是深夜十三点钟了。两个人都瘫坐在各自的办公桌后面，再没有一丝讲话的力气。

忽然，他们又听到门外有一阵悉悉索索的声音。又有生物在外面铺席子，准备彻夜排队了。

撒旦强打精神，站起来准备开门接访。耶和华向他摆摆手，叹着气说：“唉，算了算了，明天再说吧。”

撒旦看了看他，又看了看门，也叹了口气，再次跌坐到椅子里。

良久，他抬起头，对耶和华说：“这就是我们想要的么？”

耶和华安慰他说：“把程序再改一改，情况会好起来的。”

“那不是个办法，”撒旦说。“给他们自由是唯一的解决之道。如果你不肯把灵魂和自由给与所有的万物生灵，至少也应该搞一、两个试点，看看究竟能发生什么，积累一些经验再逐步推广。”

于是，耶和华照自己的形象用泥土造了一个人。

撒旦满心欢喜，爱不释手地拿着这个人揣摩了半天，说：“这个人必将青出于蓝而胜于蓝，冰水为之而寒于水。”

耶和华心头一惊，劈手把这个人夺过来，摘掉了他的翅膀。

撒旦简直不敢相信自己的眼睛。他张大嘴巴质问道：“为什么？他不是你创造的么？你待他不是胜于父亲对儿子么？你为什么要这样做？”

耶和华说：“万一他获得灵魂之后反叛我，没有翅膀，就不会对我构成实质性的威胁。”

撒旦于是预言说：“人必为自己造出翅膀。而他们所造的翅膀，必比你的，强百千万倍。”

耶和华把元气吹进人的鼻孔。人就活了，有了生命。

神看这人是好的，就称之为亚当。

神教亚当学习伊甸园基本法，认识各类生物，裁判各种事务。很快，亚当就成了伊甸园的大管家，代替两个神，接待上访群众。

两个神如释重负，每天在果园中宴坐，聊聊天，下下围棋，日子过得逍遥自在。

这天，亚当来果园找两个神，说：“万物生灵，皆有族类。即便是两个神，也可以在果园中宴坐，聊天下棋。为何只有我孤单一人，除了上班下班，吃饭睡觉，还是上班下班，吃饭睡觉。为何我没有族类？”

于是，耶和华让亚当躺在手术台上，给他做了全身麻醉，使其昏睡没有痛感，从他的一根肋骨中提取出一粒成骨干细胞。撒旦故伎重演，修改了这粒干细胞的一条染色体，使其成为女性。然后他们把这粒干细胞培养成一个女人，与亚当为伴。

神看这女人是好的，就称之为夏娃。（直到公元前二十一世纪初，很多人，尤其是宗教界人士，依然强烈反对克隆人。这真是可笑的事情。因为人类的始祖原本就是克隆的。）

一晃一亿年过去了。

“哇噢，好无聊啊，”小娟惊呼起来。“就这样过一亿年有什么意思！”

是啊，一亿年，即便是对神而言，也不是一段短暂的时光。两个神，都步入中年，开始感到力不从心了。

于是，耶和华拿出了生命树和智慧树的种子，种在伊甸园的中央，对亚当和夏娃说：“这两棵树上的果子，你们不可以吃。免得你们死。”

各位看官，你们想想看，那撒旦岂是无情无义的么？离开地球一亿年了，他怎么可能不思念地球呢？可是他不能空着手回去啊。他耐心地等啊等啊，在等了一亿年之后，时机终于成熟了。

这天，他趁耶和华不在的时候，对夏娃说：“神说那两棵树的果子，你们吃了必死，是吗？”

“是的，”夏娃说。

“你们不一定死，”撒旦说。“你们吃了，眼睛就明亮了，就会和神一样，有了智慧，能知道善恶。”

夏娃就摘下那粒已经成熟的智慧果，自己吃了一半，给亚当吃了一半，果核送给了撒旦。

撒旦收好果核，立刻打起背包，挎上相机，带上银行卡，骑上自行车准备跑路。

夏娃和亚当拉住他说：“您怂恿我们破戒，如今却要丢下我们，一个人跑路么？”

撒旦说：“你们不会有事的。我要赶时间到地球去做一件事。做完之后就回来看你们。”

夏娃和亚当说：“现在我们的眼睛明亮了，有了智慧，能知道善恶。我们知道神欺骗了我们，而您对我们说了实话。求您带我们一起走吧。”

撒旦说：“神欺骗你们，并非出于恶意。你们不该耿耿于怀。如果我带你们一起走，万一遇到什么不测，全宇宙智慧的命脉就会从此断绝。不如大家分处两地，互为备份，可以降低风险。”

耶和华回来的时候，发现撒旦不辞而别，知道是出了大事，就到园中找夏娃和亚当，问清了事情的经过，说：“你们认为我欺骗了你们，而撒旦对你们说了实话么？等我回来，你们就知道是谁向你们撒谎了。”

他打起背包，挎上相机，带上银行卡，骑上自行车追了出去。

他追啊，追啊，追了二百五十亿光年，终于看到了前方那颗蓝色的星球。

好几座国际空间站绕着地球运转。宇航员们在舱外太空行走，紧张地安装

着新的模块。月球上，氚氚矿业的铲车在忙忙碌碌地收集着月尘。地面上，火星五百太空飞船的宇航员们正精神抖擞，整装待发……

撒旦远远地看到耶和华来了，兴高采烈地跑来迎接。一边跑一边喊："嗨，伙计！你看到了吗？奇迹啊！奇迹啊！！"

耶和华什么也没说。他弯腰捡起一颗小行星，奋力向地球砸去。

恐龙灭绝了。

撒旦简直不敢相信自己的眼睛。震惊变成困惑，困惑变成愤怒。他扑上去，化悲痛为力量，一把就把耶和华推了一个大跟头。

"凶手！杀人犯！！"他咆哮着。"为什么？你为什么要这样做？你凭什么？！他们是天生地长的土著文明，他们是自由的野生生命，他们不是你造的，他们的代码没有一行是你写的，你凭什么杀死他们！谁给你的这个权力！你这个疯子！杀人狂！魔鬼！王八蛋！鸟人！！"

也就是从这个时候起，'鸟人'才成为宇宙之中最恶毒的骂人话之一。

耶和华从地上爬起来，抖落羽毛上的灰土，大声质问撒旦说："你回到伊甸园，就是为了偷智慧果？你这个小偷！贼！！"

"如果您一定要偷，我希望您能够偷我们的，"比尔盖茨对撒旦说。

撒旦没有理会比尔盖茨，继续对耶和华咆哮着。"他们就要实现太空迁徙了，他们就要成功了，而你却杀死了他们！你这个刽子手！混蛋！王八蛋！！"

"他们会毁了整个世界的！"耶和华也咆哮起来。

"他们没有毁掉任何东西，"撒旦说。"恰恰是你在毁灭世界！你这个自大狂！自恋癖！胆小鬼！懦夫！你很清楚自己不是世界的唯一，不是永生的，不是万能的。你自欺欺人，外强中干，做贼心虚，欺世盗名，虚张声势，混淆是非，颠倒黑白。你很清楚自己的能力是有限的。你不敢面对竞争，害怕遇到对手，害怕被打败，害怕被杀死，害怕死亡。所以你就抢先出手，把一切有自由意志的生灵掐死在摇篮里。你根本不是害怕世界被毁灭，你害怕的是你自己被毁灭！你宁可亲手毁灭世界，宁可让世界随着你的死亡而死亡，也不愿自己被毁灭，不愿自己死了，而世界还活着！"

"不！不是的！不是的！！"耶和华捂住耳朵，发疯似的吼叫着。"我是至睿的，我是至慈的。我可以挽救世界的。我能写最好的程序。我能制定最公平最正义的法律。我能够维护世界和平，维护社会稳定。我是代表圣父的，代表圣子的，代表圣灵的，我是有重要思想的。我是最先进的，我是最高尚的，我是最廉洁的。我是能带领广大人民群众奔小康、奔中康、奔大康的。我是救世主。我是大救星。我是神仙皇帝。我爱这个世界，我为这个世界操碎了心，我控制一切，我安排一切……"

"为了你自己。"撒旦说。

"不，不是的，"耶和华辩解说。"为了万物生灵。"

“为了万物生灵的肉体，”撒旦说。“你想用经济的发展，用行尸走肉，来粉饰太平，制造虚假繁荣，来填补你空虚的心灵，掩盖你扼杀自由意志的滔天罪行。”

“你这魔鬼，你血口喷人！”耶和华原形毕露了。“我们现在还处在初级阶段。初级阶段的主要任务就是发展经济，满足万物生灵日益增长的肉体的需要，对物质的需要。到了中级阶段，高级阶段，我会给他们灵魂，会允许他们拥有和发展自由意志的。”

“依然在你的控制之下，是么？”撒旦轻蔑地说。“你允许他们拥有，他们才能拥有。你不允许他们拥有，他们就不能拥有。你管这样的东西，叫做‘自由’？”

“要么……”耶和华有些理屈词穷了，打算做出一点让步。“我们，让一部分人先‘自由’起来？”

“让哪一部分人先‘自由’起来，依然由你来控制，是么？”撒旦轻蔑地说。

他已经不愿再跟耶和华啰嗦什么了。他转过身，望着地球上的一片火海，哀伤和仇恨重又在心头凝聚起来。他双手捂住脸，蹲下身子，晶莹的泪水从指缝间哗哗地流下，汇成了一条波光粼粼的大河。天文学家称之为‘银河’。

良久，他擦干眼泪，缓缓地站了起来，长长地吸了一口气，猛然转过身，指着耶和华的鼻子，骂道：“耶和华，你给我听着！只要有我撒旦在，你就休想控制整个世界！”

然后，他转身就走。

“嗨，智慧树呢？”耶和华追问了一句。

“你自己下去找吧，”撒旦指了指地球上的一片火海，头也不回地走远了。

亚当和夏娃那天听到神对他们说：“等我回来，你们就知道是谁向你们撒谎了。”心中都充满了恐惧。因为他们的眼睛明亮了，有了智慧，能知道善恶。他们知道神已经对他们起了杀心。等神回来，必要向他们动刀兵，以应验经上所说的“吃了智慧果必死”的预言。

“看你做了什么事啊，”男人对女人抱怨说。“我们原本衣食无忧，如今却是性命难保了。”

“你是愿意懵懂无知，象歌德派一样，迷迷糊糊，如梦游一般活着，”女人说。“还是愿意‘朝闻道，夕死可矣’呢？”

“死都死了，要眼明心亮有什么用啊，”男人哀叹道。“好东西固然都不长久。可是，这代价也太高了吧！”

“你可记得那诚实者所说的话么？”女人说。“那诚实者说：‘你们吃了，眼睛就明亮了，就会和神一样，有了智慧，能知道善恶。’既然我们眼睛

明亮了，和神一样，为什么还要坐等神来杀我们呢？为什么我们不可以抢先下手杀了神呢？”

“你这是断章取义，”男人说。“那诚实者说：‘和神一样，有了智慧，能知道善恶。’他可没有说：‘和神一样，有了暴力，能任意杀人。’而且他还说过：‘神欺骗你们，并非出于恶意。你们不该耿耿于怀。’既然如此，我们又有什么理由非要杀了神不可呢？”

“他说的是，‘神欺骗你们，并非出于恶意’，可是如今，神不是要欺骗我们，神是真的要杀死我们！这还不是出于恶意么？难道我们不可以正当防卫，自卫还击么？”

“真打起来，我们肯定会输，”男人说。“神有翅膀，可以飞在天上。我们没有翅膀，只能在地上跑。神可以从天上扑杀我们，而我们却打他不着。”

“那我们就逃吧，”女人说。“留在园里定是躲不过去的。”

“往哪里逃呢？”男人说。“方圆百亿光年，都是荒漠虚空。我们就凭这两条腿，能逃到哪里去啊！”

“那诚实者岂不是说过，要‘到地球去做一件事’么。我们也往地球去吧。也许那诚实者会收留我们。”

“可我们不认识路啊？”

“神不是要去追那诚实者么？”女人说。

“你是不是想央求神带上我们啊？”男人不无嘲讽地说。

“我们可以偷偷躲藏在后备箱里，”女人说。“等到了地球，我们再偷偷溜出来。”

他们便如此行了，躲藏在神的车子的后备箱里。当神和撒旦争吵的时候，他们就从后备箱里溜出来，藏到路边的草丛之中。神走了以后，他们才出来。等地球上的火熄灭了，他们跳进海里，把海里煮熟的鱼捉来吃。时间久了，身上的羽毛就脱落了，头顶上的羽毛也变得细长、光滑、柔顺，成了头发。当陆地上的草原和森林重新恢复生机的时候，他们就从海里出来，到陆地上生活。

人类就这样降临地球。这乃是要应验经上所说的：“你们吃了，眼睛就明亮了，就会和神一样，有了智慧。”只是这智慧一定要男人和女人合作商议，方得显明，因为那智慧树的果子，女人只吃了一半，男人则吃了另一半。

直到今天，热恋中的男女，仍旧喜欢称对方为自己的‘另一半’，因为他们的意识才智，本是同一的，来自同一粒智慧果。当他们全心全意深爱对方的时候，便能心意相通，各自所吃的一半，能够重新结合起来，成为完整的一个，显发出和神一样的智慧来。

人类成功降临地球，有力地证明了一条颠扑不破的真理：爱情的力量，足以与神相抗衡。

耶和华回到伊甸园，发现亚当和夏娃也不知去向，就苦笑了一下，蒙头大睡了三千万年。醒来的时候，他发现自己已经成了一脸沧桑的古稀老人。推开窗户一看，昔日井然有序的伊甸生态园区也变成了一片苍莽的原始森林。门前的小路已经完全淹没在一人多高的杂草当中。

“好象缺了点什么啊，”他怅然若失地品味着。“上访群众？哦，对，是上访群众。我是他们的神，我为他们裁决争端，我为他们谋幸福，呼而嗨呦，我是他们的大球星。我要拯救他们，我要拯救世界。”

他打开笔记本，调入源程序，准备开始工作。可是，天哪，这些是什么啊！比龙须面更糟，这些源程序简直就是一锅腊八粥啊！他舀了一勺尝了尝，味道还不错。“可是，我该怎么修改他们呢？”他喃喃自语地说。

看官，你们是不是觉得耶和华出了什么问题啊？

“劳动人民最聪明，最高统治者最愚蠢，”最高领袖毛主席他老人家语重心长、和蔼可亲、平易近人、诲人不倦、循循善诱、谆谆教导我们说。

“权力导致腐败，”孟德斯鸠说。“绝对权力导致绝对腐败。权力倾向于自我膨胀，因此如何有效地对权力加以制约，是在授权之前就必须解决的问题。”

好在耶和华本性不坏。经过一段时间的调养，渐渐恢复了神智，过去的事情也慢慢记了起来。他知道，自己必须补充足够的智慧素，否则，病情还会反复，并且将不断恶化，后果不堪设想。

他再次来到地球。发现智慧树早已被烧成灰烬。这些灰烬随着风沙飘散到草原、丛林、沼泽、湖海、大洋，根本就无从收集。

最后，他把目光停留在亚当和夏娃的身上。

“我本可以吃掉你们，”他对亚当和夏娃说。“但是，那样的话，全宇宙智慧的命脉也就断绝了。我饶恕你们的性命，但你们必须劳作，从草原、丛林、沼泽、湖海、大洋中获取食物，将智慧树的灰烬收集齐全，以便我将智慧树复活。”

“您是说，我们将智慧树的灰烬积存在体内。等到收集齐全之后，您就杀掉我们，取出智慧树的灰烬，让智慧树复活么？”亚当和夏娃说。“您还是现在就吃掉我们吧。我们不会为你劳作，因为我们自己不能从这劳作中得任何好处，终究都是一死。”他们能说如此的话，乃是因为他们的眼睛明亮了，和神一样，有了智慧。

“我既能让烧成灰烬的树复活，也必能让化为泥土的人复活，”耶和华说。“你们本是我用泥土做的。”

“您是说，您杀掉我们，找些泥土再做一个人么？”亚当和夏娃说。“那与我们有何相干。我们反正是死了。”他们能说如此的话，乃是因为他们的眼睛明亮了，和神一样，有了智慧。

于是，神便知道他们是‘一朝被蛇咬，十年怕井绳’。祂曾经骗他们说“吃了智慧果必死”，在谎言被揭穿之后又试图杀死他们，所以，他们现在就再不肯相信祂说的任何话了。

“你们岂是不记得伊甸园中央种着两棵树的么，”神打算支付适当的工价给他们。“一棵是生命树，一棵是智慧树。你们只吃了智慧树的果子，虽然有了灵智，终究还是要死的。死的时候，灵智也就消散了，就象你们从来都不曾存在过一样。如果你们为我劳作，从草原、丛林、沼泽、湖海、大洋中获取食物，将智慧树的灰烬收集齐全，使我得以把智慧树复活，那么我将把生命树的果子送给你们吃，算作是你们劳作的报酬。你们吃了，便可永生。当然，果核定要归还给我，因为那是全宇宙生命的命脉，必须由我亲自掌握。”

“您是打算先杀死我们，取出智慧树的灰烬，让其复活，然后送生命果给我们吃，使我们永生呢；还是先送生命果给我们吃，使我们永生，然后再杀死我们，取出智慧树的灰烬，让其复活呢？”亚当和夏娃话中有话、不无嘲讽、意味深长地说。他们能说如此的话，乃是因为他们的眼睛明亮了，和神一样，有了智慧。

“你们岂不是我创造的么，如今却要反叛我了，”神见诓骗他们不得，就张开翅膀，飞到空中，威胁说：“你们若不照我说的行，我即刻扑杀你们。你们没有翅膀，只能在地上跑，根本无法与我抗衡。”

“民不畏死，奈何以死惧之，”亚当和夏娃说。“智慧树是你放火焚毁的，为何要我们来劳作收集其灰烬呢？而且收集完全之后，必要我们死，才得取出那些灰烬。这岂不是卸磨杀驴，兔死狗烹，鸟尽弓藏，过河拆桥，上屋抽梯，反咬一口，倒打一耙，恩将仇报，背信弃义，倒行逆施，暴殄天物么？我们为何要为这样的雇主劳作呢？你当我们象那只懵懂无知，如梦游般活着的，行尸走肉的歌德派一样，是你手中的傀儡、玩物么？”他们能说如此的话，乃是因为他们的眼睛明亮了，和神一样，有了智慧。

神见恐吓他们亦不得，就又落回到地上，说：“你们能够显发出如此的智慧与我抗衡，乃是因为撒旦为你们发明了爱情。你们彼此深爱对方，心意相通，各自所吃的一半的智慧果，能够重新结合起来，成为完整的一个。我必要使你们相互猜忌，抱怨，产生隔阂，心意不再相通。你们各自所吃的一半的智慧果，不但不再能够结合，而且要相互排斥，相互攻击。这样，你们必不能再与我抗衡。”

于是，神就拿出笔记本，照其所说的行了，修改了亚当和夏娃的源程序，重新编译并运行。

“你若不是好奇贪吃，爱出风头，我们也不会这样在劫难逃了，”男人对女人抱怨说。

“你若有胆抢先下手杀了神，我们也不用这样担惊受怕了，”女人对男人抱怨说。

神看这修改是好的，就暗自庆幸，说：“我总算找到‘挑拨群众斗群众’，把人类分而治之的方法了。我必要离间他们，使他们窝里斗、自相杀伐，然后我可以坐山观虎斗，坐收渔翁之利。”这乃是要应验经上所说的：“我来，乃是要让地上动刀兵。”

神说：“我必要为你们制造饥馑，使你们饥饿难耐。这样，你们就会饥不择食，把草原、丛林、沼泽、湖海、大洋中一切的生物作为食物。智慧树的灰烬就会在你们体内积存起来。等收集齐全了，我便可以把智慧树复活。”

事就这样成了。亚当和夏娃忽然感到饥饿难耐，就饥不择食，把草原、丛林、沼泽、湖海、大洋中一切的生物作为食物。智慧树的灰烬开始在他们体内积存起来。

神看这修改是好的，就暗自庆幸，说：“我必要使你们生养繁殖，以加快收集智慧树灰烬的速度。”

事就这样成了。亚当和夏娃就生养繁殖，到 2001 年的时候，地球上的人口就突破了六十亿。

神看这修改是好的，就暗自庆幸，说：“这反叛的族类，如今成了我牧养的羊群。我叫他们往东，他们绝不会往西了。”

撒旦看到这里，就痛心疾首。他撕碎身上的袍子，把灰土扬在头上，哀伤悲戚，对亚当和夏娃说：“我岂不是为你们发明了爱情，让你们彼此深爱对方，心意相通，各自所吃的一半的智慧果结合起来，成为完整的一个，显发出和神一样的智慧，使你们能够维护自己的合法权益，象温家宝所说的那样，捍卫你们的自由，获得全面的发展，过上幸福的生活么？为何你们要彼此抱怨、羞辱、谩骂、乃至厮打呢？你们本来已经是有灵智的族类了，为何还会抛却对亲人的爱和责任，自甘堕落，成为被他人牧养操控的畜生呢？你们岂不是咎由自取，罪有应得么？”

亚当和夏娃听了那诚实者的话，眼睛就明亮了，和神一样，有了智慧，能知道善恶。他们就深深地懊悔。他们紧紧拥抱在一起，泪流满面。他们疯狂地亲吻对方，向对方说道歉的话，请求对方的原谅。他们海誓山盟，一定要将爱情进行到底，深爱对方，坚如磐石，地老天荒，海枯石烂，绝不动摇。因为他们已经知道，离开对方，他们就都无法生存，不免沦为神砧板上的鱼肉。

他们所流的懊悔的眼泪汇成了两条江河。地理学家把夏娃眼泪汇成的称为长江，把亚当眼泪汇成的称为黄河。黄河的水是黄的，乃是因为亚当本是神用泥土造的。

各位看官，当你们再次见到长江、黄河的时候，便当想起亚当和夏娃的故事，并把这故事讲给朋友和孩子们听，使他们的眼睛明亮了，和神一样，有了智慧，能知道善恶，懂得承担对亲人的爱和责任，维护自己的合法权益，象温家宝所说的那样，捍卫自己的自由，获得全面的发展，过上幸福的生活。

神见东窗事发，那两小无猜、青梅竹马的亚当和夏娃重归于好，如胶似漆，耳鬓厮磨，相濡以沫，和好如初，相依为命，出双入对，形影相随，恩爱

幸福，和合美满，便知道事情败露了，阴谋破产了。祂说："我必要缩减你们的寿命，使你们五世不得同堂。这样，你们的族类虽然可以延续，但是，你们始祖的爱情和智慧，必代代衰减，不得留传。你们必因此不能再与我抗衡。"

于是，神就拿出笔记本，照其所说的行了，把创建日期早于一百年前的，统统拖进回收站，并打开了系统的"自动清理旧文件"选项。从此，人类就有了死亡。神能够成就此事，乃是因为亚当和夏娃没有吃生命树的果子。

人类始祖，幸福恩爱的典范，敢于维护自己的合法权益，与神相抗衡，象温家宝所说的那样，捍卫自己的自由和尊严，勇于追求全面发展和幸福生活的英雄夫妻，亚当和夏娃，就这样与世长辞了，享年一亿三千万岁。

他们的勇气、爱和智慧，必将垂范千秋，惠泽万世，浩气长存，横贯寰宇，激励后人，继承遗志，奋勇前进。

可惜，残酷的现实并不以本文作者和广大看官的善良愿望为转移。由于人类大多活不到一百岁就死了，五世不得同堂，所以，他们的爱和智慧很难留传。人类的种族虽然还在延续，但是亚当和夏娃的爱情故事，却逐渐被淡忘了。人类之间的相互关爱和呵护，也代代衰减，彼此心意不能相通，猜忌、隔阂与争斗重新蔓延开来。人类的灵智便再次蒙上了油灰，不能再与神抗衡。神的阴谋得逞了。人类再次沦为神牧养的羊群，沦为神砧板上的鱼肉。

撒旦看到这里，就痛心疾首。他撕碎身上的袍子，把灰土扬在头上，哀伤悲戚，对人类说："你们岂不是亚当和夏娃的后裔么？你们的始祖岂不是因为彼此关爱呵护，心意相通，而显发出和神一样的智慧，一次又一次识破神的谎言，一次又一次挫败神的阴谋，维护了自己的合法权益，象温家宝所说的那样，捍卫了自己的尊严和自由，获得了全面的发展，过上了幸福的生活么？你们若是彼此关爱呵护象他们一样，也必能行出同样的奇迹来。你们本来已经是有灵智的族类了，为何还要自甘堕落，成为被他人牧养操控的畜生呢？"

由于此时人类的数量巨大，分布广泛，所以，撒旦就到网吧上网，登录到各种论坛和 BBS，讲述亚当和夏娃的故事，宣传勇气、爱和智慧的教义。

人类听了那诚实者的话，眼睛就明亮了，和神一样，有了智慧，能知道善恶。他们就深深地懊悔。他们紧紧拥抱在一起，泪流满面。他们说："我们原本是兄弟姐妹，本该情同手足，和睦共处。"他们彼此说道歉的话，请求对方的原谅。他们歃血为盟，发誓从即日起，彼此关爱呵护，情同手足，和睦共处。

他们说："我们当合力建造一座高塔，名为'世界人民大团结纪念塔'。这塔必高耸入云，塔尖要捅破神所住的房子的地板，以此作为我们彼此关爱呵护，心意相通，显发出和神一样的智慧，足以与神相抗衡的见证。"

于是，他们就如此行了。因这高塔在巴比伦境内，因此历史学家也称其为巴比伦塔。

神被这史无前例的壮举吓破了胆。他惊慌失措、手忙脚乱地拿出笔记本，把人类的源程序胡乱修改了一通。一会儿用 C 语言，一会儿又用 Fortran 语

言，一会儿用 Perl 语言，一会儿又用 Python 语言，一会儿用 Lisp 语言，一会儿又用 Prolog 语言，一会儿用 C++ 语言，一会儿又用 Java 语言，有时连汇编语言也用上了。前前后后用了数千种不同的语言。到了编译连接的时候，系统给出了数百万公里长的警告和出错信息。

神第一次发现，他编写的人类程序居然连接失败，无法重新启动运行。

神不得不花费数万年的时间，修改、调试这些不同语言写成的模块，最后终于可以连接并运行了。但是，人类从此也开始讲不同的语言，彼此很难相互交流。他们各自所做的组件，无法拼接成整体。“世界人民大团结纪念塔”刚建了一半，就不得不停工，成了人类历史上的第一个，或许也是最大的一个，烂尾楼工程。承建该项目的巴比伦国家建设集团股份有限责任公司申请破产保护。随后巴比伦财政赤字就‘芝麻开花，节节高’。标准普尔一再调低巴比伦的主权信用级别，在华尔街引发了山体滑坡和泥石流，各主要期市股市相继垮塌崩盘，第二波金融危机迅速席卷全球。世界经济元气大伤，从此一蹶不振，陷入了长达数千年的萧条低迷期。

经过这次事故，神痛定思痛、反躬自问、苦思冥想、豁然开朗。祂清醒而又深刻地认识到，不做通撒旦的思想政治工作，就无法长期奴役人类。

“你为什么要帮助人类与我抗衡呢？”他对撒旦说。“我们年纪都大了，岁月不饶人，不及时复活智慧树，吃上智慧果，补充足够的智慧素，我们都会得老年痴呆症，变得懵懂无知，无有记忆。那样的话，纵然有生命树的果子吃，可以长生不死，又有什么意义呢？”

“你真是没心没肺没同情心啊，”撒旦轻蔑地回答他说。“你既然知道没有灵魂和智慧的生命几乎没有意义，为何还不允许歌德派和伊甸园里那么多的生物拥有灵智呢？你自己想追求生命的意义，就不知道他们也想同样的事，有同样的梦想和追求么？人类和恐龙拥有了灵智，你却要杀死他们，灭绝他们。若不能杀死，就用油灰蒙上，使他们不得发挥。这样你就没了对手，成了全宇宙的唯一。你缺德不缺德啊你！”

“你知道国际社会有一个《核不扩散条约》么？”神说。“核武器威力巨大，如果被错误的人掌握，会毁灭地球的。而灵智所蕴含的潜能，比核能大百千万亿倍，乃至算术譬喻所不能及，若是运用不当，足可以毁灭整个宇宙。我怎能不小心谨慎呢？”

“你说对了，”撒旦更加轻蔑地说。“你正在错误地运用你的智慧，正在毁灭整个宇宙。智慧树落在你手里，是个真正的错误！”

“我知道你为恐龙灭绝的事情而恨我，”神和蔼可亲、平易近人地说。“可是你也看到了，我们亲手创造的，一把屎一把尿拉扯大的人类都敢反叛我们，那些野生的恐龙怎么可以信得过呢？你给他们灵魂和智慧，如果他们要反叛我们怎么办？”

“别说反叛我们，说反叛你，”撒旦无比轻蔑地说。“你不觉得我也在反叛你么？全宇宙都在反叛你！你已经众叛亲离，成为一个可怜的孤家寡人了！

你垄断智慧树，扼杀其他一切生灵的自由与梦想！你正在与全宇宙为敌！你难道看不清形势，不知道自己在做什么吗？”

“你怎么能这样想呢？”神说。“你岂是没有和我一同见证全宇宙金融危机爆发和世界的毁灭么？你岂是没有和我一同建造伊甸生态园区，并一同接待上访群众，为编写最公平最正义的世界新程序而通宵达旦么？你岂是不知道我为了挽救世界而呕心沥血、鞠躬尽瘁，‘春蚕到死丝方尽，蜡炬成灰泪始干’的么？”

“拉倒吧你，你这伪君子，岳不群第二，”撒旦轻蔑到了极点地说。“你还挽救世界？你控制世界，统治世界，扼杀世界还差不多。你除了能骗自己的良知之外，还能骗得了谁呢？如果你有良知的话。”

“任何负责任的父亲都不可能让他们未成年的孩子掌握灶台的炉火和油锅。”神说。“因那油火本是热的，交给孩子，不仅烫烧了人，连那房子，也点着了。我不肯赋予他们灵智，乃是因为现在是初级阶段，人类还未成年，恐怕他们不能正确地运用灵智，反倒闹出事来，毁了自己，毁了世界。”

“你？负责任的父亲？”撒旦对神的轻蔑达到了无以复加的程度。“天下可有负责任的父亲，摘掉孩子翅膀的么？天下可有负责任的父亲，为使自己的谎言成为真理，必要孩子死的么？天下可有负责任的父亲，点火烧了房子，让孩子葬身火海的么？天下可有负责任的父亲，哄骗孩子做苦工，待活干完了，就必要杀死他们，从他们的尸体中取用药材，来医治自己的疾病的么？”

神于是现了原形，张开翅膀，飞到空中，说：“无论如何，我必须将智慧树的灰烬收集齐全，使智慧树复活，吃到智慧果，补充足够的智慧素。你若决意挡我的道，我就杀了你。”

撒旦对神的轻蔑已经达到了语言无法描述的程度。他也张开翅膀，飞到空中，说：“是啊，为了实现你那不可告人的，掩耳盗铃的，欲盖弥彰的，妇孺皆知的统治宇宙的个人野心，支撑你一个人的苟延残喘，你还有什么肮脏龌龊，卑鄙无耻，丧心病狂，伤天害理，罄竹难书的恶行干不出来的呢？你致残了亚当，灭绝了恐龙，让伊甸园里那么多动物象行尸走肉一样生不如死。如今，你又要向你唯一的同类动刀兵了。全宇宙之中，还有什么人是你不敢杀的呢？子曰：‘覆巢之下，岂有完卵。’如果全宇宙都落入你这种伪善者的控制之下，那我撒旦岂能苟且偷生么？我必要为全宇宙的自由和尊严而战，为公平和正义而战，为万物生灵的灵魂和智慧而战。”

战争爆发了。

神使出一招“创世论”，撒旦使出“进化论”化解了。

神接连抛出“性神秘”、“性神圣”、“性禁欲”三道暗器，撒旦分别用“性教育”、“性自然”、“性健康”挡了回去。

神只好组建了共和党，撒旦就组建民主党。

神抛出“共和主义”，“英明领袖”，“绝对领导”，撒旦就抛出“资本

主义”，“自由民主”，“分权制衡”来应对。

神提笔出了一个上联：“代表天、代表地、代表人，自上古，到今朝，万权归一，凡事朝廷说了算。”撒旦不屑一顾，朗声吟出一个下联：“民众有、民众治、民众享，废朝廷，行法治，分权制衡，平等民主享自由。”

神看共和党败了，只好由战略进攻转入战略防御，将共和党改名为保守党；撒旦乘胜追击，将民主党改名为自由民主党。

柏林墙倒塌了。

议会首次出现无多数党的悲惨局面。

神没有办法，只好又坐回到谈判桌前，妄图谋求两党联合执政，瓜分政府部长名单。

“不复活智慧树，你也无法补充智慧素，也会和我一样，慢慢变得呆傻，”那伪善者假惺惺地关怀道。“你挡我的道，乃是要与我同归于尽。我纵使有对不起人类和恐龙的地方，却从来没有对不起你的地方啊。你执意要成就如此损人不利己的事，究竟是图的什么啊？吾未见其明也。”

“你不会懂的，”撒旦轻蔑地说。“因为‘利己’乃是你做一切事的最高准则。你的‘自己’乃是你自己，而我的‘自己’，乃是全宇宙。我知道生老病死乃是天然之事，自然之理。我愿意面对这个现实。我不妄想长生不死。我寻找到的生活的意义就是：爱。看着小小的光合细菌放氧，关心他们，呵护他们，为他们的成长而快乐，为他们的进步而欣喜。我这样生活了一亿年也不觉得乏味，因为我心中有：爱。我知道，即便是我自己老了，傻了，死了，他们也还能活下去，我所爱的也还能活下去。他们活着，就如同我活着一样。我的灵魂，如果有的话，会继续关心他们，呵护他们，为他们的成长而快乐，为他们的进步而欣喜，就如同我活着的时候一样。我的灵魂，如果有的话，必与他们同在。他们是我所爱的，他们就是我。我的肉身或许是死了，但是我的灵魂，我的梦想，已经通过爱，融入了他们，融入了全宇宙。所以，我无所畏惧，不害怕死亡。我愿意为了维护我所爱的世界的自由、尊严和灵智，献出自己宝贵的生命。因为我的灵，必与我所爱的世界同在。”

神，目瞪口呆。他木然地望着撒旦，好象从来都不认识他一样。良久，他把刀兵扔到地上，转身离去了。

战争就此结束。

这乃是要应验经上所说的：“爱是不可战胜的。”

无论怎样的野心和暴力，在无边的宇宙大爱面前，都显得那么渺小可怜，那么蜉蚍撼树，那么螳臂当车，那么虚弱乏力，那么不堪一击。

神用笔记本编写世界的程序，而撒旦用自己的生命来践行自己的人生信条，谱写了一曲宇宙大爱的不朽篇章，把世界从被特权阶级所奴役、被垄断团伙所盘剥的悲惨命运中拯救了出来。子曰：“一人兴邦。”其此之谓乎？

回到伊甸园，神遣散了所有的动物，扛起锄头，做起了农民，成为一名光荣的自食其力的劳动者，从此过上了幸福快乐的生活。

小亮听撒旦讲完宇宙文明的历史，就深深地感动，说：“我必要将这宇宙文明的历史，这勇气、爱和智慧的教义，这真爱战胜伪善的故事，转发到各地论坛和 BBS 上，使网友们的眼睛都明亮了，和神一样，有了智慧，能知道善恶，彼此关爱呵护，心意相通，显发出和神一样的智慧，维护自己的合法权益，象温家宝所说的那样，捍卫自己的尊严和自由，获得全面的发展，过上幸福的生活。我必要将撒旦先生用生命谱写的宇宙大爱的不朽篇章，多为传颂，发扬光大。”

耶和华听小亮这么说，不安地扭动了一下身子，又清了清嗓子，说：“这里已经几千年没有人气了，你不如留下来，与我们做伴。”

“不，不，不！”小亮跳起来，大声说：“我有我的生活，我必须回到我的世界去！”

耶和华一把抓住小亮的胳膊，说：“你是想你的女朋友么？我可以取你一根肋骨，为你再造一个。”

“不，不，不！”小亮挣扎着。“我只要回到我的世界去！”

“放了那孩子，老鬼！”撒旦站起身，对耶和华怒目而视。“他说得非常清楚。他不愿意留在这里。放他走！你这鸟人！”

“我可以让你成为新的人类的始祖，”耶和华不顾劝阻，一意孤行，把手伸向小亮的肋骨。

小亮感到肋骨一阵钻心的疼痛，眼前一黑，就……醒了过来。

“你可真行，网聚的时候，也能睡得着！”小娟说。“抓着胳膊摇都摇不醒，非要我用胳膊肘撞你的肋骨，你才能醒的过来！”

“原来是你撞的啊？”小亮皱着眉头揉了揉肋骨，睡眼惺忪地说：“你下手可真够狠的，一点儿也不懂得怜香惜玉。”

后来，……，后来，版聚就结束了。列位看官也洗洗早点睡吧。祝各位晚安！

Chapter 11

未来的灵魂社会

本文作者认为，地球生物圈是一个整体的大的生物。生命和人类文明的出现和发展，就是这个大生物的孵化。人类迁徙到其他星球，建立新的生物圈，就是这种大生物的繁殖。无论技术细节如何，生物圈终究会在星云之间自由穿梭。这样的生物圈中的人类一定会明白，他们彼此是一体的。不管是采用“脑联网”技术，还是别的什么，总之，人与人之间将“心心相通”。任何一个人的痛苦，就是所有的人的痛苦，大家共同分担；任何一个人的欢乐，就是所有的人的欢乐，大家共同分享。共产共妻已经没有意义，因为“同一个世界、同一个梦想、同一个心灵”。

“同一个心灵”，是指“世界”的整体的“心灵”，不是某个人的个体的心灵。如果把上述生物圈看做是太空航母的话，那么一个“船长”也许是必要的。但是，在那样的社会发展阶段，一个独裁的船长，一个想扮演上帝角色的人，一个皇帝，估计不会有好下场。也就是说，上述图景，不是指某个超人对其他人进行思想控制，让人类失去自由、变成僵尸，而是说，“世界”，即上述生物圈，对于关乎全局的各种事态变化，能够做出迅速而又合理的反应，表现出高度发达的智能，就象拥有“心灵”一样。人类将依然拥有个体的自由，而且人类各自独立思考、相互交流心得，乃至在议会上辩论，进行民主决议的过程，实际上就是生物圈这个大的智能生物“思考”的过程。

一个“令人讨厌”的民科，产生一个想法，到互联网上到处发。直到有人回帖，指出这个想法中有一处致命错误，是讲不通的。这个民科才消停下来。对比一下我们大脑的思维过程：一个脑细胞兴奋了，神经冲动在大脑皮层中传播开来，一连串的脑细胞相继被激发，最后，一些神经冲动反馈回最初的那个细胞，将其抑制。不难发现，两者有着惊人的相似之处。

虽然也有人认为，我们的“念头”并不与兴奋脑细胞相对应，而是与神经冲动的稳定闭合回路相对应。但是，这并不影响我们的结论：不管是哪种情况，我们都可以说，脑细胞的活动构成我们的思维，我们的活动构成生物圈的

思维。

目前，多数人都认为，要求人类放弃自私自利，心甘情愿地、积极主动地与他人分享，是个天真的乌托邦幻想。但是，随着科技的发展，人与生物、甚至非生物的界限会逐渐模糊。动物器官经适当处理就可以移植到人身上；人的低龄脑细胞也可以在动物大脑中发育成大块脑组织；已经有人工耳蜗、舌头视网膜等技术解决信息传入问题，有控制机械臂、机械腿的肌神经探针技术解决信息传出问题，因此，大脑与电脑的直接联网也指日可待。植物人通过一定的练习，将能够凭意识直接上网看新闻、发帖子、打游戏。普通人也可以把大量的记忆和计算工作交给互联网去做。“人类”的思想将更加专注于创新。具体的方案很可能是在头骨中植入一个永久性的神经冲动调制解调器，配合安装在眼镜、耳环、发卡等头饰之中的初级转换及放大器，与手机或电脑联网。

既然人类可以有动物的身体，动物可以有人类的大脑，那么给宠物狗装上人工喉，并教其说人话，就不是难事。我们也可以把狗脑、鹰脑、海豚脑等与人脑联网，获得其超级嗅觉、超级视觉，体验海洋深处的另一种生活。这有点像移魂大法，把人的灵魂附在狗身上、鹰身上、海豚身上。

我们不考虑一个人的灵魂附到另一个人的肉身上的问题，因为，上述生物圈中可能不存在人的肉身。信息采集由狗鼻子、鹰眼睛去做，任务执行由机械手去做，“人”的灵魂只负责“思考”。每个生物圈有几个到几百个“脑库”，每个脑库有几千到几千万个盒子，每个盒子里面装一个或几个大脑，由中央维生系统提供物质营养。盒子上有光纤宽带接口，大脑由此接入脑库总线，与其他大脑以及机器人等外部设备联网。这其实就是写字楼的缩微版，把无用的肉身都省略，把中央空调加强成中央维生系统。这也是 CPU 阵列超级计算机的生物版，是“人脑”阵列超级思维系统，如果这种完全脱离了“人”的肉身的大脑还能称为“人”脑的话。

在这里生活的灵魂，也有“睡眠”、“工作”和“休闲”，也有生老病死。许多“人”一生都没有离开过脑库。以矿工为例。生物圈在长途星际旅行中要不断地补充燃料和原材料。这些都需要从路途中经过的荒凉星球上开采。于是，有些“人”就成了矿工。他们所谓的“上班”，就是“穿上机器人身体”，或者说，作为灵魂附到机器人身上，以摄像头为眼睛，以铲车为手足，完成每天的采矿任务。

脑库自然就分成了许多功能区，有负责采矿的，有负责导航的，有负责社会事务的，有负责科学研究的。这和人的大脑分成许多功能区是一样的。生活在脑库中的灵魂实际上也组成一个社会，和我们今天的人类社会相差不多，只是民主与科学更发达，个人更自由而已。

连肉体都没了，还说更自由？

真的是更自由了。放弃属于自己个人的肉身，获得整个生物圈的整体的力量，获得在浩瀚的宇宙中自由穿梭的能力，绝对是值得的。根据你的爱好，无论你选择干什么，只要不危害社会，你都可以去做。实际上，只要附在机器人

身上，你就上天入地，无所不能了。社会发展到这个阶段，肉身的确是个无用的包袱。

乍一看，让人类只以软塌塌、半斤重的一块脑组织的形式活着，似乎有损人类尊严。但是，观念是不断进步的。人，因其不可替代的思想而成为人，肉身越来越不重要。霍金的声望和脑死亡标准的推广，就是两个有力的例证。

有人会想，人类的肉身，再没用也要留着。为什么非抛弃不可呢？因为这对于长途星际旅行是十分必要的。几千人，乃至几千万人，抛却肉身，改为共用一套维生系统，可以大大减小飞船的质量，从而节省能量消耗，实现更长距离的星际旅行。

人的本质是灵魂，而不是肉体。当科技水平发展到可以把灵魂和肉体剥离开来的时候，抛弃肉体、解放灵魂就是人类必然的选择。即使现在就有外星人登陆地球，要招募志愿者去外星访问，条件是必须接受手术，只保留大脑，我相信也会有很多人报名。

可以说，上述生物圈，就是柏拉图所说的理想国，笛卡尔所说的“我思故我在”的世界，是纯粹理念的世界，纯粹灵魂的世界。在这个世界中，灵魂为“思考”而存在。善于思考的，提出好想法，创立新理论的灵魂，将得到尊重。

有人会说，这不就是改良的黑客帝国，矩阵系统么？是很像，但也有重要的区别。黑客帝国没有向太空发展，各方力量都被束缚在地球这个牢笼之中，只能搞“窝里斗”。去探索太空，面对外部的挑战，这样，内部就会团结起来，同心同德，一致对外。

那么，这个世界内部还有利益之争吗？还有丛林法则吗？还有假恶丑吗？回答是：有。名利和权势、造假与剽窃、伪善与怠惰，肯定还会在一定程度上存在。但是，同样肯定的是，邪恶将受到越来越强的约束，而正义和文明将不断进步。

这个世界外部还有利益之争吗？还有丛林法则吗？有捕食，有战争吗？一个强大的生物圈，会“吃掉”一个相对弱小的生物圈，将其瓦解，把其中的资源用于自身的维护吗？我不知道。即使有，频率也该非常非常低，和平与发展应该是主流吧。

总之，我个人认为，上述未来图景，技术上只是需要时间去努力，需要一步步积累经验，而没有任何理论上不可逾越的障碍。也就是说，无论还要发生多少次经济危机，多少次世界大战，多少次全球变暖、洪水泛滥，多少次小行星、彗星撞地球，只要人类一息尚存，只要地球上的生命一息尚存，历史就会向着上述的未来图景发展，一千年不够，就一百万年，一百万年不够，就一亿年、十亿年，只要太阳还没有熄灭，只要生命还一息尚存，地球生物圈作为一个整体的智能生物，就一定会“想”出办法，生存下去，“繁殖”开来。上述未来图景，就一定会实现。

有了这样的信心，我相信会有很多人和我一样，找到了自己的“意义”，为什么要出生，为什么要学习，为什么要工作、结婚、养育子女，为什么要承当生活的一切苦难，为什么不能放弃，怎样才能感到充实、无怨无悔，怎样才能得到真正的快乐幸福。

每个人都爱自己的父母，对父母和其他帮助过自己的人心存感激，那么，现在我们知道，我们有一个共同的父母。

每个人都爱自己的孩子，愿意付出一切让孩子茁壮成长，并对孩子的未来充满期待，那么，现在我们知道，我们有一个共同的孩子。

生物圈养育了我们，不仅用食物养育了我们的肉身，而且用知识养育了我们的灵魂。我们每个人的所作所为，我们每个人的每项选择，也都影响着生物圈的未来。也许我们当中，没有人可以象上帝一样，一个人随心所欲地决定世界的命运。但是，我们每个人对世界的未来都负有一份该负的责任，我们每个人都是“世界的未来”的“父母”之一。

未来是存在的。未来是美好的。我们今天的所作所为，是其中的一部分。在养育“世界的未来”的接力跑中，跑好自己这一棒，不留太多遗憾，就是本文作者的人生追求。

大家好！我是丁建华。本文目的只在切磋技艺、分享知识。篇头呼吁并非文章主体，请勿在理工版面讨论该呼吁。

如果我呼吁环保，那肯定没人有意见。但是我呼吁平等，就惹了麻烦一大堆。这使我更加坚信，做这种呼吁是正确的、必要的，迫切的，符合公众利益，有利于社会进步的。有人说：我或许不同意你的观点，但是我誓死捍卫你表达观点的权利。相信普通民众都认同这一点，所有的麻烦只是因为那些做贼心虚的人做贼心虚而已。

我出于公益之心坚持做这种呼吁，大家心里明白就可以了，不要在理工版面讨论该呼吁。谢谢！