High School Mathematics Done Right

中学数学这样学

吴金闪

为了理解数学而写：知识简单，思考深刻，体现数学

从现实到结构，从结构到现实

思维的语言，严密的论证，纯粹理性的愉悦

有系统，有联系，知识网络

教的更少，学得更多

List of changes

批注清单

1 插图解释交换律和结合律 35
2 插入图，用所有自然数，骰子等集合当例子，做一个用点的，一个用点加上记号的 40
3 插入图，集合包含关系，补集的包含关系 44
5 考虑是否印单独一页 46
6 插入集合的 Venn 图，交集和并集的 Venn 图表示。 65
7 插入子集、全集、补给、包含关系的 Venn 图。 67
8 插入一张给各个区域用集合运算符号表示的 Venn 图，例如 $X \cap Y, X \cup Y, X \sup \overline{Y}$ 等等 68
9 编制一些集合交和并运算，补集运算的习题。 68
10 插入符合函数的图示 74
11 插入圆环上顺时针编号，逆时针旋转的示意图。 78
12 插入 g 的两种映射的示意图。 79
13 插入群的两个视角的示意图：三个集合 T，两次映射操作，整合起来，成为复合起来的映射，显示每个元素；三个点，每个点代表一个集合，两个映射箭头符号，不显示元素，直接作用到集合上，合起来成为复合映射；一个点，两个映射箭头符号，合起来成为一个符合映射的箭头。 80
14 插入苏格拉底是要死的的示意图，两个具有包含关系的集合，一个元素，标上图中的符号。 86
15 编制一些三段论的应用题 86
16 插入 $A \subset B$ 和 $\overline{B} \subset \overline{A}$ 的示意图。 87
17 编制一些反证法的应用题 88
18 插入平面直角坐标系，标注一个点 $(1, 1)$ 当例子。 95

19 插入描点画图过程。 100
20 插入函数 $y = ax^2 + bx + c$ 和直线 $y = y_*$ 有和没有交点的示意图。 102
22 考虑是否印单独一页 108
23 （引用直线的定义） 118
24 插入 PMQ 直线示意图 119
25 插入描点法找 $(x = x_0, y = y_0)$ 的图 121
26 插入图：上半部分和下半部分两个半圆不同颜色，代表不是函数，分开成两个函数 122
27 插入图，编一些解析几何证明平面几何的题 123
28 插入图 124
29 编一些用几何方法求解代数问题的题 124

目录

第零章　对小学数学的要求和复习 **29**

0.1　对小学数学的要求 29

0.2　小学数学复习 33

0.2.1　从数数到数和运算律 33

0.2.2　现实和数学，从关系到数学结构 39

0.2.3　数学知识的系统性 39

0.2.4　从小学数学学习和复习到理解型学习 45

0.3　小学数学知识总结以及本章小结 45

第一部分　初中篇 **49**

第一章　数轴上的点和数 **51**

1.1　非负整数的数轴表示 51

1.2　分数的数轴表示 51

1.3　从 $\sqrt{2}$ 不是分数到 $\sqrt{\frac{q}{p}}$ 51

1.4　无限不循环小数的数轴表示? 51

1.5　有理数、无理数和实数 51

1.6　数轴上还有其它数吗? 52

1.7　代数式 52

1.8　推荐阅读材料 52

1.9　作业 52

1.10　本章小结 52

第二章　幂运算 **53**

2.1 数的正整数次幂运算 53
2.2 数的负整数次幂运算 53
2.3 数的零次幂 53
2.4 数的分数次幂 53
2.5 数的幂运算的性质的字母表示 54
2.6 完全平方和完全立方公式 54
2.7 平方差和立方差公式 54
2.8 推荐阅读材料 54
2.9 作业 54
2.10 本章小结 54

第三章 变量和常数 **55**
3.1 完全已知的数 55
3.2 给定的（当作已知）的数 55
3.3 未知的或者重点关注的数 55
3.4 单项式的元数和阶数 55
3.5 多项式的元数和阶数 55
3.6 推荐阅读材料 55
3.7 作业 55
3.8 本章小结 55

第四章 因式分解 **57**
4.1 乘出来和变成多个算式相乘 57
4.2 为什么需要因式分解 57
4.3 简单因式分解 58
4.3.1 基于乘法对加法分配律的因式分解 58
4.3.2 基于完全平方公式的因式分解 58
4.3.3 基于平方差公式的因式分解 58
4.3.4 基于立方差公式的因式分解 58
4.3.5 基于完全立方公式的因式分解 58
4.4 几种类型混合的因式分解 * 58
4.5 推荐阅读材料 58

4.6 作业 58
4.7 本章小结 58

第五章 一元一次方程和一元一次不等式 59
5.1 复习：运算律、等式的性质 59
5.2 一元一次方程的求解 59
5.3 一元一次方程的应用 59
5.4 复习：不等号和不等式的性质 59
5.5 一元一次不等式的求解 59
5.6 一元一次不等式组的求解 59
5.7 一元一次不等式的应用 59
5.8 基于完全平方的恒成立不等式及其应用 59
5.9 推荐阅读材料 60
5.10 作业 60
5.11 本章小结 60

第六章 集合初步 61
6.1 集合的朴素概念 61
6.2 集合的交和并 64
6.3 子集、全集和补集 65
6.4 不等式的解当作集合 68
6.5 推荐阅读材料 68
6.6 作业 68
6.7 本章小结 68

第七章 映射初步 69
7.1 从命名、分类、操作到映射 69
7.2 映射 71
7.3 映射的复合，最简单的群和表示理论 73
7.4 三个元素的集合到自身的映射 * 77
7.5 推荐阅读材料 80
7.6 作业 80

7.7 本章小结 80

第八章 逻辑学初步 81

8.1 从集合到命题 81
8.2 什么是和为什么需要逻辑? 82
8.3 逻辑三律和演绎逻辑 85
8.3.1 三段论 86
8.3.2 逆否命题的等价关系和反证法 86
8.3.3 逻辑变量的运算 88
8.4 逻辑是共通的普适的吗? 89
8.5 逻辑和批判性思维 89
8.6 逻辑和数学、逻辑和科学 91
8.7 推荐阅读材料 92
8.8 作业 92
8.9 本章小结 92

第九章 一元函数 93

9.1 一元函数当作一个吞吐机 93
9.2 一元函数当作一个关系 94
9.3 平面直角坐标系 94
9.4 一元函数当作一个图形 95
9.5 一元一次函数 96
9.6 一元一次方程的几何含义 98
9.7 一元二次函数和一元二次方程 99
9.8 负一次方函数的定义 102
9.9 为什么需要函数? 102
9.10 推荐阅读材料 102
9.11 作业 102
9.12 本章小结 102

第十章 二元一次方程和二元一次不等式 103

10.1 换元法：化二元一次方程为一元一次方程 104

10.2 消元法：等式的形式和代数式的计算 104
10.3 二元一次方程的几何含义 104
10.4 二元一次方程的应用 104
10.5 二元一次不等式组的求解 104
10.6 二元一次不等式组的几何含义 104
10.7 二元一次不等式组的应用 104
10.8 推荐阅读材料 104
10.9 作业 104
10.10 本章小结 104

第十一章 一元二次方程和一元二次不等式 105
11.1 一元二次方程的几何含义 105
11.2 配方法求解一元二次方程 105
11.3 因式分解法求解一元二次方程 105
11.4 一元二次方程的应用 106
11.5 一元二次不等式的几何含义 106
11.6 一元二次不等式组的求解 106
11.7 一元二次不等式组的应用 106
11.8 推荐阅读材料 106
11.9 作业 106
11.10 本章小结 106

第十二章 直角三角形和锐角三角函数 107
12.1 复习：勾股定理 107
12.2 点和距离的定义 107
12.3 直角三角形的等比例扩大 107
12.4 锐角的正弦余弦函数 109
12.5 三角形中的正弦定理 109
12.6 三角形中的余弦定理 109
12.7 三角函数的应用 109
12.8 圆和角的定义 109
12.9 推荐阅读材料 109

12.10 作业 109
12.11 本章小结 109

第十三章 平面几何：三角形的全等 111
13.1 三角形定义 111
13.2 三角形全等的定义和判定 111
13.3 为什么首先研究三角形? 111
13.4 尺规作图 111
13.5 全等三角形的应用 111
13.6 推荐阅读材料 112
13.7 作业 112
13.8 本章小结 112

第十四章 平面几何：平行 113
14.1 平行和垂直的定义 113
14.2 平面几何基本公理 113
14.3 平行线的判定和性质 113
14.4 推荐阅读材料 113
14.5 作业 113
14.6 本章小结 113

第十五章 平面几何：圆 115
15.1 圆的定义 115
15.2 圆中三角形的相关定理 115
15.3 共圆定理：把不是圆的问题变成圆的问题 115
15.4 弧度、弧长和扇形面积 115
15.5 推荐阅读材料 115
15.6 作业 115
15.7 本章小结 115

第十六章 平面几何：解析方式 117
16.1 代数表达式和几何图形 117
16.2 通过把几何问题转变成代数表达式来解决几何学问题 122

16.3 通过把代数表达式转变成几何问题来解决代数问题 123
16.4 推荐阅读材料 125
16.5 作业 125
16.6 本章小结 125

第十七章 平面几何：综合 **127**
17.1 平行线基本定理 127
17.2 三角形的相似 127
17.3 相似三角形的应用 127
17.4 数学解题 WHWM 四问和概念地图用于几何证明 127
17.5 几个平面几何知识体系赏析 127
17.5.1 Euclid 体系 127
17.5.2 Hilbert 体系 128
17.5.3 SMSG 体系 128
17.5.4 本书的体系 128
17.6 推荐阅读材料 128
17.7 作业 128
17.8 本章小结 128

第十八章 概率论初步 **129**
18.1 古典概型的例子 129
18.2 事件和事件集合的运算 129
18.3 概率：从事件集合到 $[0,1]$ 的映射 129
18.4 独立事件和互斥事件 129
18.5 初识条件概率 129
18.6 概率论的应用 129
18.7 推荐阅读材料 130
18.8 作业 130
18.9 本章小结 130

第十九章 统计学初步 **131**
19.1 描述性统计 131

19.2 统计推断和人类知识 131
19.3 从抽样获取整体性质 131
19.4 条件概率用于统计推断 131
19.5 统计学的应用 131
19.6 推荐阅读材料 131
19.7 作业 131
19.8 本章小结 131

第二十章 排列组合 133
20.1 组合 133
20.2 排列 133
20.3 从排列组合到概率 133
20.4 排列组合的应用 133
20.5 推荐阅读材料 133
20.6 作业 133
20.7 本章小结 133

第二十一章 逻辑学和数学论证 135
21.1 命题 135
21.2 论证 135
21.3 原命题和逆否命题的等价性 135
21.4 三段论 135
21.5 数学归纳法 136
21.6 不完全归纳法 136
21.7 推荐阅读材料 136
21.8 作业 136
21.9 本章小结 136

第二十二章 数学建模 137
22.1 从数学建模的角度看数学知识的发展 137
22.2 从把数学知识用于描述现实看真正的数学应用题 137
22.3 数学建模五步及其应用 137

22.4 推荐阅读材料 137
22.5 作业 137
22.6 本章和本卷小结 137

第二部分 高中篇 139

第二十三章 函数之幂函数、指数函数、对数函数 141
23.1 幂函数 141
23.2 指数函数 141
23.3 对数函数 141
23.4 函数和反函数的图形和表达式 141
23.5 推荐阅读材料 141
23.6 作业 141
23.7 本章小结 141

第二十四章 函数之半圆、半椭圆函数 143
24.1 半圆函数 143
24.2 半椭圆函数 143
24.3 圆 143
24.4 椭圆 143
24.5 双曲线 143
24.6 推荐阅读材料 143
24.7 作业 143
24.8 本章小结 143

第二十五章 函数之三角函数 145
25.1 余弦函数 145
25.2 余弦定理 145
25.3 正弦余弦函数关系 145
25.4 正切反切函数 145
25.5 周期函数 145
25.6 推荐阅读材料 145

25.7 作业 145
25.8 本章小结 145

第二十六章 平面和空间矢量 147
26.1 直角坐标系和矢量的分量形式 147
26.2 矢量加法和数乘 147
26.3 矢量内积 147
26.4 坐标变换和描述坐标变换的矩阵 147
26.5 矢量的抽象记号语言 147
26.6 矢量的应用 147
26.7 推荐阅读材料 147
26.8 作业 147
26.9 本章小结 147

第二十七章 复数 149
27.1 负数的开平方 149
27.2 负数的开三次方 149
27.3 复数当作矢量使用 149
27.4 Euler 公式 149
27.5 复数的加法和乘法 150
27.6 复数的应用 150
27.7 推荐阅读材料 150
27.8 作业 150
27.9 本章小结 150

第二十八章 不等式初步 151
28.1 几个恒成立不等式 151
28.2 从恒成立不等式到函数极值 151
28.3 不等式和极值的应用 151
28.4 推荐阅读材料 151
28.5 作业 151
28.6 本章小结 151

第二十九章 导数和极值初步 **153**
29.1 导数的定义初步 153
29.2 简单函数的导数 153
29.3 函数的导数用于求极值 153
29.4 导数和极值的应用 153
29.5 推荐阅读材料 153
29.6 作业 153
29.7 本章小结 153

第三十章 离散变量的概率论和统计学 **155**
30.1 从观测数据得到频率再到概率 155
30.2 从已知离散概率到观测数据的频率 155
30.3 用 Monte Carlo 和 Bootstrap 配合数值计算来沟通频率和概率 156
30.4 概率论和统计学的应用 156
30.5 推荐阅读材料 156
30.6 作业 156
30.7 本章小结 156

第三十一章 立体几何 **157**
31.1 空间直线的平行和垂直 157
31.2 空间平面的平行和垂直 157
31.3 推荐阅读材料 157
31.4 作业 157
31.5 本章小结 157

第三十二章 表示论：数学是对操作的研究 **159**
32.1 当作对象属性或者过程的数 159
32.2 数当作加法的表示 159
32.3 数当作乘法的表示 159
32.4 数和矩阵当作旋转的表示 159
32.5 数学是对操作的研究对关系的研究 159
32.6 把数学看作是操作的研究的意义 160

32.7 推荐阅读材料 160
32.8 作业 160
32.9 本章小结 160

第三十三章 数学知识网络和学科大图景 161
33.1 数学一二层知识的系统性 161
33.2 数学学科大图景，及其和一二层知识的联系 161
33.3 从数学知识体系到科学知识体系 161
33.4 高等数学学什么 161
33.5 推荐阅读材料 162
33.6 作业 162
33.7 本章小结 162

第三十四章 再论数学建模 163
34.1 从数学建模的角度看数学知识的发展 163
34.2 从把数学知识用于描述现实看真正的数学应用题 163
34.3 数学建模五步及其应用 163
34.4 推荐阅读材料 163
34.5 作业 163
34.6 本章和本卷小结 163

参考文献 165

名词索引 167

人名索引 169

插图目录 169

举例目录 170

作业目录 170

献给

吴立心、吴逸兮，以及所有的中学生们。愿你们永远可以体验到思考的快乐，讲道理的快乐，把知识搞通透的快乐，运用知识解决问题的快乐，而不是被教“就是这样，你记住”，“你只需要会做题”。

致谢

序

前言

本书是《小学数学这样学》[1] 的后续篇，也是“人类知识高速公路上以高层知识生成器为目标的理解型学习”（简称“理解型学习”[2]）应用于具体学科的另一个例子。在这里我们把中学，包含初级中学和高级中学，的数学知识做了梳理，用新的方式做了反映数学学科大图景的呈现。一个学科的学科大图景指的这个学科的典型研究对象、典型研究问题、典型思维方式、典型分析方法。以及这个学科和世界以及其他学科的关系。在这个新的呈现里面，我们强调了学习是为了创造知识、创造性地使用知识，以及欣赏知识的创造和创造性使用，并且企图通过理解型学习来通过具体知识的学科之中体会好数学学科大图景，从而达成上面的学习的目的。

我们来补充一点点关于理解型学习的知识。在这个学习方法之中，最重要概念是知识的 $1+5$ 个层次：第零层前知识，经验和体验；第一层，事实性程序性知识；第二层，学科概念知识；第三层，学科大图景知识；第四层，一般性人类思维；第五层，教和习的方法。在学习三角形之前就看过摸过三角形，在学习加法之前就在生活中把两堆东西和起来数过，这成为经验和体验。乘法口诀、三角函数公式、题型和解题步骤口诀属于第一层知识。乘法是重复加法的简便记号、三角形是三条线段首尾相连构成的封闭图形、三角函数的定义以及各个三角函数之间的关系属于第二层知识。数学是思维的语言，加法来自于对“合起来数一数”的抽象，数来自于对“数 + 量 + 物”的模式的抽象，每个题型及其解题步骤的梳理总结的过程（提出或者被提出问题；找到这个问题中的对象，识别对象之间的关系，把关系转变成数学表达式，把所要解决的问题变成数学问题；求解问题；检验答案；进一步推广到一类问题。有的时候也被称为数学建模五步，或者简称数学五步 [1]）属于第三层知识。例如从加法（和起来数一数，从经验和体验到学

科概念）和乘法（重复的加法，为了更加简便的语言和记号）的概念的定义方式体会到抽象这个思维方式在数学中的重要性，体会到概念之间的联系，体会到不同层次知识之间的联系，进一步看到系联性思考，从证明中体会到批判性思维（在数学上，任何公理以外的命题必须经过我自己的严格证明，才能被我接受），进而把抽象、系联性思考、批判性思维迁移到其他学科的知识的理解和创造之中，属于第四层知识。通过知识之间的联系从而实现上下左右贯通来学习，通过具体知识的学习体会到学科大图景，以及从是否能够把学科大图景体现好的角度来看每一个具体知识的学习，是有关“理解型学习”这个教和习的方法的知识（甚至，从这样通过具体数学知识的学习体会到理解型学习，从理解型学习的角度来设计具体数学知识的学习，也是理解型学习的例子），属于第五层知识。

更进一步，我们把这样的通过不同层知识之间的关系来学习和思考，也就是从对相对低层知识的总结提炼中学习、体会甚至创造相对高层知识，从相对高层知识来指导相对低层知识的学习、理解和创造，称为上下贯通；我们把通过同一层知识之间的联系来理解好这一层知识，以及把来自于一个领域的高层知识用于学习着解决或者创造着解决另一个领域的问题称为左右贯通。我们可以借助概念地图（概念和概念通过连词相连，例如“加法–是–合起来数一数”、“合起来数一数–基于–数”、“合起来数一数–基于–数数”、“数–来自于–对‘数 + 量 + 物’这个模式的抽象–体现了–抽象”）来呈现和运用上下左右贯通。各个学科的概念地图的总和被称为人类知识高速公路。正是由于第三层和第四层知识具有帮助解决本领域或者其他领域问题，创造知识的能力，我们把它们称为高层知识生成器。

因此，知识的层次中的第三、四、五层知识，也就是高层知识生成器；人类知识高速公路，也就是各个学科的概念地图的总和；上下左右贯通，也就是从低层知识中总结体会到高层知识，从高层知识来看低层知识，运用高层知识来学习着或者创造着解决本领域或者其他领域的问题从而学习或者创造知识，是理解型学习最重要的核心概念和理念。更多信息可参见吴金闪的《教的更少，学得更多》[2]。

因此，本书的第二个目的就是用理解型学习来帮助学习者学科中学数学，从而也学会理解型学习，进而用理解型学习来学习其他学科，以及给创造和创造性使用各个学科的知识做好准备。

在这样的一个定位和写作意图的过程中，我们设想的直接的读者是教师、家长和教育管理者。希望这些读者能够从本书所选择的教什么、怎么教之中体会到理解型学习，进而来开展理解型学习指导下的课程教学、课程设计和考察方式。但是，为了达到这个目的，我们希望能够通过“做中学、创造中学”来帮助教师、家长和教育管理者，也就是从看如何把理解型学习的理念和工具用于实际课程的设计来领会理解型学习，而不是理解型学习的理念和概念的直接传达。因此，我们打算通过在理解型学习的指导下来具体编撰这样一本中学数学教材，当例子，来帮助这些读者得到更大的收获，而不是被说教。

另一方面，我们也希望写出来一本中学数学教材，能够帮助中学生对数学具体知识的理解更透彻，甚至能够进一步理解到数学的学科大图景，学会像一个数学家一样来思考，或者至少能够欣赏数学家的思考，而不是仅仅学会了一堆算来算去的知识。这就要求我们第一、把数学材料梳理好，突出联系，可以帮助学习者更容易更透彻地实现上下左右贯通，看到数学学科大图景；第二、帮助学习者学会理解型学习，甚至在其中一部分知识的学习上，设计一些问题来做“创造体验式学习”——通过“假装着”把所要学习的知识创造出来而学会这个知识，从而更高地为将来创造知识做准备。

因此，本书的前言分成了“写给中学生们的”和“写给教师、家长和教育管理者的”两个部分。请读者们各取所需。

写给中学生们的

上面是写给中学生们的前言。下面是写给教师、家长和教育管理者的前言。

写给教师、家长和教育管理者的

本书所体现的数学学科大图景

前面我们提到了本书的核心思想通过“人类知识高速公路上以高层知识生成器为目标的理解型学习”来实现“教和学的更少，学得更多”。对于

数学学习，这里的高层知识生成器指的是第三层知识数学学科大图景和第四层超越学科的一般性人类思维，以及教和学的方法。为了实现这个“人类知识高速公路上以高层知识生成器为目标的理解型学习”，我们大概需要完成以下步骤 [2]：

1. 梳理这个学科的学科大图景；
2. 选择体现这个学科大图景的核心概念、概念间关系来构建学科概念网络；
3. 选择合适的案例来体现每个概念的提出和运用，促进对概念的理解；
4. 标注书籍、论文、项目、案例、习题等到这个学科概念网络上；
5. 结合概念网络上的分析算法（例如学习顺序算法、自适应诊断性检测算法）来设计和开展教学实践。

这其中最关键的就是对学科大图景的梳理。那么，数学学科大图景，也就是数学的典型研究对象、典型研究问题、典型思维方式、典型分析方法、和世界还有其他学科的关系，到底是什么呢？我们在《小学数学这样学》[1] 中有专门的一章来展开讨论。在这里，我们把其中的主要意思复述一下。更完整的版本请参考吴金闪的《小学数学这样学》[1]。

- 数学观
  - 数学是思维的语言
  - 数学是描述世界的语言，“双向”（从现实提炼数学结构，把数学结构用于描述现实）数学建模
  - 数学是研究关系和模式的学科
- 数学思维
  - 抽象是数学的核心思维
  - 批判性思维在数学中的地位：证明和计算是数学最重要的论证方式

  - 表示数学结构和研究表示间的转换是的数学研究的目标和研究方式
  - 分解和综合是数学结构研究的典型方式
  - 数学知识有系统：公理化体系，层次性和跨越层次的联系，不同数学分支的融合
  - 化归的思想

- 分析方法
  - 数学建模五步
  - 数学解题 WHWM 四问，及其嵌套合并
  - 试错和分析错误是数学的重要分析方法之一
  - 近似是数学重要分析方法之一
  - 微积分——累积量和变化率的关系
  - 无穷小和极限的概念
  - 连续性及其分析方法
  - 变换和不变量的思想和相应的分析方法
  - 约束下求极值

- 最基础的概念（其他高级结构的基础）
  - 集合和映射
  - 范畴论
  - 数理逻辑：什么是演绎证明，为什么能用

本书会围绕以上内容来决定教什么怎么教，在可能的情况下把数学学科大图景体现在具体知识的学习之中，把理解型学习体现在数学的学习之中。

其实，这个具体知识和学科大图景以及超越学科的一般性人类思维还有教和学的方法的关系，就是系统思维 [3]：通过直接和间接联系，从个体看到整体，从整体的角度来看个体。我们把它叫做“系联性思考”。不仅仅

是学科学习，在面对任何系统的任何问题的时候，系联性思考——也就是通过把这个系统在各个层次做分解和综合，看到在各个层次上各个部分之间的联系，进而从整个系统的整体功能的角度来看其每一个部分的作用和地位，从这个系统的每一个部分来看到这个系统的整体的功能——往往可以给我们提出和解决问题的启发。而且，更进一步，系统科学还在此思维方式的基础上，发展了一些分析方法，解决了一些典型问题 [3]。因此，我们也希望通过这本书，你不仅仅学会数学，体会到数学学科大图景，学会理解型学习，还能体会到系统思维的核心 “系联性思考”，而且都是通过做中学、创造中学，而不是通过这些高层知识的被说教，或者被强迫记住，来学习。

第零章　对小学数学的要求和复习

在我们开始中学数学的学习之前，我们希望你最好有下面的小学数学的知识和思维的基础。如果没有这个基础，你可以先试着学下去，遇到问题了再想办法补上，或者你也可以先去学习《小学数学这样学》[1]，然后再来看本书。

0.1　对小学数学的要求

首先，我们希望你知道的第一、二层知识有：

1. 认识非负整数（正整数和零），会数数；
2. 会整数的加法和乘法的计算，理解整数的加法和乘法的含义；
3. 会整数的减法和除法的计算，理解整数的减法和除法的含义；
4. 整除、因数、倍数、最大公约数、最小公倍数的概念和简单计算；
5. 从非负整数通过不断的减法拓展到包含负的整数的所有整数；
6. 通过除法从整数拓展到分数，以及分数的四则运算；
7. 从分数到等价的循环小数（包含有限小数——循环位是 0，和无限循环小数），以及小数的四则运算；
8. 平方、立方，开平方，开立方的含义和简单计算；
9. 从 $\sqrt{2}$ 不是分数，初步了解无理数（无限不循环小数）的存在性；

10. 可以用十进制来表达数，通过数十进制表达来理解乘法竖式计算；
11. 整数、分数、循环小数的数轴表示；
12. 加法和乘法的运算律，以及通过运算律来给算式做变形；
13. 用字母表示数，从关注算出来的得数，到关注代表事物的数量的字母之间的关系，以及关系对应的数学操作；
14. 等式的性质；
15. 通过等式的性质来求解一元一次方程；
16. 三角形、四边形（包含其特殊情况梯形、平行四边形、长方形、正方形）的内角、周长、面积的定义和计算，圆的周长和面积的定义和计算；
17. 长方体、正方体的表面积和体积的定义和计算；
18. 具有等几率基本事件（古典概型）的事件的概率的定义和简单计算；
19. 最基本的描述统计（均值、直方图）的定义和计算；
20. 把以上知识用于求解应用题，或者自己提出和解决日常生活中的问题。

但是，更加重要的是，获得这些第一二层知识的过程，或者说这些一二层知识之间的联系及其背后的三四层知识。具体来说，

1. 从“数 + 量 + 物”经过“抽象”，也就是忽略“量”和“物”（从而适用于任何“量”和“物”），而得到非负整数；
2. 更进一步从上面的非负整数的抽象过程，领会到数学概念、数学思维和现实的关系——受现实启发；
3. 从“合起来数一数”到加法，从重复多次的加法到乘法，体现了数学对重复模式的抽象：抓住重复模式的核心并给个名字；
4. 建立加法和减法，乘法和除法之间的互为逆运算的关系，进一步对比从加法到乘法、从减法到除法的这两个逻辑关系的关系；

5. 从加减乘除的定义、相互关系，体会到逻辑关系在数学中的核心地位，体会到数学知识的系统性；

6. 从整个非负整数到包含负整数的所有整数、分数、小数的过程，体会到运算的封闭性的重要性，结合出现这些运算和数的场景，体会到数学和现实的关系——现实使得这些运算会出现，于是要求数学给这些运算的结果一个表示；

7. 从整数和小数的加减乘除四则运算总结出来有且只有加法和乘法的四则运算规律的过程，体会到数学知识的系统性，体会到抽象、模式发现和总结这样的思维在数学中的重要性；

8. 从分数和循环小数的等价性的证明中体会到逻辑思维在数学中的重要地位；

9. 从把运算封闭性运用于平方和平方根从而得到 $\sqrt{2}$ 的含义和存在性的过程，体会到上下贯通，也就是从之前的学习中得到了运算的封闭性，现在把运算的封闭性用于新的运算；

10. 从把运算封闭性运用于平方和平方根从而得到 $\sqrt{2}$ 的含义和存在性的过程，结合实际场景（面积等于 2 的正方形的边长），体会到数学和现实的关系——数学需要给现实提供描述它的数学结构；

11. 从 $\sqrt{2}$ 不是分数的证明中，以及从 $\sqrt{2}$ 不是分数这个事实到提出无理数的过程中，中体会到逻辑思维在数学中的重要地位；

12. 通过可以用数的十进制来理解乘法竖式计算步骤进一步领会到数学知识的系统性；

13. 通过数的数轴表示初步体会到数和形这些看起来不同的数学对象之间的关系，从而更进一步体会到左右贯通——不同领域之间的知识之间的关系；

14. 依托于从四则运算中抽象总结出来四则运算律及其字母表示的过程，更进一步地理解抽象在数学中的重要地位；

15. 从用字母表示任意的数代入算式反而集中体现了算式中的运算，也就是关系，这一点，体会到数学开始从关注数，到开始关注关系；

16. 从运用等式的性质来解一元一次方程，进一步领会到数学知识的系统性；

17. 从几何初步进一步领会到数学知识的系统性，而且是从数的知识到形的知识的迁移，左右贯通这个系统性；

18. 在求解数学问题过程中，不依靠题型和口诀，而是每次去做数学解题 WHWM 四问：这个问题里面有哪些东西（What，有什么），这些东西之间具有什么关系（是如何相互联系的，How），这个联系如何用数学来表达（How），为什么它们之间有这个关系（Why），为什么这个关系对于解决这个问题是重要的（Why），通过解决了这个问题我学习到了什么体会到了什么，这个问题对我来说意味着什么（Meaningful）；

19. 通过把数学用于提出和解决实际生活中的问题，以及前面提到的这些数学概念和数学思维的提出和发展的过程，领会到数学建模五步：提出有关实际对象的问题，把问题转化为数学问题，求解数学问题，通过实验或者实践来检验得到的答案，对问题、答案、解答、这个过程中的分析方法思维方式做总结提炼然后尝试系统化和推广；

20. 从以上的证明、抽象、解决实际问题等过程中体会到数学是思维的语言，数学是描述世界的语言。

如果更进一步，你已经掌握了用包含字母的算式来表示运算律，并且还能够从以下几条基本的运算律“加法结合律、加法交换律、乘法结合律、乘法交换律、乘法对加法的左侧分配律（乘法在左，加法在右）”推导出来（也知道推导是什么意思，也就是依靠定义和更加基本的已经知道是正确的命题，加上逻辑，来得到新的命题）更多的运算律，例如乘法对加法的右侧分配律，见例 0.1，那就更好了。如果你还进一步掌握了集合的概念，明白了实际上证明的过程依靠的是逻辑学当中的叫做演绎逻辑的东西，而这个东西的背后是集合那就更更好了。不过，这些知识——包括前面的列表中的一二层和三四层知识，以及这里的字母表示数和算式、带字母的算式表

示运算律、从定义到基本运算律、从基本运算律到导出运算律（以及背后的知识的系统性）、演绎逻辑用于证明数学命题、演绎逻辑和集合的关系、集合我们都会在本章以及本书后续章节中再一次学习。

例 0.1 (乘法对加法的右侧分配律). 已知加法结合律、加法交换律、乘法结合律、乘法交换律、乘法对加法的左侧分配律，证明乘法对加法的右侧分配律 $(a+b)\times c = a\times c + b\times c$。

证明.

$$
\begin{aligned}
(a+b)\times c &= c\times(a+b) && \text{(乘法交换律)} && (1)\\
&= c\times a + c\times b && \text{(乘法对加法的左侧分配律)} && (2)\\
&= a\times c + b\times c && \text{(乘法交换律)} && (3)
\end{aligned}
$$

□

0.2 小学数学复习

这一章，我们从重新构建整个小学数学学科概念知识以及学科思维等学科大图景的角度来复习一下小学数学。如果通过这个复习，你发现仍然有不太能理解的地方，请参考《小学数学这样学》[1]。

0.2.1 从数数到数和运算律

从数数的经验体验，到自然数，非负整数，所有整数，分数，小数，无理数的存在，实数初步；从概念到基本运算律，到导出运算律；用字母表示数，用字母表示算式（算式是数）。

我们假设咱们的读者已经具有数 $1,2,3$ 三个数的经验和体验，也就是看到篮子里有一个苹果，知道对应的符号是 1，单位是个；看到盘子里面有两块饼干，知道对应的符号是 2，单位是块；看到桌子周围坐了三个人，知道对应的符号是 3，单位是个；看到餐桌上放了三双筷子，知道对应的符号是 3，单位是双，等等等等。从这个朴素的数的概念开始，我们每次给前面的数增加 1 来获得一个新的数——指的是含义上。知道含义之后，我们给

每一个这些新增加的数一个名称，例如 $3+1$ 记作 4，读作四；$4+1$ 记作 5，读作五；$5+1$ 记作 6，读作六；$6+1$ 记作 7，读作七；$7+1$ 记作 8，读作八；$8+1$ 记作 9，读作九。原则上，我们还可以有 $9+1, 9+1+1, \cdots$。但是，我们稍后再给它们一套记号和读法。也就是说，所有的自然数的含义都可以基于 $1, 2, 3$（其实，$2, 3$ 都不用，只需要知道朴素的 1 的含义和加 1 的朴素含义——一个数后面加上 1 所代表的对应到实际对象的含义，不需要知道严格的 1 的定义和加法的定义），但是 9 以后的自然数的符号和名称我们会稍后再介绍。

有了 $1, 2, 3, \cdots, 9$ 这九个符号以后，我们来看这些符号在记号、比大小和加法上如何发挥作用。首先，我们说，有了每一个数所代表的在特定单位下的某种东西的数量的含义，我们就可以在同样的单位下把同质的东西和起来看数量。例如，一篮子里面装了 2 个苹果，另一个篮子里面装了 3 个苹果，把两篮子苹果装在同一个篮子里面，就得到这个篮子里面有 5 个苹果。在这里，实际上，我们都不用数数：我们把 5 的含义——在“个”当作单位来看“苹果”时候的数量——用于这个合起来的篮子里面的苹果，则自然就是 5 个，然后，我们发现，实际上这个两个篮子合到同一个篮子里面的操作，就是把两堆苹果合并。当以后，这样的合并的操作我们见多了，就会发现，背后的数之间存在着脱离开了对象的关系——$2+3=5$。这个表达式，无论是苹果，还是饼干，还是筷子，无论是相应的个、块还是双，都是成立的。于是，我们把这个对应着合起来以后再看数量的操作定义为加法。

有了加法，也就是合起来数一数，自然我们就可以考虑反过来的操作，整体中去掉一部分，被称为减法。减法也是可以超越对象和单位的，只要坚持同一个单位同质的对象。例如，从一个装了 3 个苹果的篮子里面拿出区 2 个苹果，则剩下的是一个苹果，算式记作 $3-2=1$。同时，我们也能够理解，减法是加法的逆运算，反向操作——因为把剩下的和拿走的合起来的数量自然就是一开始的整体的数量。

一旦有了加法我们就可以不断地加下去，所以，我们可以在含义上得到从 1 开始的所有的自然数。类似地，一旦有了减法，我们就可以不断地减下去，从而得到从例如 3 开始的所有的数。我们试试：$3-1=2$，$2-1=1$，$1-1=?$。我们发现，我们还没有这个 $1-1$ 的含义的记号和读音。我们给这个含义发明一个记号和读音，记作 $1-1=0$，读作零。我们其实可以继

续减下去，例如 $0-1, 0-1-1, \cdots$。于是我们就有了负的整数。对应到显示当中，负数当作具有参考点和方向的数的表示是最好理解的。例如我们规定某个温度叫做零度，则比零度大的用正数表示，比零小的用负数表示。例如，我们规定某个高度叫做零点，更高的用正数，更低的用复数。例如，我们如果规定往东为正方向用正数表示，那么往西就用负数表示。一旦有了负数，则减法就也成了加法，相当于加上一个负数。例如，$4-2=4+(-2)$。

有了加法，我们可以从加法的含义得到加法的结合律和交换律：对于合起来数一数，如果要把三堆东西合起来数那么先把哪两堆合起来数一数不影响得到总数；如果是把两堆东西合起来数一数，则先数清楚哪一堆也不影响得到的总数。例如，$(1+2)+3=6=1+(2+3)$，$2+3=5=3+2$。这些运算律，在将来我们学习了用字母表示数（任何一个数都可以代入到这个字母的位置上去）之后，就可以表示为 $(a+b)+c=a+(b+c)$，$a+b=b+a$。

有了整数和加法（注意减法已经通过负数统一为加法），我们来看乘法。当我们需要把多个同样数量的对象合起来数一数的时候，我们可以给这个“重复多次的的加法”一个自己的名字——乘法。例如，一箱子羽毛球，每个箱子里面有 50 筒，每筒有 12 个羽毛球。我们可以算 49 次加法来得到羽毛球的总数，也可以用这个新的计算——乘法来表示和得到相应的规则，然后重复使用这个规则来完成羽毛球总数的计算。这个规则在个位数乘以个位数的时候被称为乘法口诀表。例如 $2\times 3=4$（二三得六），$3\times 7=21$（三七二十一）。这些口诀本身是通过加法来得到的，例如 $3\times 7=7+7+7=21$。

有了乘法，我们可以从乘法的含义得到乘法的结合律和交换律：$(a\times b)\times c=a+(b+c)$，$a\times b=b\times a$。在这里我们直接用字母来代笔了一个任意的整数[1]。乘法的交换律可以通过把乘法算式 $a\times b$ 看作是一个平面上有排好的 a 行 b 列个同样的东西要合起来数一数来理解。这样一个平面上的所有东西，也可以通过两种方式来计算，得到相同的结果：第一种，把每一行看作一个单元——每一行里面有 b 个，总共有 a 行，于是总数为 $a\times b$；第二种，把每一列看作一个单元——每一列里面有 a 个，总共有 b 列，于是总数为 $b\times a$。插图解释交换律和结合律 。乘法的结合律可以通过把乘法算式 $(a\times b)\times c$ 看作是一个三维的长方体上有排好的 a 行 b 列 c 层个同样的东西要合起来数一数来理解。这样一个长方体上的所有东西，也可以通过两

[1] 未来会发现拓展到分数小数之后，这些运算律仍然成立

种方式（其实还有第三种）来计算，得到相同的结果：第一种，先计算 a,b 所在的面，总共有 $a\times b$ 个，然后这样的面有 c 层，于是总数为 $(a\times b)\times c$；第二种，先计算 b,c 所在的面，总共有 $b\times c$ 个，然后这样的面有 a 层，于是总数为 $a\times(b\times c)$。如果你还是不习惯同字母表示数，你可以通过把字母换成具体的数的方式来理解上面的论证过程。

顺便，我们来看看乘法对加法的左侧分配律，也就是：$a\times(b+c)=a\times b+a\times c$。这同样可以通过乘法和加法的含义来理解。我们这样来看 $a\times(b+c)$——每一个单元里面有 $(b+c)$ 个同样单位下的同样的东西，这样的单元有 a 个。自然把每一个单元拆分为 a 个东西和 b 个东西这两个小单元，然后分别计算 a 份这样的两种小单元内的东西的数量，再合起来，得到 $a\times b+a\times c$。自然，这样的拆分不影响总量。

有了乘法，我们可以来定义乘法的逆运算——除法。也就是，如果 $3\times 7=21$，那么，我们就可以问 $3\times?=21$ 呢？答案就是 7，记作 $21\div 3=7$。类似地，$2\times 3=6$ 用除法表示就是 $6\div 3=2$。我们把 $2\times 3=6$ 对应到现实场景——把两个每个篮子里面装了 3 个苹果的篮子合起来的苹果的总数，把 $6\div 3=2$ 看作是，把总数是 6 的苹果分成每个篮子里面放 3 个苹果可以放几个篮子，或者说每次从 6 中取出来 3 可以取多少次。从另一个角度，我们可以把 $6\div 2=3$ 看作把 6 个苹果平均放到两个篮子里面，每个篮子里面会有几个苹果。把这个平均分的问题和前面的每次取出多少可以取多少次联系起来的办法是：每次我们取出来足够的苹果给每个篮子里面放一个，由于总篮子数量是 2 则，每次需要取出来 2 个。这就回到了从整体中每次取出一部分可以取多少次的问题。于是，就可以用除法来计算了。

当然，反过来，把每个篮子里面相同数量的苹果合起来数一数，自然就是乘法。于是，乘法和除法就互为逆运算。注意，我们约定除法的除数不能为零——无论多少个零加起来还等于零，不会等于一个非零的数，因此 $\frac{q}{0}$ 当 $q\neq 0$ 的时候没有意义[2]。

从另一个角度，我们来看 $2\times 3=6$。我们把这 6 个苹果当成整体，我们说，每一个篮子里面的苹果，占这个总苹果的多少呢？我们发现，这里有两个篮子，其中的一个篮子自然就是占两个之中的一个。把一个整体分成 n 份，取其中的 m 份，就记作 $\frac{m}{n}$。例如这里的 $\frac{1}{2}$。$\frac{m}{n}$ 也可以看作是 $m\times\frac{1}{n}$

[2]未来会学习到无穷大，以及 $\frac{0}{0}$ 的情形。

——从整体 n 份取出 m 份相当于是 m 次重复地取出来整体 n 份中的 1 份再合起来。这样我们就有了分数的定义，包含真分数 $m < n$ 两者都是整数，以及假分数 $m > n$。其中假分数相当于重复取 n 份中的 1 份的次数非常多，多过了 n 次。这个时候，我们需要想象以下拿出来另一个整体也分成 n 份，然后来满足这个取 m 次（$m > n$）的要求。

下面，我们来把分数和除法联系起来——$\frac{q}{p} = q \div p$ $(p \neq 0)$。我们以假分数并且可以整除为例，不能整除和真分数的的情形道理相同但是论证过程会稍微复杂一点。$q \div p$（$q > p$ 并且可以整除）有一个性质 $(q \div p) \times p = q$。这是上面的乘法和除法就互为逆运算的直接结果。或者说，先把一个整体（q 个东西）平均分成 p 份，得到每一份的数量，然后再取这样的数量的东西 p 份，合起来自然还是 q 个东西。$\frac{q}{p}$ 也具有同样的性质：$\frac{q}{p} \times p = q$。我们把一个整体分成 p 份取其中的 q 份，然后把 p 个这样的份合起来，自然就是 p 个原始的整体。我们可以这样来理解：先把 p 当作 1，自然 $\frac{1}{p}$ 就是一个整体的 q 份中的 1 份，然后我们现在有 q 份合起来，自然就得到了整体；接着，如果我们一开始取走的不是一份而是 p 份，则我们需要把前面取走一份的操作重复 p 次合起来，自然就得到原始的整体的 p 倍。现在我们得到两个对象具有同一个性质：$(q \div p) \times p = \frac{q}{p} \times p$，两个数乘以同一个不等于零的数得到的得数相同，则这两个数也相同。

现在，除法和分数已经联系起来了，我们说，除法和乘法就统一了，$q \div p = \frac{q}{p} = q \times \frac{1}{q}$。这样任何一个除法算式（只要分母不为零）就可以转化为乘法算式了。我们也可以来计算分数乘法了。例如，$\frac{q}{p} \times \frac{m}{n} = q \times \frac{1}{p} \times m \times \frac{1}{n} = m \times q \times \frac{1}{p} \times \frac{1}{n} = (m \times q) \times \left(\frac{1}{p} \times \frac{1}{n}\right) = (m \times q) \times \left(\frac{1}{p} \times \frac{1}{n}\right) = (m \times q) \times \left(\frac{1}{p \times n}\right) = \frac{m \times q}{p \times n}$。其中，我们用到了 $\frac{1}{p} \times \frac{1}{n} = \frac{1}{p \times n}$。这个可以从分数的含义上来理解：把一个整体先分成 p 份取其中的 1 份，然后，把这个取好的部分再看作整体分成 n 份取其中的 1 份，自然相当于把原始的整体分成了 $p \times n$ 份取其中的 1 份。

分数的加法在同样的分母下就很容易计算，相当于直接计算分子的加法。对于分布不相同的分数，我们只需要把分母变成相同的数——称作通分，就可以回到分母相同的情形。通分需要计算整数的最小公倍数，以及需要分数的性质——分子分母同时乘以一个数分数值不变，也就是 $\frac{q}{p} = \frac{q \times a}{p \times a}$ $(a \neq 0)$。回到分数和乘法的统一，我们有 $\frac{q \times a}{p \times a} = q \times a \times \frac{1}{p \times a} = q \times a \times \frac{1}{p} \times \frac{1}{a} =$

$q \times \frac{1}{p} \times a \times \frac{1}{a} = q \times \frac{1}{p} \times \left(a \times \frac{1}{a}\right) = \frac{q}{p}$。反过来，如果我们的分数的分子和分母之间存在这公因数，则我们需要通过计算最大公因数的方法来对分数作化简，得到分子和分母没有公因数的最简分数。其原理仍然是分子分母同时乘以一个不等于零的数，分数值不变。

这样我们就有了分数的加法和乘法的计算方法，也有了分数就是除法（分数线就是除号）。于是，除法就通过分数成了乘法，也就是 $q \div p = q \times \frac{1}{p} = \frac{q}{p}$。

这样，我们就从数一二三的经验开始构建了整数、整数的加减乘除、负数以及通过负数把减法统一到加法、分数及其加减乘除运算以及通过分数把除法统一为乘法。对于分数，或者说除法，我们可以通过除法计算规则把分数转化为一个包含整数部分和小数部分，中间通过小数点隔开的小数。小数的计算可以采用两种方式。一种是参考整数的计算，也就是基于数位对齐来作计算。一种是转化为分数来作计算。如果所有的小数都可以转化为分数，那后者就可以用来计算所有的小数运算。但是，我们发现，所有的分数都可以转化为有限小数或者无限循环小数（这个可以通过除法计算得到[3]），但是，并不是所有的小数都可以转化为分数——那种无限不循环小数不能转化为分数，并且这样的数存在，例如 $\sqrt{2}$。

分数实际上在更小的单位下就可以看作是整数。例如 0.1 元是 1 角，0.2 千克是 200 克，$\frac{1}{4}$ 元是一个四分元（Quater）如果我们引入“四分元（Quater）”这个货币单位的话。因此，既然分数可以转变为整数，很容易可以理解之前证明过的加法结合律、加法交换律、乘法结合律、乘法交换律、

[3]举例来说，$\frac{a}{3} = a \div 3$。我们忽略整数部分，直接从小数部分开始。这里的余数有三种情况 $0, 1, 2$。第一种情况，整数部分除法得到的余数为 0。那直接就没有小数部分，于是得到一个有限小数。第二种情况，整数部分除法得到的余数是 1，那需要计算下一位小数——通过在被除数后面加零来计算下一位，也就是得到 $10 \div 3$。这个时候，余数还是有 $0, 1, 2$ 三种可能。只要这里出现的余数之前出现过，那么，我们往后得到的的商和余数就都相同。实际上我们得到余数 1，已经出现。因此，后面继续除下去得到的商和余数都会重复。第三种情况也一样，整数部分除法得到的余数为 2。计算小数部分得到的余数也是 2，重复出现。实际上，更一般地，每一次相除得到的余数都会除数，因此只有有限多个可能，只要不断地计算小数部分，余数总有一天要么是零，要么会重复出现。因此，分数都会成为有限小数或者无限循环小数。为了语言简单，我们可以把有限小数——循环了 0 和无限循环小数合起来成为循环小数。实际上，还可以证明循环小数必定可以表达为分数。见《小学数学这样学》[1]。

乘法对加法的左侧分配律在分数上仍然成立。但是，由于 $\sqrt{2}$ 这样的数无限不循环不能写成分数，因此在任何一个单位下，这个数都是无限长的。这就使得这样的无限不循环小数很难用单位变换来理解。历史上，人们把可以通过单位变换联系到整数的数称为有理数，不能的成为无理数。大约就是不可理解的数的意思。

无理数和有理数合起来称作实数。证明实数上仍然有加法结合律、加法交换律、乘法结合律、乘法交换律、乘法对加法的左侧分配律就不是这么简单的事情了。将来大学阶段以及以后可能会学习到。但是，我们暂时接受这些运算律仍然成立。

有了数和运算律，我们还缺一个等式的性质，也就是：等式的两边同时作加法和乘法（包含除以同一个不为零的数）等号仍然成立。而等式的性质的理解就特别简单了：无论一个算式多么的复杂，等号成立，意味着左边的得数等于右边的得数。对于已经相等的两个数，或者说实际上同一个数，则无论对这个数作什么样的运算，只要这个运算合法而且结果唯一，得到的结果左右两边仍然相同。

最后，运算律和等式的性质是对任何数都成立的，因此有必要从以具体数为例的表示过渡到可以适用于任何数的表达式。为了这个目的，我们用字母来表示任意的一个数，就有了字母所表示的运算律和等式的性质。在中学阶段，这个用字母表示的运算律和等式的性质是我们数学学习的起点。

0.2.2 现实和数学，从关系到数学结构

数学 WHWM 四问，数学五步，数学建模

0.2.3 数学知识的系统性

集合、逻辑（三段论、反证法）、证明，数学知识网络

集合是一个满足下面的一个叫做“确定性”的要求的一堆东西：给定任何一个东西，我们必须能够明确地判断这个东西是否属于这一堆。未来在大学阶段我们会学习更加抽象的定义。在那之前，我们先采用这个描述性定义。我们还给这个描述性定义的核心“是否属于”一个数学符号：属于用 $\in$，不属于用 $\notin$。例如，1 是自然数之一，也就是属于自然数，记作“$1 \in \mathbb{N}$”（其

中 $\mathbb{N}$ 代表了所有自然数)；1 不是质数 (也不是合数)，记作 “$1 \notin \mathbb{P}$” (其中 $\mathbb{P}$ 代表了自然数中的所有质数)。于是，我们可以用这个符号来描述集合：一个集合 A 就是给定任何对象 x，我们都可以得到 $x \in A$ 和 $x \notin A$ 有并且只有一个成立。

这个确定性的要求其实很高。例如，秃头、高个子、好苹果、水果等等都不是一个集合。比如，尽管我们大多数时候可以对于一个人是否属于秃头、高个子达成共识，但是也会经常出现分歧。甚至有些词语本身的含义就是不够明确的，例如，智慧、聪明、才干、气质等等 (注意这里的 “等等” 的含义，原则上本来是应该清楚的，也就是那些含义不明确的词语，但是在生活中往往 “等等” 不一定能够具有这样明确的意思，甚至 “含义不明确” 这个短语本身的含义也不一定是明确的)。那么，什么才是真的明确呢？最严格的标准就是可操作可测量的。例如，如果我们规定只有头发少余 1000 根的叫做秃头，身高大于 180cm 的叫做高个子，那没问题，就是明确的。类似地，我们可以说，智商测量量表不一定测量的就是聪明程度，而且不同的量表可能测出来的东西和数值都不一样也不能转换，但是，如果我们忽略这些问题，仅仅用某一套量表来测量，然后规定一个阈值，我们就有了 “聪明人” 的明确含义，就可以构成集合了。因此，实际上，集合背后是可操作可测量。

如果我们把集合 A 包含的所有的对象——称作集合 A 的元素——写下来，我们用下面的记号 $A = \{a, b, c\}$。一般来说，我们用大写字母表示集合，小写字母表示元素。例如，一个骰子扔下去可能出现的向上的面的集合记作 $D = \{1, 2, 3, 4, 5, 6\}$。注意这里的六个数仅仅是记号，表示哪一个面向上。如果我们约定每个出现的面可以获得和面的符号数量相同的钱 (一人民币元为单位)，那么，这里的六个数就有了可以比大小和作加减运算的含义。

我们还可以通过一个叫做 Venn 图插入图，用所有自然数，骰子等集合当例子，做一个用点的，一个用点加上记号的 的东西来把集合可视化。我们画一个圆圈表示集合 A，用点 (可以带上记号) 来表示集合里面的元素。为了表示属于和不属于集合 A，我们画一个全集 I，其中的一部分被包含在 A 之中，另一部分没有被包含在 A 之中，称作 A 的的补集 $\bar{A}$。其含义是 $\bar{A} = \{x | x \notin A\}$。一般来说，当我们说补集的时候，我们会约定一个全集

(这可能是因为简单地说“所有的其他对象”所指的对象不够明确，也就是不一定构成集合) I。于是，我们通常把补集写作 $\bar{A} = \{x|x \notin A, x \in I\}$。其中 $\{x|p(x)\}$ 的后半部分 $p(x)$ 表示 x 所需要满足的条件。

定义 0.1 (补集). 给定一个集合 A，以及一个当作全集的集合 I，集合 A 的补集定义为 $\bar{A} = \{x|x \notin A, x \in I\}$。

到此，我们就有了集合、补集和全集的概念。下面我们来学习交集和并集——关于两个集合的关系。我们定义两个集合的交集为同时属于两个集合的的元素。

定义 0.2 (交集). 给定两个集合 A, B，其交集定义为 $A \cap B = \{x|x \in A \text{ 并且} x \in B\}$。

其中“并且”可以用记号 $\wedge$ 或者 & 来表示。

定义 0.3 (并集). 给定两个集合 A, B，其并集定义为 $A \cup B = \{x|x \in A \text{ 或者} x \in B\}$。

其中“或者”可以用记号 $\vee$ 或者 | 来表示。

很容易我们就可以把交集和并集通过两两集合的操作来定义多个集合的交集和并集。

如果我们发现，两个集合的交集和其中之一相等，例如 $A \cap B = A$，那就是 A 完全被包含于 B 之中，或者说，所有 A 中的元素肯定都在 B 中。这个可以通过画图来理解。将来我们也可以严格证明。比如说，我们假设有至少一个元素在 A 中但是没有在 B 中，我们就可以证明这个元素不会在 $A \cap B$ 中，于是 $A \cap B \neq A$。这里我们用到了反证法。后面我们会回到用集合的语言来理解反证法。我们把这样的关系称为包含关系。

定义 0.4 (被包含于). 给定两个集合 A, B，我们称 A 被包含于 B，记作 $A \subseteq B$，如果 $x \in B, \forall x \in A$。

其中“$\forall$”的含义是“任意”，也就是随便取一个满足后面的条件。数学家经常把条件放后面，因此，这里的含义是“取任何一个集合 A 中的元素 x，都满足 x 在集合 B 中”。反过来，称作 B 包含 A。

定义 0.5 (包含). 给定两个集合 A, B，我们称 A 包含于 B，记作 $B \supseteq A$，如果 $x \in B, \forall x \in A$。

这里的包含实际上既可以是真包含也可以是相等，也就是 $A \subseteq B$ 可以是 $A \subset B$ 还可以是 $A = B$。其中前者表示 B 中有不属于 A 的元素。

顺便，我们增加一个两个集合相等的定义和一个集合相等和集合包含

关系之间的对应关系的定理。

定义 0.6 (集合相等). 两个集合完全相等的含义是两个集合中的元素完全相同，也就是 $x \in B, \forall x \in A$ 并且 $x \in A, \forall x \in B$。

从这个定义我们就看出来，刚好这意味着 $A \subseteq B$ 和 $B \subseteq A$。

定理 0.1: 集合相等和双向包含

集合相等和双向包含] $A = B \iff (A \subseteq B) \wedge (B \subseteq A)$.

证明. $A = B$ 意味着 $x \in B, \forall x \in A$，也就是 $A \subseteq B$。$A = B$ 还意味着 $x \in A, \forall x \in B$，也就是 $B \subseteq A$。
反过来，$A \subseteq B$ 意味着 $x \in B, \forall x \in A$；$B \subseteq A$ 意味着 $x \in A, \forall x \in B$。和起来，正好就是 $A = B$。 □

注意集合相等的定义是一个额外强加的要求。不一定任何对象都满足这个要求。例如，如果我们用一个人的祖先构成的集合来代表这个人。那么，兄弟姐妹具有完全相同的祖先，也就是集合完全相同，但是兄弟姐妹并不一定只有完全相同的一个人。当然，我们仍然可以说，这些兄弟姐妹有完全相同的祖先的集合（甚至这些祖先的顺序）。因此，集合相同仅仅表示集合相同，不一定能够延伸到这个集合整体所指代的对象之间的关系，如果集合整体所指代的对象有一个定义的话。这个问题牵涉到把集合当作一个更大层次的集合的元素，然后建立映射的关系。未来我们会再来学习这个集合的集合以及映射。很有意思的是，除了未来往往在研究生阶段会学习到非常特殊的情况（叫做范畴），绝大多数数学对象的定义都可以建立在集合的基础上。也就是粗糙地说，数学基本上只需要集合（以及集合的集合，集合的集合的集合，这些仍然都是集合）就够了。集合在数学语言中具有非常强大的威力。

有了上面这一点点集合论的知识，我们来理解一下演绎逻辑。首先，我们来看三段论。三段论的原始形式是这样的：大前提，小前提，结论。其中大前提就是类似于“所有的人要是会死的”这样一个对集合里面的所有元素的断言。小前提是说某个对象属于这个集合，例如“苏格拉底是人”。结论就是把对象和大前提的断言联系起来，例如“于是，苏格拉底是会死的”。用集合论的的语言，现在我们就很好理解这个三段论了：我们从一个集合

中的所有的元素都具有某性质开始，自然也就得到任何一个这个集合中的元素具有这个性质。注意，这看起来就跟废话一样，但是，正是这样的看起来特别简单自然的命题和推理成了构建宏伟的数学知识大厦的基石。

理解了三段论之后,我们来用集合理解逻辑论证中的另外一个主题——命题、逆命题、否命题和逆否命题及其之间的关系。还是用“所有的人要是会死的”、“苏格拉底是会死的”当作原始命题，否命题就是“不是所有人（有的人）都不会死”，“不是苏格拉底就不会死”；逆命题就是“会死的是人”、“会死的是苏格拉底”；逆否命题是“不会死的不是人”、“不会死的不是苏格拉底”。通过对这些话的具体含义作思考，我们自然可以来判断这些命题是否成立。例如，我们发现，逆命题不对，会死的不仅仅是人，以及人之一苏格拉底，还有猫、马、兔子等等。否命题也类似，例如苏格拉底之外的人也会死。比较麻烦的是逆否命题。不过，如果我们可以使用反证法，我们可以这样来论证：假设你否命题不成立，也就是“不会死的还是人”，根据“所有人都会死”，我们得到“不会死的是人，会死”。这里“不会死的会死”直接就矛盾了，因此，这个原始的假设，也就是你否命题不成立和原命题直接矛盾。因此，如果原命题为真，则逆否命题肯定为真。也就是说，原命题和逆否命题完全等价。但是，是不是我们可以脱离开这些命题的实际内容，不需要去找出来满足或者不满足这些命题的事实，直接来作判断，或者至少找到作判断的方法呢？这就是逻辑学所要解决的问题。

我们用集合来当作逻辑学的基础来解决这个原命题、否命题、逆命题、逆否命题之间的关系的问题。我们说一个命题必然有一个前提 A 一个结论 B，记作 $A \Rightarrow B$。我们把所有的满足前提 A 中的要求的对象组成的集合也记作 A[4]，满足结论 B 中的性质的对象组成的集合也记作 B[5]。我们说，原命题 $A \Rightarrow B$ 就意味着集合包含关系 $A \subseteq B$，也就是任何一个 A 中的元素必然属于集合 B。集合的包含关系就是命题中的导出关系。这一点非常重要。通过集合论的语言，我们就可以把命题中的条件到结论的导出关系变成集合的包含关系。

---

[4]或者 $\mathscr{A} = \{x | P_A(x) = 1\}$ 表示满足 $P_A(x)$——也就是 x 满足条件 A——的所有的 x 的集合，但是只要没有歧义我们就用同样的符号指代条件和集合

[5]或者 $\mathscr{B} = \{x | P_B(x) = 1\}$ 表示满足 $P_B(x)$——也就是 x 满足性质 B——的所有的 x 的集合，但是只要没有歧义我们就用同样的符号指代性质和集合

有了这个导出关系和包含关系的对应，我们就可以来看原命题、否命题、逆命题、逆否命题之间的关系了。我们假设原命题 $A \Rightarrow B$ 成立，也就是包含关系 $A \subseteq B$ 成立。插入图，集合包含关系，补集的包含关系 。我们说“否”就对应着补集，否命题就对应着补集之间的包含关系。也就是说，否命题就是 $\bar{A} \Rightarrow \bar{B}$，对应着 $\bar{A} \subseteq \bar{B}$。但是，由于集合 A 被包含在 B 的内部，补集 $\bar{A}$ 实际上会把补集 $\bar{B}$ 包含在内部。因此，我们有 $\bar{A} \supseteq \bar{B}$，或者说 $\bar{B} \subseteq \bar{A}$。对应到命题，我们有 $\bar{B} \Rightarrow \bar{A}$。因此，我们得不到否命题，但是能够得到逆否命题。也就是说，逆否命题和原命题等价，但是，原命题和否命题之间没什么直接的关系，原命题成立的同时逆命题可以成立也可以不成立。

我们再来看原命题和逆命题之间的关系。首先，我们从逆命题和否命题之间的关系是逆否命题关系，我们就可以得到逆命题和否命题完全等价。于是，原命题和逆命题之间的关系，就是原命题和否命题之间的关系，也就是没什么直接关系，原命题成立的同时逆命题可以成立也可以不成立。其次，我们也可以直接从集合来看原命题和逆命题之间的关系。原命题 $A \Rightarrow B$ 成立意味着集合包含关系 $A \subseteq B$ 成立；而逆命题 $B \Rightarrow A$ 成立意味着集合包含关系 $B \subseteq A$ 成立；但是，$A \subseteq B$ 既不能保证也不能否定 $B \subseteq A$ 成立。

当集合 $A = B$ 的时候，我们既有 $A \subseteq B$ 也有 $B \subseteq A$，于是我们有 $A \Rightarrow B$ 和 $B \Rightarrow A$，也就是原命题和逆命题同时成立。反之亦然。这个时候，我们称 A 和 B 等价。

总结一下：通过集合的语言，以及把集合和命题相结合，我们得到——逆否命题和原命题等价，当且仅当对应的集合 $A = B$ 的时候，原命题和逆命题也同时成立，逆命题和否命题由于是逆否命题关系也等价。有了这个知识，我们就可以更加方便地来处理逻辑演绎的问题了。例如，我们可以不加判断就得到“苏格拉底是要死的”和“不会死的不是苏格拉底”完全等价，以及如果要判断逆命题“会死的就是苏格拉底”、否命题“不是苏格拉底就不会死”是否和原命题“苏格拉底是要死的”等价需要去检查集合“苏格拉底”、“会死的”是否完全相同。这个时候，我们自然发现，两个集合不同，于是不等价。

有了这个命题和逆否命题等价，我们就可以进一步来证明反证法是可以用的一种逻辑演绎证明方式。反证法指的就是，为了证明 $A \Rightarrow B$，我们先假设 B 不成立，也就是从 $\bar{B}$ 开始，然后证明一个和 A 直接矛盾的结果，

也就是证明 $\bar{A}$，也就是得到了 $\bar{B} \Rightarrow \bar{A}$。这正好就是 $A \Rightarrow B$ 的逆否命题。因此，反证法相当于通过证明逆否命题成立来证明原命题成立。回到 $\sqrt{2}$ 不是有理数的证明，我们从假设 $\sqrt{2} = \frac{q}{p}$ 是一个有理数并且整数 p, q 之间没有公约数（有就化简掉，直到最简分数）开始，证明了 p, q 之间有公约数 2。这个结果和假设有直接矛盾。因此，假设不能成立，也就是说，$\sqrt{2}$ 不是有理数。

另外，初看起来，这里好像没有了前提。实际上，这里的前提是 $a^2 = 2$，也就是说这里的命题的完整表述是：$a^2 = 2 \Rightarrow a \notin \mathbb{Q}$（$\mathbb{Q}$ 表示有理数），或者说 $a^2 = 2 \Rightarrow a \neq \frac{q}{p}, \forall p, q \in Z$。在我们的证明过程中，我们需要多次使用这个前提条件。

在这一小节，我们学习了一点点集合的语言，然后我们用集合的语言理解了三段论，理解了原命题、否命题、逆命题、逆否命题之间的关系，理解了反证法。三段论、反证法和命题之间的关系是我们后面要用到的演绎证明的基础。将来我们还会把更多的其他的数学概念也建立在集合之上。这体现了集合这个数学基本概念的重要性，也体现了数学知识的系统性。

0.2.4 从小学数学学习和复习到理解型学习

人类知识高速公路上以高层知识生成器为目标的理解型学习

0.3 小学数学知识总结以及本章小结

我们还可以把以上的知识之间的联系通过一张图来表达，如图 1。这张图可以用来帮助你检查一下是否具备了学习本书所要求的小学阶段的知识。如果你差不多能够走通这张图上的各条连边（概念间关系）和各个顶点（概念，数学学科概念和学科思维），则你就基本上具有了学习本书的基础。

在本章中，我们明确了为了学习本书所需要的小学阶段的数学知识，给出了要求列表，做个这些知识的复习。我们再一次强调，本书的这个前置知识的要求不算低，主要问题不在数学计算上，而是在于对数学知识以及这些知识的被创造的过程还有过程背后的思维的理解。如果从列表和复习中你发现好多知识对你是新的，你可以选择把这部分复习多看几遍，也可以

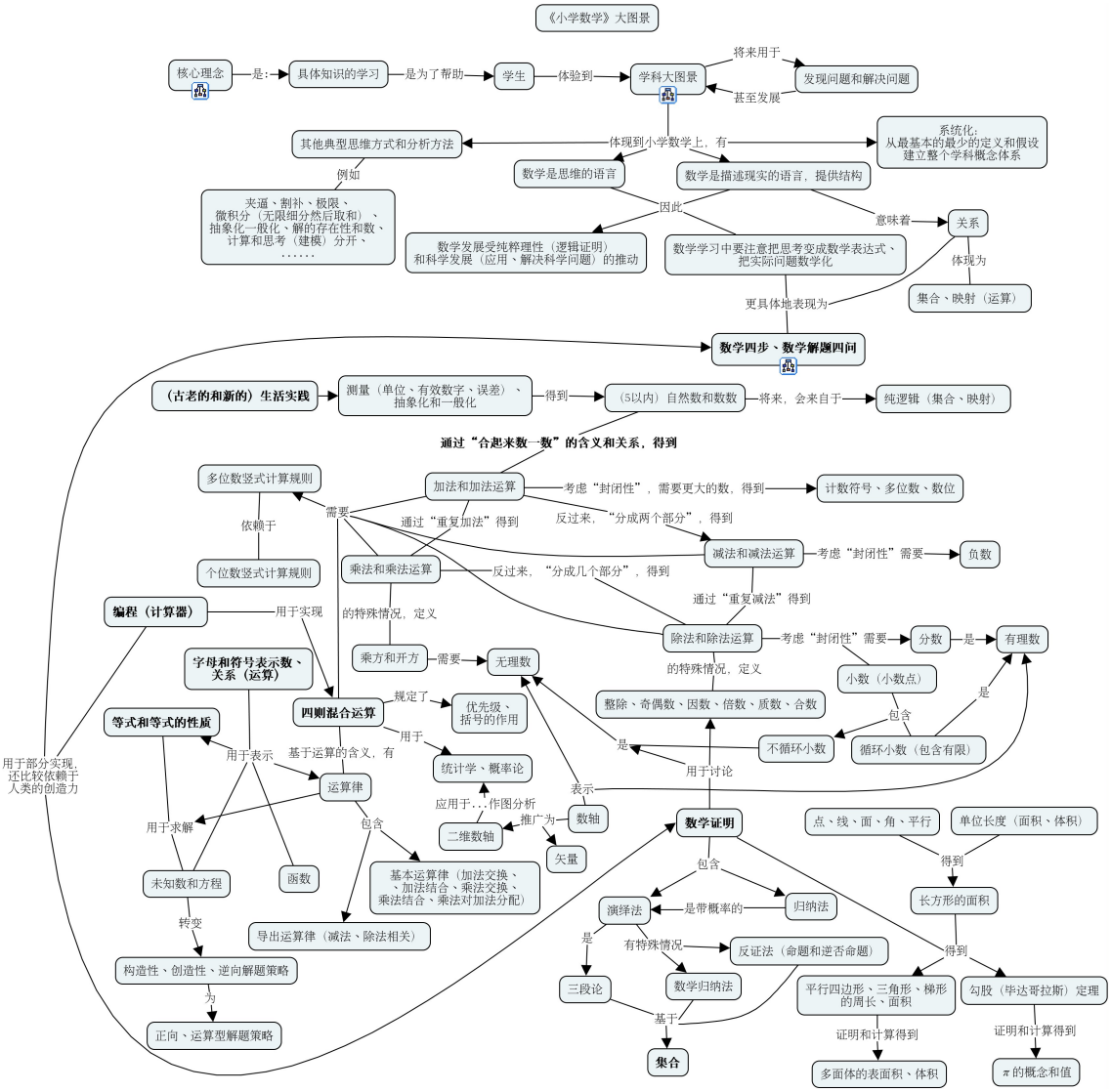

图 1: **小学数学整体概念地图**。可以用来检查一下是否具备了学习本书所要求的小学阶段的知识。考虑是否印单独一页

回去看一下《小学数学这样学》[1]，也可以顺着本书先看下去，然后时不时地回到本章再一次来阅读。

第一部分

初中篇

第一章　数轴上的点和数

1.1　非负整数的数轴表示

1.2　分数的数轴表示

1.3　从 $\sqrt{2}$ 不是分数到 $\sqrt{\frac{q}{p}}$

$\sqrt{2}$ 存在，从几何和代数两个角度，从现实的角度。$\sqrt{2}$ 不是分数的证明。$\sqrt{4}, \sqrt{5}$ 呢?

$\sqrt{\frac{q}{p}}$ 是分数的条件。

运算封闭性的要求。

1.4　无限不循环小数的数轴表示？

数轴上还有其他的数吗？这些数对加减乘除加上乘方开方运算封闭吗?

1.5　有理数、无理数和实数

有理数、无理数、实数的含义和名称。

1.6 数轴上还有其它数吗？

填满了吗？任意两个实数之间必然还有实数。不可能再有其它数了吗？

暂时接受点和“实数”的一一对应关系，未来会有机会进一步来学习、质疑和思考它。

1.7 代数式

用字母表示任何一个给定的数。

用字母和运算符号构成算式，强调关系。任何一个具体数代入字母都可以。

从代表一个给定的数到一个给定但是未知的数，到变量。

1.8 推荐阅读材料

1.9 作业

1.10 本章小结

第二章　幂运算

2.1　数的正整数次幂运算

2.2　数的负整数次幂运算

2.3　数的零次幂

2.4　数的分数次幂

数的无理数次幂呢?

2.5 数的幂运算的性质的字母表示

2.6 完全平方和完全立方公式

2.7 平方差和立方差公式

2.8 推荐阅读材料

2.9 作业

2.10 本章小结

第三章　变量和常数

3.1　完全已知的数

3.2　给定的（当作已知）的数

3.3　未知的或者重点关注的数

3.4　单项式的元数和阶数

3.5　多项式的元数和阶数

3.6　推荐阅读材料

3.7　作业

3.8　本章小结

第四章　因式分解

4.1　乘出来和变成多个算式相乘

4.2　为什么需要因式分解

解方程（代数基本定理，解的数量和方程的阶数相同），多个独立因素联合发挥作用，或者反过来，把某个多个因素联合发挥作用的问题变成每个因素独立发挥作用，然后各个解决。

分而治之。

4.3 简单因式分解

4.3.1 基于乘法对加法分配律的因式分解

4.3.2 基于完全平方公式的因式分解

4.3.3 基于平方差公式的因式分解

4.3.4 基于立方差公式的因式分解

4.3.5 基于完全立方公式的因式分解

4.4 几种类型混合的因式分解 *

4.5 推荐阅读材料

4.6 作业

4.7 本章小结

第五章　一元一次方程和一元一次不等式

5.1　复习：运算律、等式的性质

5.2　一元一次方程的求解

5.3　一元一次方程的应用

5.4　复习：不等号和不等式的性质

5.5　一元一次不等式的求解

5.6　一元一次不等式组的求解

5.7　一元一次不等式的应用

5.8　基于完全平方的恒成立不等式及其应用

凑完全平方来求极值

5.9 推荐阅读材料

5.10 作业

5.11 本章小结

第六章　集合初步

这一章我们来学习集合这个概念。将来继续学习数学之后，我们会发现，集合是绝大多数数学概念的基础，以前学过的数学概念都可以用集合以及下一章要学习的定义在集合之上的映射的语言来重新表述和理解。而提出集合这个概念不过就是为了一个特别简单的目标，为了把意思说清楚，而且仅仅是为了把意思中的那个对象说清楚。这也体现了数学就是思维的语言，就是描述世界的语言。

6.1　集合的朴素概念

在我们用语言来表达我们的意思的过程中，我们往往先要确定一个对象，然后再来说这个对象怎样怎样。例如，“这个苹果很新鲜”这里的对象就是“这个苹果”。这个对象有一个“很新鲜”的属性。在这里，我们其实还藏起来了一个新鲜程度的含义甚至确定这个新鲜程度的操作性测量——例如拿眼睛看如果发现外皮蔫了等特征就说不新鲜，或者用仪器测量这个苹果的水分含量和这个苹果或者这一类苹果刚采摘下来的时候的水分含量做一个比较从而得到一个比例值，然后规定这个比例值小于多少就是不新鲜——的方法。这里的一个测量的方法以及一个判断的标准，就给出了明确的是否新鲜的含义。更进一步，我们还可以说“这一箱子苹果、这一堆苹果很新鲜”，“这种苹果很甜（同样地通过尝一下，或者测量糖分含量，建立标准）”来把我们所指的对象从一个苹果变成一批苹果，一种苹果。在对象上，“一箱子苹果”就从特指的单个的苹果变成了特指的若干个苹果；“一种苹果”就从单个的具体的苹果变成了抽象的一类苹果——这一类可能指的

是和苹果的 DNA（脱氧核糖核酸[1]）以及 DNA 所决定的外形、颜色等属性，以及和这一种苹果的生长环境例如产地等因素有关的品种。也就是说，“一种苹果” 其实是一个抽象概念，它不是由具体的每一个苹果来承担的，而是由这一类苹果的共性特征来承担的。

顺便，沿着这个一个和一类的思路，当我们写作的时候说 “这个人很可爱” 和 “这样的人很可爱” 的时候，其实，我们是做了抽象的，原则上，这个时候我们必须至少在自己心中是明确地把这个人的某个特征提炼为这样的人的特征的。否则，我们最好还是就这个具体对象来讨论，而不是抽象为对其所代表的一类对象的讨论。因此，当你说 “你这人怎么这样” 的时候，你对这个人的批评远远没有 “你这种人，就会这样这样” 来得狠。好了，说了这么一大段语言和表达经常遇到的事情之后，让我们回到数学。

为了更明确地表达出来所指的对象，以及用于指代一群对象或者一类对象，人们提出了 “集合” 的概念。一个集合指的是这样一个群体的概念：我们完全可以判断任何一个对象是否属于这个群体。这个性质被称为集合元素的确定性。群体本身的含义是一个或者多个对象合起来，往往人们还会给这个群体一个合适的名称。例如，$\{1,3,5,7,9\}$ 是一个集合，一个群体，我们可以称之为 “十以内的所有奇数”。例如，所有的自然数也是一个集合[2]。例如，一个硬币（通常抛出去以后落下来的硬币只会显示为两面之一，立着的情况不算）的两面——正面或和反面，也是集合。

那么，是不是人们日常生活中所用的所有的类别名词都满足确定性，都可以算是集合呢？不是的。例如，胖子、秃头、美景、好人都不算集合。我们很难说清楚体重大于多少，或者体重身高比例大于多少算胖子。类似地，秃头、美景、好人也一样。当然，实在有必要的时候，也可以定义一个测量方式和判断的阈值。例如，身体质量指数（Body Mass Index，BMI）就是

[1]一个生物的脱氧核糖核酸决定了这个生物的大部分性状，是这个生物的可以遗传给后代的一定程度上决定了后代的性状的遗传物质。通常同类的生物具有相似的脱氧核糖核酸。这个类的概念在生物学中具有层次性，例如界、门、纲、目、科、属、种。更细的种类上相同的物种一般具有更大的遗传物质上的相似性。

[2]这是一个有趣的故事，有些数学书约定零属于自然数，有些约定 $1,2,3,\cdots$ 才是自然数零不是。因此看起来好像一个数是否属于自然数，当这个数是零的时候，是不明确的。因此，自然数不满足前面提到的确定性要求。但是，实际上，任何一本书，这个约定是唯一的。因此，只要在那本书本身的范围内，自然数仍然是满足确定性要求的。

这样的一个判断肥胖的方式，

$$BMI = \frac{M}{H^2}, \tag{6.1}$$

其中 M 是以公斤为单位的体重，H 是以米为单位的身高。当 $BMI \geqslant 28$ 的时候被称为肥胖。这个时候，忽略测量误差带来的不确定性，我们也就有了满足集合元素确定性的“胖子”集合的定义。

习题 6.1 (BMI 的定义). 请思考为什么 BMI 这样来定义，而不是体重除以身高或者身高的三次方，以及这个定义出来的 BMI 通常被用于哪些地方。

也就是说，由于我们在表达中对明确性有要求，我们尽量从集合开始来描述我们的对象。注意，这不是集合定义，而是给出来集合必须满足的一个要求——元素的确定性。其含义是：给定任意一个对象，我们可以判断这个对象是否属于这个集合。用数学符号这样来表达，就是如果 X 是一个集合，则 $\forall x$（“对于任意一个 x”的意思，读作“任意 x”），$x \in X$（读作“x 属于 X）或者 $x \notin X$（读作“x 不属于 X”）两者有且只有一个成立。也就是，要么 x 属于 X，要么 x 不属于 X，必须有而且只能有其中一个成立，不能两个都不成立，也不能两个都成立。

将来我们会进一步学习到不满足这个确定性要求的数学对象，以及如何来定义集合才能使得它满足这个确定性的要求。

习题 6.2 (理发师悖论). 有兴趣的同学可以查一查“理发师悖论”。思考一下这个理发师悖论和集合的定义以及集合的元素的确定性的关系。

除了确定性，有的时候我们也说集合元素具有互异性和无序性。互异性的意思是，集合的任意两个元素是不同的。无序性的意思是，集合元素的出现顺序无关紧要。于是，$\{1,1,2,3\}$ 和 $\{1,2,3\}$、$\{1,3,2\}$ 是同一个集合。

在这里，我们用到了集合的一种描述方式：列出来这个集合的所有元素，习惯上用一对大括号 $\{\cdot\}$ 来把这些元素包含在里面。对于有限个元素的集合，这可以做到。对于无限个元素的集合，这需要一些定义这个集合的人和阅读这个定义的人之间的一些共同的知识基础。例如，我们说这个符号 $\{0,1,2,3,4,\cdots\}$ 代表了所有的非负整数，这个符号 $\{\cdots,-4,-3,-2,-1,0,1,2,3,4,\cdots\}$ 代表了所有的整数，以及 $\{1,3,5,7,\cdots\}$ 代表了所有的奇数。由于我们经常用到所有整数和所有非负整数这两个集合，我们还给了它们自己的名字和

符号，

$$\mathbb{Z} = \{\cdots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \cdots\}, \tag{6.2}$$

$$N = \mathbb{Z}^* = \{0, 1, 2, 3, 4, \cdots\}. \tag{6.3}$$

集合还可以用另一种方法来描述，给出来所有的元素满足的一个性质，只要这个性质是否满足的条件对于任何对象都是具有明确的结论的，就行。例如，所有的奇数我们可以写成 $\{2n+1, n \in \mathbb{Z}^*\}$。

现在我们有了对集合的几种描述方式：给出来所有的元素，写出来一个所有元素需要满足的性质，约定一个符号或者名称。

有了集合，也就是指代明确的对象（们），仅仅是我们准确表达思想，表达思考过程，表达我们的意思的第一步。未来我们还会学习到如何表达这个对象具有什么性质，这样的判断句——将来会被成为命题，以及对象之间的关系——将来会被称为映射。为了给命题和映射的学习作准备，我们来学习一点点集合本身的运算。

顺便，在这里学习集合这个概念的过程中，你已经看到了我们——其实是人类历史上的数学家们——是如何尽力把意思表达明确的。数学就是一个帮助人们把意思表达明确的语言。

6.2 集合的交和并

其实，我们应该早就发现了，集合 $\mathbb{Z}, \mathbb{Z}^*$ 之间，以及它们和奇数集合之间是有一定的关系的。例如，非负整数集合 $\mathbb{Z}^*$ 其实是整数集合 $\mathbb{Z}$ 的一部分，奇数集合也是整数集合的一部分，并且奇数集合可以通过非负整数的集合来定义。这一节，我们来定义集合之间的关系，以及基于这些关系的运算。

给定任意两个集合 X, Y，可能有元素同时属于这两个集合，也就是 $\exists x \in X, x \in Y$。我们把满足这个条件的元素放到一起，构成一个新的集合，也就是 $\Omega = \{x | x \in X, x \in Y\}$。首先，我们来看看是否这可以看作一个集合，也就是是否满足元素的确定性。对于任意的一个元素 y，其是否属于集合 X, Y 只有四种情况：第一种，不属于两者之中的任何一个；第二种，属于 X，而不属于 Y；第三种，不属于 X，而属于 Y；第四种，属于 X 也属于 Y。我们发现，以上四种情况的任何一种都可以判断这个元素 y 是否属于

Ω。因此，满足元素的确定性。例如，只有第四种情况下，$y \in X$ 并且 $y \in Y$ 则 $y \in \Omega$，其他情况下都有 $y \notin \Omega$。直观上，这个集合 Ω 相当于把两个集合共同的元素都找了出来。

为了把这个直观上的关系清楚地表示出来，我们引入一个集合的图形表示，Venn 图：我们用一个圆形、长方形、椭圆形之类的闭合图形来表示一个集合，把集合内的元素表示成这个闭合图形内的点，集合之外的元素表示为这个闭合图形之外的点。

在这个 Venn 图的表示下，我们发现，前面的集合 Ω 就相当于两个集合 X, Y 的 Venn 图表示的交叉的部分，共同的部分。因此，我们把 Ω 称作集合 X 和 Y 的交集，记作 $X \cap Y$，也就是，

定义 6.1 (交集). $X \cap Y = \{x | x \in X \text{ 并且} x \in Y\}$.

有了交集的概念，我们就很容易定义并集，

定义 6.2 (并集). $X \cup Y = \{x | x \in X \text{ 或者} x \in Y\}$.

可以发现，前面四种情况的后三种都满足并集所要求的性质，只有第一种不满足。因此，其也满足元素的确定性要求。在直观上，并集就相当于把 X, Y 两个集合覆盖的元素合并起来，因此被称作并集。

插入集合的 Venn 图，交集和并集的 Venn 图表示。

6.3 子集、全集和补集

有了交集和并集的概念，我们就可以来判断两个集合的包含关系了。如果两个集合 X, Y 的交集就是其中之一，例如 X，我们说，X 被包含在 Y 之中，X 是 Y 的子集。其中，X 是 Y 的子集的含义是 X 被包含在 Y 之中，也就是

定义 6.3 (子集). $X \subset Y$ 指的是 $\forall x \in X, x \in Y$.

我们还给出如下的一个判断句，或者说命题，如果 X 是 Y 的子集，则 X, Y 的并集是 Y。我们甚至说，反过来也如此，也就是如果 X, Y 的并集是 Y，则 X 是 Y 的子集。

我们来看看这些判断句是否正确：第一，$X \cap Y = X \Rightarrow X \subset Y$；第二，$X \subset Y \Rightarrow X \cap Y = X$；第三，$X \cup Y = Y \Rightarrow X \subset Y$；第四，

$X \subset Y \Rightarrow X \cup Y = Y$。这里符号“⇒”的意思就是左边的条件成立则右边的结论也成立。我们一个个来检查这些判断句是否成立。

通过这个检查，你也会发现集合的语言的另一个作用：集合和逻辑论证是紧密相连的。实际上，把含义表示明确，把对象描述明确，再加上把逻辑论证表述明确，正好就是集合的语言的作用和魅力。

定理 6.1: 集合包含关系和交集一

集合包含关系和交集一]$X \cap Y = X \Rightarrow X \subset Y$。

证明 6.1. 根据交集的定义 $X \cap Y = \{x | x \in X \text{ 并且} x \in Y\}$。现在已知 $X \cap Y = X$，那么就可以推断出来，没有任何一个 Y 之外的属于 X 的元素。因为，如果有 $z \in X, z \notin Y$，那么，这个 $z \notin X \cap Y$，于是 $X \neq X \cap Y$。因此，$\forall z \in X$，必然有 $z \in Y$。根据子集的定义，$X \subset Y$。

定理 6.2: 集合包含关系和交集二

集合包含关系和交集二]$X \subset Y \Rightarrow X \cap Y = X$。

证明 6.2. 一方面，根据交集的定义 $X \cap Y = \{x | x \in X \text{ 并且} x \in Y\}$。另一方面，根据子集的定义，$X \subset Y \Rightarrow \forall x \in X, x \in Y$。也就是说，在 $X \subset Y$ 的条件下，交集的条件中的第二条 $x \in Y$ 已经自然被满足。因此，这个时候交集的条件只剩下 $X \cap Y = \{x | x \in X\} = X$。

第三个和第四个判断句的证明交给读者。有了这四个命题，我们就可以非常好地建立起来 Venn 图的直觉图像和集合的交集、并集、子集之间的对应关系了——你直接上看起来图像上的包含关系、合并图形、交叠图形，正好就是集合之间的相应的子集、并集、交集的关系。

习题 6.3 (集合包含关系和并集一). 证明 $X \cup Y = Y \Rightarrow X \subset Y$。

习题 6.4 (集合包含关系和并集二). 证明 $X \subset Y \Rightarrow X \cup Y = Y$。

对于两个具有包含和被包含关系的集合，$X \subset Y$，或者说 $Y \supset X$（开口方向的含义类似于大于小于号），如果我们把更大的那个集合 Y 称作全集，则那个更小的集合 X 就称作子集。而那部分在 Y 中但是不在 X 中的元素构成的集合则称作子集 X 在全集 Y 中的补集，记作

定义 6.4 (补集). 当 $X \subset Y$ 的时候，子集 X 在全集 Y 中的补集 $\overline{X} \triangleq \{z | z \notin X \text{ 并且} z \in Y\}$。

有了子集的概念，我们可以来讨论两个集合的相等和子集的关系，也就是

定理 6.3: 子集和集合相等

子集和集合相等]$X = Y \iff X \subset Y$并且$Y \subset X$。

证明 6.3. 按照集合的定义，两个集合相等 $X = Y$ 的含义是两个集合中包含的所有元素相等，也就是 $\forall x \in X, x \in Y$，$\forall y \in Y, y \in X$。其中，前半句正好就是 $X \subset Y$ 的定义，后半句正好就是 $Y \subset X$ 的定义。因此，$X = Y \Longrightarrow X \subset Y$并且$Y \subset X$。

反过来，按照包含关系的定义，$X \subset Y$ 就是 $\forall x \in X, x \in Y$，$Y \subset X$ 就是 $\forall y \in Y, y \in X$，两者和起来正好就是 $X = Y$ 的含义。因此，$X \subset Y$并且$Y \subset X \Longrightarrow X = Y$。

其中双向箭头符号 $\iff$ 的含义是，既有 $\Longrightarrow$ 还有 $\Longleftarrow$。也就是，命题的两侧互为结论和条件，也就是两者相互等价。因此，$\iff$ 也被读作“等价于”。在证明等价关系的过程中，我们要作双向的命题的证明。注意，命题并不都是双向的。

有了这个集合相等和子集之间的关系，我们在以后证明两个集合相等的过程中，经常会通过证明这两个存在相互包含关系来完成。

更进一步，为了区别两个集合之间的包含关系和相等关系，我们把前面所说的“包含”关系区分为“包含”（$\subseteq$）和“真包含”（$\subset$）关系。两者之间的区别是“包含”关系允许两个集合之间是相等的关系或者真包含关系，而“真包含”关系不允许两个集合之间是相等的关系。也就是说，以后 $X \subset Y$，意味着 $\forall x \in X, x \in Y$，而且，$\exists y \in Y, y \notin X$。$\subseteq$ 和 $\subset$ 就类似于小于等于号（$\leqslant$）和小于号（$<$）的关系。

插入子集、全集、补给、包含关系的 Venn 图。

到此为止，我们就学完了基本的集合知识了，包含集合要满足的元素的明确性（以及不太重要的互异性、无序性），集合的交和并运算，子集（包含和被包含关系）、全集和补集，以及集合相等和相互包含的关系。为了帮助大

家熟悉这些概念和符号，我们编制了需要大家完成的一些练习题，还把在一个全集中的两个集合的一般关系的每一个区域画成了一张 Venn 图（插入一张给各个区域用集合运算符号表示的 Venn 图，例如 $X \cap Y, X \cup Y, X \sup \overline{Y}$ 等等 ）。

再次回到为什么我们要有集合的语言：为了我们在思考和表达中所指代的对象更加明确，以及形成可以直接指代一堆或者一类对象的语言和符号，也就是用来描述对象之间的层次关系、交叉关系的合适的语言和符号。我们也初步看到了，集合这个语言在表达和证明命题中的作用。

编制一些集合交和并运算，补集运算的习题。

6.4 不等式的解当作集合

6.5 推荐阅读材料

6.6 作业

6.7 本章小结

第七章　映射初步

这一章我们来学习数学的另一个抽象概念——映射。我们将会发现，映射是描述对对象的操作——例如给对象命名、分类，以及把一个对象和另一个对象联系起来——的最基本的方式。而对象和操作是我们人类思考和面对世界的典型方式。因此，作为思维的语言和描述世界的语言的数学就有这个责任来给对象和操作准备好语言。上一章，我们学会了用集合来明确指代对象，这一章，我们来学习用映射指代操作。

7.1　从命名、分类、操作到映射

有三个世界时时刻刻伴随着我们：现实世界、概念世界（即存在于人类的认知结构中，也存在于各个学科的知识构成的空间中）、语言世界。我们表面上用语言世界（可以包含人类的自然语言以及数学语言、编程语言）去描述现实世界，实际上，我们是用概念世界去描述现实世界，但是需要通过语言世界这个媒介。因此，对于现实世界的对象，我们往往就得给这个对象一个语言世界的名称以及一个概念世界的含义。这个把现实世界的对象和其在概念世界和语言世界的名称和含义联系起来的过程，就是映射。

例如，如果我们有两支铅笔，而且我们经常需要按照某种方式——例如使用目的或者个人偏好——来选择用其中的哪一支。现在假设每次我们有另一个人替你去拿笔，你必须想办法明确告诉那个人每次你需要拿哪一支。实际上，你也可以把“你”看作你的大脑，把“另一个人”看作你的手。我们来设想一个如何完成这个明确告诉另一个人的过程。

首先，你可能由于某种原因选择了其中那支黑色外壳蓝色笔芯的铅笔，

而不选用另一支是白色外壳黑色笔芯的圆珠笔。接着，你怎么告诉另一个人我们的这个选择呢？按照所选择的这支笔的某种特征，而且是可以和另一支相区别的特征。例如，请你帮我取一下那支黑色外壳的笔，那支铅笔，那支蓝色笔芯的笔，以及它们这些特征的组合。不过在这里，我们不能说“黑色的笔”，这可能意味着黑色外壳，也可能意味着黑色笔芯。我们还可以先跟另一个人做一个约定，把“黑色外壳蓝色笔芯的铅笔”称作“一号笔”记作①，把“白色外壳黑色笔芯的圆珠笔”称作“二号笔”记作②。只要在这个约定下，我们就可以很方便地达到告知另一个人我们的选择的目的。

我们说，在这个过程中，你建立了映射，也就是从这两支笔到这两支笔的被指代方式之间的一个映射。例如，从这两支笔到这两支笔的两种外壳颜色的映射，从这两支笔到这两支笔的笔芯颜色的映射，从这两支笔到这两支笔的笔头材料（铅笔还是圆珠笔）的映射，从这两支笔到这两支笔的记号之间的映射。我们在下一节再来明确地学习什么是映射。但是，通过这个例子，我们大概知道了映射就是给一群东西里面的每一个东西一个名字，或者说一个记号。在下一节中，我们会发现，这个指代明确的记号名字的映射通常要求是一对一的。

其实在上面的为了指代对象明确而作的命名中，我们实际上用到了分类。例如，所有的黑色外壳的笔，所有的铅笔，所有的蓝色笔芯的笔等等。因此，当我们给一个东西一个名字，其实往往这个名字和这个对象的某种属性有关，并且满足这个属性特征的对象可能还不止这一个。这时候，我们说，其实我们对对象做了分类。从映射的角度，分类就是建立了一群对象中的每一个对象和类别名称以及这个名称代表的类别属性之间的映射。具体来说，在这里，就是建立了第一支笔和铅笔之间的映射，第二支笔到圆珠笔之间的映射。就像我们已经知道的，有很多笔都是铅笔，有很多笔都是圆珠笔，因此，这个映射实际上是给了很多东西一个共同的名字，然后把这个共同的名字叫做这些很多个东西的类别。在下一节中，我们会学习到，这就是多对一的映射。

下面我们来假设一个更加复杂一点场景：我们有一把削铅笔的刀，它可以用于削所有的铅笔但是不能用于任何一支削圆珠笔。那么，这个削笔刀就是一个映射，把一支通常笔头磨损了不方便书写的铅笔变成一支具有再次方便书写的笔头的铅笔。任何一支钝笔头铅笔和这个削笔刀结合（其

实还要加上削的过程，但是我们先忽略这个过程），就可以得到一支好笔头铅笔（假设这支铅笔的长度使得它还能和值得被削）。为了说起来简单，我们允许削笔刀去削本来就不钝的铅笔，也就是只要来一支铅笔，就削一下得到一支好笔头铅笔。我们还可以让这个削笔刀更加智能一点，给它增加一个判断器——如果来的是铅笔，则削，如果是圆珠笔，则不削。实际上这个判断器也是一个映射，遇到“铅笔”对应着“使用削笔刀”，遇到“圆珠笔”对应着“不使用削笔刀”。原来的削笔刀和这个判断器的组合成了一个有点智能的削笔刀。实际上，你还可以增加一个映射——一个钝笔头判断器。但是，为了说起来简单，前面我们已经允许削笔刀削本来就笔头不钝的笔。然后，我们来把这个智能削笔刀映射和前面的取用笔的映射结合起来：先取出来笔，然后，把这个取出来的笔投入到智能削笔刀。这时候，我们就得到了一个更加智能的取用削笔刀，我们只要告诉它“帮我取黑色外壳的笔”，它就会取出来，判断一下是否是铅笔，如果是，削好比头，递过来给我们用。稍后我们会发现，用映射的语言，这叫做映射的复合。

通过这一节，我希望你已经领会到，明确对象以后，对对象的命名、分类和操作这些典型的认识和改变世界的行为，实际上都是映射。现在，我们来明确定义映射。数学就是给你认识和改变世界准备的语言。

7.2 映射

定义 7.1 (映射). 两个集合 X, Y 之间的映射 f，记作 $f : A \mapsto B$，满足如下条件：第一，对于任意一个 X 中的元素 x $(\forall x \in X)$，存在着唯一的一个 Y 中的 y $(\exists x \in X)$，使得 $f(x) = y$。

这个定义就是把我们前面已经形成的关于映射的直觉，从一个集合 X 里面取出来一个元素 x，把这个元素联系到另一个集合 Y 的一个元素 $y = f(x)$。我们把取出来元素 x——称作原像——的集合称作定义域（domain），联系到的元素 y——称作像——的集合称作陪域（codomain）。这个联系要满足一个条件：任何一个定义域中的元素都可以在陪域中找到一个而且只有一个像。

映射的定义对于是否会有两个不同的 $x_1, x_2 \in X$ 都被映射为同一个 $y \in Y$ 没有约束。如果这个条件满足，则映射 f 称作单射，

定义 7.2 (单射). 单射是满足如下条件的映射：$f : A \mapsto B$，对于 X 中的任何两个元素 x_1, x_2 $(\forall x_1 \neq x_2 \in X)$，$f(x_1) \neq f(x_2)$。

映射的定义对于是否每一个 $y \in Y$ 都被映射 f 用到没有约束。如果这个条件满足，则映射 f 称作满射，

定义 7.3 (满射). 满射是满足如下条件的映射：$f : A \mapsto B$，对于 Y 中的任何一个元素 y $(\forall y \in Y)$，存在着一个 X 中的 x $(\exists x \in X)$，使得 $f(x) = y$。

如果一个映射既是单射又是满射，则被称为可逆映射。可逆映射的含义是，可以构造出来一个从 Y 到 X 的映射 $g : Y \mapsto X$，使得任何一个 $f(x) = y$ 都有 $g(y) = x$。以后我们会看到为什么我们特别关注映射的可逆性。现在，我们来看一下，为什么既是单射又是满射的映射是可逆的。映射的核心就是原像必然找得到唯一的像。要是把 $f : X \mapsto Y$ 反过来成为 $g : Y \mapsto X$，那就是要求，任何一个 y 也有且只有一个 x 满足 $g(y) = x$，而且这里的 $g(y) = x$ 正好就是 $f(x) = y$。因为映射是满射，所以，任何一个像 y 都找得到至少一个原像 x 使得 $f(x) = y$。我们就差一个唯一性了，也就是要求不存在这样的 $z \neq x \in X$ 使得 $f(z) = y$。这正好就是单射的要求：两个不同的原像必然导致不同的像。这个时候，我们把逆映射 g 也记作 f^{-1}，也就是 $g = f^{-1} : Y \mapsto X$，而 $f : X \mapsto Y$。

一旦两个集合之间可以建立起来可逆映射，则这两个集合的元素之间就有了一一对应关系。这就好像是我们之前讲到的取名字的场景，任何一个被取名字的对象都有了唯一的名字，任何一个名字都有了唯一的对象。

那既然可逆映射这么好，就好像完全就是给对象取个名字一样，为什么我们不直接要求所有的映射都是可逆映射呢？可逆映射没法满足分类的要求。比如说，我们通过身份证给每一个人一个的号码，通过这个号码也找得到对应的人，但是，我们就不能通过身份证号来把人分成男人、女人了[1]。那实际上，身份证号确实就区分标记了男人和女人。怎么做的呢？第一，我们建立一个集合 S 其中有两个元素 F, M 分别代表女性和男性。第二，我们把所有的人的集合 P（或者你们班里的所有同学们）通过映射 $\mathscr{S}$ 映射到这个性别集合 S，使得对于任何一个 $p \in P$，$\mathscr{S}(p) = F$ 或者 $\mathscr{S}(p) = M$（映射的定义已经要求两者不能同时成立）。当然，我们知道这个映射本身是通

[1]假设我们在这里暂时仅仅考虑男人和女人两种性别。

过每一个人的 DNA 来判断的，或者大多数时候可以通过性别特征来作粗糙的判断。因此，这个分类的背后其实有更加深刻的概念。这个分类也会带来一些额外的信息，例如可能男性普遍来说身高更高一点，力气更大一点，因此有些任务可能更加适合男性。也就是说，把人分成男人和女人，是有意义的。好了，我们已经看到了不可逆的映射，那些把一类的 x 映射为某个 y，把另一类的 x' 映射为另一个 y' 是有意义的。实际上，将来我们会发现，可逆映射反而太平庸，这样的相当于抽象和分类的映射在更深刻的问题上才更有用。

因此，在这里，我们先定义了一般的映射，然后在定义单射和满射，再合起来定义了可逆映射。实际上，你可以追问，那是不是可以允许有些 $x \in X$ 找不到像，以及有些 $x \in X$ 找得到多个像。对于前者，我们重新定义一下集合 X 就可以。对于后者，可能我们就需要想办法把这样的多个像用某种标准分开，然后分开以后各自成为一个映射。例如，4 的平方根有两个 ± 2，但是，我们可以区分出来其中哪些大于零，哪些小于零，然后，分别构建映射。这就是正的平方根和负的平方根的映射，也就是 $\sqrt{x}$ 和 $-\sqrt{x}$，对于任何 $x > 0$。也就是说，很有可能，基于这里定义的映射，我们可以很好地处理各种类型的关系。这也有待于我们的后续学习和运用数学中来进一步检验这一条。

7.3 映射的复合，最简单的群和表示理论

有了集合和映射，我们可以重新定义大部分的数学概念。例如，在小学阶段，我们学习数学是从自然数开始的。我们把自然数和用自然数来数数当作“自然的天生就有的”东西，然后，从数数开始我们可以建立加法的概念，以及加法之上的减法、乘法，还有乘法和减法之上的除法的概念。接着，我们从具体运算中总结出来了运算律。最后，到了用字母表示数。这基本上就是我们目前阶段已经学习到的算术和代数。今天，我们非常粗糙和初步地来体会一下，数其实不是数学里面最重要的最基础的概念，也不是天生的自然就有的，反而是运算和关系决定了数。在高等数学里面，这叫做表示理论。我们不会去学习这个名词，但是，我们会通过例子来帮你体会一下什么是表示理论，为什么世界往往可以由数（或者说代表数的字母，

字母参与的算式）来描述。

为了这个目的，我们先要来学习一下映射的复合。给定两个映射 $f: X \mapsto Y, g: Y \mapsto Z$，我们可以定义一个新的映射 $g \circ f: X \mapsto Z$：对于任意 $x \in X$，$g \cdot f(x) = g(f(x))$。其中我们用到了 $f(x) \in Y$，因此，$g(f(x))$ 存在且唯一。这满足了 $g \circ f: X \mapsto Z$ 是一个映射的要求。复合函数中先作用的函数一般写在右侧。这样它会先作用于写在右侧括号中的原像上。

前面我们提到了给笔编号的映射（记为 $L: P \mapsto N$，从笔的集合到编号的映射），判断是否铅笔并且如果是则削尖铅笔的映射（记为 $S: P \mapsto P$，从笔的集合到笔的集合的映射），我们还提到了我们可以组合成一个决定使用哪一支笔，告诉另一个人需要的笔的编号，然后判断一下是否铅笔，如果是就削尖的智能映射（记为 $IS: N \mapsto P$，从编号的集合到笔的集合的映射）。这其实就是映射的复合。

具体这是如何复合起来的呢？我们看到，复合映射 $IS: N \mapsto P$，把一个数字编号变成一支笔而且是削尖了的笔。因此，它肯定先要根据数字编号来找到相应的笔，也就是利用 $L^{-1}: N \mapsto P$，它是映射 $L: P \mapsto N$ 的逆映射。这也是为什么在之前的场景中，我们要求编号和笔之间具有一一对应关系的原因。一一对应关系是可逆的。找到笔之后，就可以接着用上削笔映射 $S: P \mapsto P$ 了。因此，$IS = S \circ L^{-1}$。

插入符合函数的图示

现在，我们可以来描述一个原始场景下完全和数没有关系的问题了——我们来考虑对一个硬币的状态的可能的操作。忽略抛出以后立起来的硬币，硬币的状态只有两个，我们记作 $\mathbb{S} = \{H, T\}$。注意，这里的 H, T 不是数，就是代表硬币两个状态的一套符号。你可以理解为“头、尾”状态或者“正、反”状态。给定一个抛出以后静止在桌子上的硬币，其状态肯定是 H, T 两者之一，也就是 $S = H$ 或者 $S = T$。

从日常生活中，我们知道，对于一个硬币，我们只有两种操作，不翻转（保持硬币当前状态不动）和翻转。实际上，这两个操作是映射。我们分别记作 $\hat{I}$ 和 $\hat{X}$。上面加一个箭头符号表示这是映射不是集合。把这两个映射明确写出来，我们有，$\hat{I}(H) = H, \hat{I}(T) = T$，$\hat{I}(H) = T, \hat{I}(T) = H$。

我们来考虑一下所有的可能的从 $\mathbb{S}$ 到 $\mathbb{S}$ 的映射。由于 $\mathbb{S}$ 只有两个元素，因此，从 $\mathbb{S}$ 到 $\mathbb{S}$ 的映射只能有四个：除了上面的 $\hat{I}$ 和 $\hat{X}$，还有 $\hat{H}$ 和 $\hat{T}$，前

者把所有的状态都映射为 H，后者把所有的状态都映射为 T。后两者不是满射，我们暂时仅仅考虑前面两个满射，$\hat{I}$ 和 $\hat{X}$。

我们来作一下 $\hat{I}$ 和 $\hat{X}$ 的复合映射。通过画下面的图，我们很容易就能发现，$\hat{I}\circ\hat{I}=I, \hat{I}\circ\hat{X}=X, \hat{X}\circ\hat{I}=X, \hat{X}\circ\hat{X}=I$。从直觉上这也很好理解，两次不翻转合起来是不翻转，先翻转后不翻转合起来是翻转，先不翻转后翻转合起来是翻转，两次翻转合起来是不翻转。到现在为止，我们的硬币问题仍然没有出现任何数。但是，你有没有发现，这里的 $\hat{I}$ 和 $\hat{X}$ 的复合映射的关系，使得 $\hat{I}$ 和 $\hat{X}$ 看起来很象很象 1 和 -1：$1\times 1=1, 1\times(-1)=-1, (-1)\times 1=-1, (-1)\times(-1)=1$。在这个意义上，我们就用 1 和 -1 分别来表示不翻转和翻转操作 $\hat{I}$ 和 $\hat{X}$。到这里，忽然我们的硬币翻转问题出现了数和乘法，而且这个数和乘法正好描述了对硬币的操作以及这些操作之间的复合映射关系。我们已经提过，映射代表了对对象的操作，远远要比数的运算适用范围大得多。但是，你看，从映射出发，考虑复合映射，自然地我们回到了数以及数的运算。这是为什么世界往往由数和运算来描述：对这个世界中的对象的操作，以及这些操作之间的复合关系，导致了数和运算的出现，或者说往往被数和运算来表示。

为什么我们优先讨论这两个可逆映射，而不是把这四个都放在一起来讨论呢？将来我们会学习到一个叫做群的东西，就是对一个对象的所有的可逆操作的集合，操作之间具有复合关系。在数学和物理学的研究中，我们往往需要用到合适的数学结构来表示群（以及建立在群的基础之上的更加复杂的数学结构）。

当然，实际上，我们可以尝试给更加一般的映射找到合适的数学结构。例如，

习题 7.1 (全部四个映射之间的复合关系). 请你来写出来全部四个映射之间的复合关系，然后，试一试是不是能够找到合适的数学结构来表示这四个映射，从而这些数学结构之间的某种乘法运算正好就是这四个映射之间的复合关系。这个问题有点难，你可能得学一点点矩阵运算之后再来找这样的数学表示。

小知识 7.1 (Dirac 符号). 在学习矩阵之前，我们还可以用基于前面提到的 H,T 的符号来给这四个映射找到一个数学表示。将来，这一套表示会在认识世界中，尤其是量子的世界中，发挥巨大的作用。我们考虑一个从

$\mathbb{S}=\{H,T\}$ 到 $\{0,1\}$ 的映射。稍后我们会发现为什么需要这个映射。同样，我们只考虑到满射。定义域和陪域都只有两个元素，因此满射只有两种：$H\to 1,T\to 0$，以及 $H\to 0,T\to 1$。

我们把前者记作 $\langle H|$，后者记作 $\langle T|$。稍后我们也会看到为什么我们用这个记号。然后，我们引入两个新的记号，$|H\rangle$ 和 $|T\rangle$。前者代表硬币状态 H，也就是从定义域中选取了 H；后者代表硬币状态 T，也就是从定义域中选取了 T。然后，我们约定下面的计算规则，

$$\langle H|\ H\rangle=1,\langle H|\ T\rangle=0, \tag{7.1}$$

$$\langle T|\ H\rangle=0,\langle T|\ T\rangle=1. \tag{7.2}$$

在这个运算规则下，我们发现，$\langle H|$ 就起到了第一个映射的作用，$\langle T|$ 就起到了第二个映射的作用。也就是说，运算规则算式 (7.1) 和算式 (7.2) 下，$\langle H|$ 和 $\langle T|$ 就是前面提到的映射的一种表示。

注意，我们最终的目标，不翻转和翻转操作，是从 $\mathbb{S}=\{H,T\}$ 到 $\mathbb{S}=\{H,T\}$ 的映射。现在我们有了从 $\mathbb{S}=\{H,T\}$ 到 $\{0,1\}$ 的映射的表示 $\langle H|$ 和 $\langle T|$，我们能不能在此基础上，再把上面的不翻转和翻转映射也构建起来呢？

我们这样来想，$\langle H|$ 把 H 映射为 1，因此，如果再配上一个 $|H\rangle$ 在后面，那就得到了把 H 映射为 H，$|H\rangle\langle H|$。

类似地，$|H\rangle\langle T|$ 把 T 映射为 H；$|T\rangle\langle H|$ 把 H 映射为 T；$|T\rangle\langle T|$ 把 T 映射为 T。

有了这四个操作，我们终于有了翻转和不翻转的表示，

$$\hat{I}=|H\rangle\langle H|+|T\rangle\langle T|, \tag{7.3}$$

$$\hat{X}=|H\rangle\langle T|+|H\rangle\langle T|. \tag{7.4}$$

习题 7.2 (Dirac 符号下算赋的验证 1)**.** 请验证算式 (7.3) 和算式 (7.4) 的算赋，结合运算规则算式 (7.1) 和算式 (7.2)，正好就是映射不翻转 $\hat{I}$ 和翻转 $\hat{X}$ 作用到硬币状态 H,S 上的效果。

习题 7.3 (Dirac 符号下算赋的验证 2)**.** 请验证算式 (7.3) 和算式 (7.4) 的算赋，结合运算规则算式 (7.1) 和算式 (7.2)，正好就是映射不翻转 $\hat{I}$ 和翻转

$\hat{X}$。这次你来通过计算这些映射之间的复合关系完全相同的方式来验证。也就是，验证 $\hat{I}\hat{I}=\hat{I},\hat{I}\hat{X}=\hat{X},\hat{X}\hat{I}=\hat{X},\hat{X}\hat{X}=\hat{I}$。

一旦我们作了这个验证，那我们就可以再一次得到 $\hat{I},\hat{X}$ 可以分别用数 $1,-1$ 以及它们之间的乘法来表示。你看，我们再一次从完全没有数的一堆 H,T 之类的符号，得到了数和运算，并且用这些数和运算来表示对硬币的翻转和不翻转操作。

这就是数学，给现实世界准备好描述它们的数学结构。

7.4　三个元素的集合到自身的映射 *

上一节我们已经学习了两个元素的集合上的映射。这一节，我们来试试三个元素的集合上的映射。记这个三元素集合为 $\mathbb{T}=\{1,2,3\}$。其中 $1,2,3$ 仅仅是代表这个集合中的三个不同元素的记号。它们之间没有任何关系。

跟之前一样，我们主要考虑满射。由于需要是满射，从三个元素映射到三个元素，必然也是单射。否则，当有两个或者以上的元素映射为同一个元素的时候，则肯定有的元素没有原像，也就不可能是满射。于是，这样的映射既是满射又是单射，合起来必须是可逆映射。那么，$\mathbb{T}$ 到其自身的映射可能有哪些呢？我们一个个来列举一下。

例如我们从元素 1 开始，其像可能为 $1,2,3$ 三者之一。确定了 1 的像之后，我们来确定 2 的像。这时候有两个不同选择——1 选中的像就不能再选为 2 的像了。最后，3 的像只能是除了前两个像之外的第三个元素。也就是只有一种选择。合起来，我们有 $3\times2\times1=6$ 种映射。我们把它们都写出来，并且给一个记号。

映射 $e:\mathbb{T}\mapsto\mathbb{T}$ 的原像和像分别为：$e(1)=1,e(2)=2,e(3)=3$。映射 $g_1:\mathbb{T}\mapsto\mathbb{T}$ 的原像和像分别为：$g_1(1)=2,g_1(2)=3,g_1(3)=1$。映射 $g_2:\mathbb{T}\mapsto\mathbb{T}$ 的原像和像分别为：$g_2(1)=3,g_1(2)=1,g_1(3)=2$。映射 $f_1:\mathbb{T}\mapsto\mathbb{T}$ 的原像和像分别为：$f_1(1)=1,f_1(2)=3,f_1(3)=2$。映射 $f_1:\mathbb{T}\mapsto\mathbb{T}$ 的原像和像分别为：$f_2(1)=3,f_2(2)=2,f_2(3)=1$。映射 $f_3:\mathbb{T}\mapsto\mathbb{T}$ 的原像和像分别为：$f_3(1)=2,f_3(2)=1,f_3(3)=3$。

我们直觉上可以看到，e 相当于不变映射；g_1 大概相当于把这三个数做了一个往前推一步，当得到的结果大于 3 的时候，把得到的结果减去 3，

重新找到一个原始集合 $\mathbb{T}$ 的元素；g_2 大概相当于把这三个数做了一个往前推两步，当得到的结果大于 3 的时候，把得到的结果减去 3，重新找到一个原始集合 $\mathbb{T}$ 的元素；f_1 相当于把第一号元素保持不变，其他两个元素互换；f_2 相当于把第二号元素保持不变，其他两个元素互换；f_3 相当于把第三号元素保持不变，其他两个元素互换。

现在我们来考虑这些映射之间的复合。首先，由于 e 是不变映射，因此任何其他映射 g_l 或者 f_l 和 e 的复合，都得到那个映射它自己，也就是相应地 g_l 或者 f_l。也就是 $e\circ g_l = g_l\circ e = g_l$，$e\circ f_l = f_l\circ e = f_l$。

其次，$g_1\circ g_1 = g_2$，连着两次往前推一步，合起来是推两步。类似地，$g_1\circ g_2 = e = g_2\circ g_1$，推三步相当于不变；$g_2\circ g_2 = g_1$，推四步相当于推一步。于是，我们发现，$e, g_1, g_2$ 之间的关系很象 $1, q, q^2$，只要满足 $q^3 = 1$ 就行。当然，这里我们还不知道什么样的 q，除了 $q = 1$[2]之外还满足 $q^3 = 1$。未来等我们学习到复数，就可以发现这样的 q 确实存在。假设我们找到了这样的 q，我们就再一次从映射、映射的复合走到了数和数之间的乘法运算。

其实，直觉上，我们还可以构建另一个表示：在圆环上，选择三个距离相等的点，顺时针地记作点 $1, 2, 3$。然后，e 就是不转动圆环，g_1 就是逆时针转动一步，g_2 就是逆时针转动两步。转动三步的时候，自然就回到原始的样子，也就是相当于没有转动。

插入圆环上顺时针编号，逆时针旋转的示意图。

加入 f_l 之后，复合映射的关系会复杂一些，我们把它编制成下面这张表。其中行对应的映射放在左边（后作用），列对应的映射放在右侧（先作用）。我们发现，这个表的每一行和每一列，每一个映射必须而且只能出现一次。我们来尝试理解一下这是为什么。

我们给这个 6 个元素的集合一个记号，$G_6 = \{e, g_1, g_2, f_1, f_2, f_3\}$。首先，我们发现，在复合映射下，每一个映射都可以看作是 G_6 到 G_6 的映射。例如，$e : G_6 \mapsto G_6$ 的规则是 $e(g) = e\circ g = g, \forall g \in G_6$。换一个试试，$g_1 : G_6 \mapsto G_6$ 的规则是 $g_1(g) = g_1\circ g, \forall g \in G_6$。我们说，$g_1(g)$ 不会重复出现，也就是不存在 $g \neq g' \in G_6$ 满足 $g_1(g) = g_1(g')$。

我们来用反证法，假设存在这样的 $g \neq g'$，使得 $g_1\circ g = g_1\circ g'$。由于 g_1 是

[2]当 $q = 1$ 的时候，e, g_1, g_2 都被表示为 1。这个显然不是理想的情况，因为我们知道 e, g_1, g_2 各不相同。

可逆映射，也就是在像和原像之间存在着一对一关系，我们发现，g 必须等于 g'，否则，只要某一个 $j \in \mathbb{T}$ 有 $g(j) \neq g'(j)$，则必然有 $g_1(g(j)) \neq g_1(g'(j))$。

不会重复出现，又要和每一个映射去复合，因此，只能每一个都出现一次。也就是说，在复合映射下，每一个映射也都 G_6 到 G_6 的可逆映射。

列的情况也一样。因此，我们得到了每一行每一列每一个 G_6 的元素必须而且仅仅出现一次的结果。你有没有发现，这个和“数独”（Sudoku，每行每列每个宫格每个数只出现一次的含义）游戏很像很像。你可能觉得数独就是一个编出来的游戏而已。现在，我们似乎看到了数独后面的数学结构。

习题 7.4 (数独游戏). 请你查一查，数独游戏是什么，其背后的数学都有什么。提示：例如，你可以查“数独”（或者“*Sudoku*”）、“数独和群”（或者“*Sudoku and group*”）。

注意，这里的每一个映射有了两种含义：第一，在原始的语言下，它是 $\mathbb{T}$ 到 $\mathbb{T}$ 的映射；第二，在复合映射的语言下，它是 G_6 到 G_6 的映射。并且，看起来，我们完全可以脱离开 $\mathbb{T}$ 而直接在 G_6 讨论这些映射以及这些映射的表示。

插入 g 的两种映射的示意图。

将来我们会学习到，G_6 这样的集合——这个集合上有复合映射关系，并且在每行每列中每个元素在这个复合映射关系下出现而且仅仅出现一次——叫做群。群通常建立在对一个更加基本的对象——例如这里的集合 $\mathbb{T}$——之上，但是，也可以忘掉这个更加基本的对象，仅仅从复合映射关系出发来研究群。

仔细观察 G_6 的复合映射结果列表，我们发现，一定程度上，这个列表是分块的：e, g_1, g_2 基本上构成了一个内部封闭的区域，在这个小区域内不牵涉到 f_1, f_2, f_3；另一方面，一旦加入了 f_1, f_2, f_3，则这分开的两片又混合起来了，联系起来了。

在本小节中，我们用集合 $\mathbb{T}$ 来定义其上的映射，然后把这些映射再看做一个集合 G_6，接着再用复合映射来定义 G_6 上的映射，而且这些映射正好是 G_6 的元素自己。最后，通过这个复合映射我们初步体验了群这个数学结构，尤其是其结构上的两面性：第一面，G_6 建立在对更加基本的对象 $\mathbb{T}$ 的操作之上，也就是 G_6 中的任何一个映射把一个 $\mathbb{T}$ 元素变成 $\mathbb{T}$ 的元素，

维持 $\mathbb{T}$ 整体不变；第二面，G_6 可以完全脱离其基本作用对象 $\mathbb{T}$，单纯从复合映射的角度来研究。

我们已经看到群可以从依赖对象的操作和操作之间的复合关系来看，还可以直接从操作之间的复合关系来看，仅仅名义上需要一个对象实际上不再需要对象（也就是不再需要明确地看到集合 $\mathbb{T}$ 内部的元素在映射 g 之下导致的原像-像关系）；而且，这两个视角之间还可以转换。这一点非常重要，未来会启发一个非常非常的数学结构“范畴”的提出。

插入群的两个视角的示意图：三个集合 T，两次映射操作，整合起来，成为复合起来的映射，显示每个元素；三个点，每个点代表一个集合，两个映射箭头符号，不显示元素，直接作用到集合上，合起来成为复合映射；一个点，两个映射箭头符号，合起来成为一个符合映射的箭头。

7.5 推荐阅读材料

7.6 作业

7.7 本章小结

可逆映射和解方程。

第八章　逻辑学初步

8.1　从集合到命题

一个命题是关于对象之间关系的一个判断句。我们已经知道集合可以用来清楚地指代对象。那么，集合可以用来指代命题吗？可以，通过集合之间的关系。既然命题是对象之间的关系，对象对应着集合，则对象之间的关系就成了集合之间的关系。

让我们从日常语言的例子开始。注意，日常语言的模糊性比较大。例如，我们说，“人总是要吃饭的”[1]。这是一个判断句，是“人”和“需要吃饭的”之间的关系的一个命题。在这里，假设“人”这个集合是明确的，并且进一步明确为“活着的和还希望继续活着的人”。也就是说，我们假设这个集合是可以定义的，具有明确的元素的[2]。我们再来看看“需要吃饭的”这个集合是不是可以定义。首先，我们需要把“饭”拓展和明确一下。原则上，如果指的是大米饭，或者中餐、法餐等任何一个具体的餐饮的种类，则人其实可以不吃这样的具体的饭的种类的。因此，我们把“饭”看作是“能量输入”[3]。于是，“人总是要吃饭的”就成了“活着和还希望继续活着的人总是需要摄入能量的”。这里还有一个程度副词“总”需要进一步明确一下。“总”在这里有两个意思：“都”和“总有一天，也就是很长时间以后，最终”。因

[1]这就是“人是铁，饭是钢，一顿不吃饿得慌”这句话的朴素说法。

[2]实际上，还真的不一定满足明确性的要求。

[3]或者也可以更具体一点“生物形式的能量输入”，从而来和“电能”之类的能量输入区别开来。不过，为了讨论的方便，放大一点到“能量输入”也可以。

此合起来就是，“每一个活着和还希望继续活着的人过了足够长时间以后是需要摄入能量的”。我们把“每一个活着和还希望继续活着的人”看作集合 A，把“过了足够长时间以后需要摄入能量的”事物看作集合 B。然后，我们说，这个命题被表示为“$A \Rightarrow B$”，其含义是任何一个来自于 A 中的元素，必然属于 B。也就是集合具有包含关系 $A \subset B$。

那么，这个命题是否成立呢？也就是集合 A, B 是否真的具有包含关系 $A \subset B$ 呢？这是另外一个问题，称作命题的检验或者证明。这可以通过逻辑检验和事实性检验来完成。我们稍后正式学习逻辑的时候，我们再来讨论如何判断命题是否成立的问题。

让我们再举一个数学语言表达的例子——“所有的偶数都是整数”。其含义，用集合的语言就是，偶数的集合（记作 $A_{2\mathbb{Z}}$）中的任何一个元素，必然是整数的集合（记作 $\mathbb{Z}$）中的元素，也就是 $A_{2\mathbb{Z}} \subset \mathbb{Z}$。

通过这两个例子，我们发现，命题就是集合之间的包含关系。

8.2 什么是和为什么需要逻辑？

让我们从一个日常生活的例子开始。我们假设“天下雨，地就湿”成立，来看看“天不下雨，地就不湿”、“地湿了，天下雨了”、“地没有湿，天没有下雨”这三句话是否成立。

注意，我们没有一开始就去检验“天下雨，地就湿”是否成立哦，而是假设“天下雨，地就湿”成立。当然，如果你愿意，我们也可以先来做一个“天下雨，地就湿”是否成立的事实性检验。既然“天下雨”了，肯定有降水。这个水肯定都有地方去。当然，我们可以把降水区域全都保护起来，把雨水收集起来，一点都不让这些个雨水接触地面。但是，一般情况下，这样做的成本也太高了，也不太可能做到对降水区域的完全覆盖。因此，在日常语言的范围内，我们认为“如果天下雨了，那么（总有那么一小块）地（地球的表面）就湿（沾了雨水）”成立。你看，一旦进入到事实性检验，我们需要关注很多细节。不过，让我们暂时接受“天下雨，地就湿”成立。

我们关注的问题是，假设“天下雨，地就湿”成立，来看看“天不下雨，地就不湿”、“地湿了，天下雨了”、“地没有湿，天没有下雨”这三句话是否成立。

“天不下雨，地就不湿”是否成立，我们从逻辑上不能做判断。但是，从事实上，洒的水浇的水，以及更多其他方式，都可以导致地湿。因此，“天不下雨，地就不湿”是不成立的。但是，我们还是没有回答，是不是可以从假设“天下雨，地就湿”成立来得到“天不下雨，地就不湿”。我们稍后回到这个问题。

“地湿了，天下雨了”是否成立，我们从逻辑上不能做判断。但是，从事实上，地湿了可以来自于洒的水浇的水，以及更多其他方式，而不一定是天下雨。因此，“地湿了，天下雨了”是不成立的。但是，我们还是没有回答，是不是可以从假设“天下雨，地就湿”成立来得到“地湿了，天下雨了”。我们稍后回到这个问题。

“地没有湿，天没有下雨”是否成立，我们来试试从逻辑上做一个判断。这个命题在同样的前提（“地没有湿”）下，如果不成立，则肯定结论是“天下雨了”。但是，根据“天下雨，地就湿”，我们发现，这个前提“地没有湿”就不成立。任何一个命题，前提的成立是一个假设，也就是“地没有湿，天没有下雨”的含义是“如果地没有湿，那么天没有下雨”，因此前提不能不成立的[4]。于是，我们得到了“地没有湿，天没有下雨”不成立的结论是不成立的。也就是，“地没有湿，天没有下雨”只能是成立的[5]。

那么，是不是可以把前面的四个命题之间的关系，尤其是，其中第一个可以用来证明第四个这样的关系，推广到一般的命题上去，从而更多地时候我们可以不依赖于事实性检验来判断命题的真假呢？也就是说，如果我们从事实性判断已经肯定了某个命题的成立[6]，则我们就可以从这个命题出发来论证其他命题是否成立。

这样大大增加了判断命题是否成立的效率。这也使得研究者可以实现分工，一部分人从事事实性检验的工作，另一部分人从事按照已知甚至已经假设成立的命题出发来论证其他命题的工作。其实，后一部分就是数学家和各个学科的理论研究者，前一部分是各个学科的实验研究者。当然，除了进行对后一部分的研究者假设的和得到的命题做检验，前一部分研究者

[4]这里，我们用到了反证法，我们会稍后用集合的语言来展示反证法确实是成立的。

[5]这里我们用到了逻辑的一个性质：一个命题只能要么成立要么不成立，两者必须而且只能取其一。后面，我们也会用集合的语言来说明这个性质。

[6]事实性判断是否可以推断出概念性命题成立，这是一个问题。一般来说，事实性判断只能推断一个概念性命题不成立。后面我们会回到这个问题。

还给后一部分研究者提供启发，构建新的命题体系的启发。

逻辑就是由于这个更加高效率地判断命题是否成立，从而尽可能地依靠非事实性检验，而建立起来的知识体系。回到前面提到的四个命题，用数学符号来表示就是，原命题 $R=1\Rightarrow W=1$，否命题 $R=0\Rightarrow W=0$，逆命题 $W=1\Rightarrow R=1$，逆否命题 $W=0\Rightarrow R=0$。其中 $R=1$ 表示下雨，$R=0$ 表示没有下雨，$W=1$ 表示地湿，$W=0$ 表示地不湿。更一般地，从一个原命题 $A\Rightarrow B$ 出发，我们把否定了前提否定了结论的命题 $\overline{A}\Rightarrow\overline{B}$ 称作否命题，把颠倒了前提和结论的命题 $B\Rightarrow A$ 称作逆命题，把否定了前提否定了结论的又颠倒了前提和结论的命题 $\overline{B}\Rightarrow\overline{A}$ 称作逆否命题。我们可以看到，逆否命题的逆否命题就是原命题。也就是原命题和逆否命题互为逆否命题，或者称作具有逆否命题关系。可以检验，否命题和逆命题之间也是逆否命题的关系。

习题 8.1 (否命题和逆命题). 证明否命题和逆命题之间互为逆否命题。

稍后我们就会看到，逻辑告诉我们说，原命题和逆否命题完全等价。那么，逻辑是如何来展示这样的等价性的呢？或者说，更一般地，是如何从一个命题来证明或者导出另一个命题的呢？提供方法来帮助研究命题之间的关系，就是逻辑。

从上面这些例子中，我们粗糙地看到了什么是逻辑以及为什么我们需要逻辑。人们研究科学根本上，是为了得到一些和现实相符的命题，甚至在这个现实真的出现之前。我们固然可以通过把这个现实实现，然后再来和这个命题相比较的方式来检验每一个这样的命题。但是，我们还可以从基于每一个命题涉及到的具体事实的判断走到基于超越事实的一般性的判断方法——这就需要逻辑。效率更高，或者说，把逻辑判断和事实判断分开进行。如果逻辑判断都通不过，则不再需要进行事实判断。或者说，只要从事实开始构建出来了一个和当前所有已知的事实不矛盾的命题的体系——所谓体系就是有一些基本的命题当作基础其他的命题建立在这些基本的命题之上，则先运用这个命题的体系来判断命题的真假。最后，我们只需要把这些基本的命题做一个事实性检验，而不需要每一个命题都做事实性检验。

当然，在关于现实世界对象的科学研究中，我们强调每一个依赖逻辑体系得到的命题，仍然得做事实性检验；同时，我们还强调，基于已有事实构建的逻辑体系本身一旦发现了新的和这个逻辑体系冲突的新的事实，则

得重新构建包含了新的事实的逻辑体系。

8.3 逻辑三律和演绎逻辑

前面，我们已经看到，命题就是集合包含关系，命题可以成立——称做这个命题的真值为真，也可以不成立——称做这个命题的真值为假。下面，我们来介绍逻辑的基本概念。

让我们首先从逻辑变量开始。我们把一个命题 $A \Rightarrow B$ 的真值当作一个逻辑变量 l，其可能为真记作 $l = T$ 或者 $l = 1$，可能为假记作 $l = F$ 或者 $l = 0$。用集合的语言，$l = 1$ 意味着 $A \subset B$，$l = 0$ 意味着 $A \not\subset B$，也就是存在着 $\alpha \in A, \alpha \notin B$。

我们首先说，任何一个逻辑变量 l 其取值必须是 $l = 1$ 或者 $l = 0$，并且不能两者同时成立。前者称为排中律，后者称为无矛盾律。注意 $l = 1$ 或者 $l = 0$ 互为否定（$l = 1 \Leftrightarrow \bar{l} = 0$），如果两者可以同时成立，就产生了矛盾。排中律的意思是，原命题（$l = 1$）和否命题（$\bar{l} = 1$，也就是 $l = 0$）必须有一个成立。我们可以直接从逻辑的基本性质出发来构建逻辑的知识。不过，既然我们学习了集合，我们还可以从集合来理解或者说建构逻辑的知识。

为什么 $l = 1$ 和 $l = 0$ 必须而且只能取其一？让我们回到命题 $A \Rightarrow B$ 的集合表示 $A \subset B$，也就是判断是否 $\forall \alpha \in A, \alpha \in B$ 的问题。首先，根据集合元素的确定性，$\forall \alpha \in A$，$\alpha \in B$ 的问题是一个可以明确回答的问题。如果“是”，则 $l = 1$，如果“否”则 $l = 0$。而且，这个答案不能是“是、否”两者都成立，因为集合的元素确定性说了“一个元素 α 是否属于一个集合 B 有且只有一个答案，要么是要么否”。因此，逻辑中的排中律和无矛盾律都来自于集合元素的确定性。

我们还注意到，集合元素还有一个互异性，也就是一个东西只在一个集合中出现一次，只要是这个东西（对应着的记号、名称）那么就是这个东西。这就是逻辑的同一律。

好了，我们有了逻辑三律——同一律、排中律和无矛盾律，并且它们都来自于集合的性质，也就是集合元素的确定性和互异性。

8.3.1 三段论

有了作为逻辑知识的基础的逻辑三律实际上来自于集合知识这个认知，我们就可以来讨论三段论这个演绎证明的形式了。在后面的讨论中，绝大多数情况下，我们都从集合的知识开始。

让我们从三段论的一个著名例子开始：

大前提：所有人都是要死的，
小前提：苏格拉底（Socrate）是人，
结论：所以苏格拉底（Socrate）是要死的。

所有的三段论都可以被标准化为这样的“大前提、小前提和结论”的三段的格式，所以被称为三段论。用集合的语言，我们重新表述一下，就是

大前提：$H \subset M$，
小前提：$s \in H$，
结论：$s \in M$。

其中 H 表示所有人的集合，M 表示所有要死的（具有有限的寿命的）事物的集合，s 代表苏格拉底（Socrate）。

插入苏格拉底是要死的的示意图，两个具有包含关系的集合，一个元素，标上图中的符号。

我们来看看为什么这个成立，从集合的角度。首先，已知 $H \subset M$，也就是 $\forall h \in H, h \in M$。其次，我们还已知 $s \in H$。于是，自然地，$s \in M$，也就得到了最后的结论。也就是说，所谓逻辑中的三段论演绎法不过就是集合的包含关系的基本定义。只要假设集合元素的确定性（以及平庸的互异性和无序性）是正确的，那么，自然三段论演绎法就是正确的。

注意，三段论演绎法的每一段话都是一个命题，并且从前两个命题，得到了后一个命题。也就是说，三段论演绎法可以用来建立命题之间的导出关系，做命题的论证。

编制一些三段论的应用题

8.3.2 逆否命题的等价关系和反证法

反证法是逻辑证明中另一个常用的方法。我们来看看它的基础是什么，为什么成立。为此，我们需要先讨论逆否命题的等价关系。

我们已经知道，原命题 $A \Rightarrow B$ 的集合表示是 $A \subset B$。同样地，命题 $\overline{B} \Rightarrow \overline{A}$ 的集合表示是 $\overline{B} \subset \overline{A}$。其中，$\overline{A}, \overline{B}$ 分别是集合 A, B 的补集。注意，既然可以讨论补集，则肯定有一个天然定义好的全集 I。例如，在“天下雨”的例子中，“天下雨”的反面就是“天不下雨”，合起来的全集就是“天下雨或者不下雨”。下面，我们来证明 $A \subset B$ 就是 $\overline{B} \subset \overline{A}$。

我们先通过画图来建立一个直觉，然后再用集合的语言来证明。首先，从 $A \subset B$ 看到，集合 A 被包含在 B 里面，两者都在全集 I 里面。于是，我们自然可以看到 A 的补集更大，包含了 B 的补集。一个集合的补集是那些不在这个集合但是在全集里面的元素的整体。

插入 $A \subset B$ 和 $\overline{B} \subset \overline{A}$ 的示意图。

定理 8.1: 逆否命题集合关系

逆否命题集合关系]$A \subset B \Rightarrow \overline{B} \subset \overline{A}$。

证明 8.1. 选择任意一个 $\overline{B}$ 中的元素 x，$x \in I, x \notin B$。我们来证明 $x \in \overline{A}$。这就要求我们证明 $x \notin A, x \in I$。第二条自然成立。我们来看第一条。由于任意的 x，两种关系中 $x \in A$ 和 $x \notin A$ 有且只有一种成立。我们只要证明了 $x \in A$ 不成立，则肯定 $x \notin A$ 成立。

根据 $A \subset B$，任意的 $x \in A$ 有 $x \in B$。我们这里已经有了 $x \notin B$，所以，$x \in B$ 不成立。因此，$x \in A$ 也不成立。

在证明过程中，我们多次用到了集合元素的确定性，或者说等价地，逻辑三律。

定理 8.2: 逆否命题等价性

逆否命题等价性]$A \Rightarrow B \Longleftrightarrow \overline{B} \Rightarrow \overline{A}$。

证明 8.2. 根据命题和集合的关系，$A \Rightarrow B$ 等价于 $A \subset B$，$A \subset B \Rightarrow \overline{B} \subset \overline{A}$，而 $\overline{B} \subset \overline{A}$ 和 $\overline{B} \Rightarrow \overline{A}$ 等价，因此，$A \Rightarrow B \Longleftrightarrow \overline{B} \Rightarrow \overline{A}$。

或者，连起来写成

$$A \Rightarrow B \Longleftrightarrow A \subset B \Longleftrightarrow \overline{B} \subset \overline{A} \Longleftrightarrow \overline{B} \Rightarrow \overline{A}. \tag{8.1}$$

或者，也可以写成更方面阅读的形式，

$$(A \Rightarrow B) \Longleftrightarrow (A \subset B) \Longleftrightarrow \left(\overline{B} \subset \overline{A}\right) \Longleftrightarrow \left(\overline{B} \Rightarrow \overline{A}\right). \tag{8.2}$$

既然原命题和逆否命题是完全等价的，那么，当我们证明原命题 $A \Rightarrow B$ 有困难的时候，我们就可以尝试来证明逆否命题，也就是 $\overline{B} \Rightarrow \overline{A}$。也就是先试着假设结论不成立，看看能否推导出来前提也不成立。如果可以，则仍然相当于证明了原命题。

编制一些反证法的应用题

8.3.3 逻辑变量的运算

有的时候，我们需要把若干个命题合起来组成一个整体的命题，例如“天下雨，地就湿，而且地就滑”[7]，例如“偶数是整数，并且不是奇数”，“合数是整数，并且不是质数”，“偶数是整数，并且是合数”，“偶数是质数，或者合数”等等。这时候，我们就需要来做若干个逻辑变量的运算。基本的逻辑变量的计算包含三种，非 $(\neg l)$，与 $(l_1 \wedge l_2)$，或 $(l_1 \vee l_2)$。它们的计算规则分别是

$$\neg l = \begin{cases} 0 & l = 1 \\ 1 & l = 0 \end{cases} \tag{8.3}$$

就是把逻辑变量的值反过来。

$$l_1 \wedge l_2 = \begin{cases} 1 & l_1 = 1, l_2 = 1 \\ 0 & \text{其他} \end{cases} \tag{8.4}$$

就是只有两个逻辑变量的取值都为真的时候，才为真，否则为假。

$$l_1 \vee l_2 = \begin{cases} 0 & l_1 = 0, l_2 = 0 \\ 1 & \text{其他} \end{cases} \tag{8.5}$$

就是只需要两个逻辑变量的任意一个取值都为真的时候，都为真，否则为假。

[7]同样地，如果要成为命题，首先得是集合，含义要明确，因此，如果我们真的来讨论这个命题我们还得定义“滑”。

8.4 逻辑是共通的普适的吗？

在接受集合知识的前提下，只有这一种，是普适的。可能是人类的大脑受到现实的刺激而演化从而自洽形成的。

还有人在追求可能的另外一套逻辑。同样，只要是基于公理化体系的，在公理本身被接受的范围内，可以自洽。但是，迄今为止，没找到这样一套不同的逻辑。

8.5 逻辑和批判性思维

学会了逻辑，首先就是用来给命题做效率更高更准确的判断的。也就是前面提到的通过命题来证明命题，而不是每个命题都依赖于事实性检验。毕竟，事实性检验的成本高很多，但是通过命题来证明命题是可以在我们的脑子里面来完成的，往往成本低很多。

但是，学会了逻辑，还有可以支撑一个更加重要的思维方式或者说思维习惯，那就是批判性思维，也就是“我从来不把没有经过我自己的理性检验的命题放入到我的认知结构之中”。这里的理性检验包含逻辑论证和事实性检验。事实性检验一般通过做实验来进行。

例如，我可以告诉你“地球是圆的”、“地球围着太阳转”、“三角形的内角和是 180°”、“两直线平行则同位角相等”等等一系列的命题，但是，只有经过你自己的逻辑论证可以被证明的，或者是通过了事实性检验的命题[8]，你才能接受它们，把它们放到你的脑子里面，成为你进一步提出问题和解决问题的基础。

那么，如何来做这个逻辑检验呢？这就要靠知识的公理化体系。在数学和科学知识中，人们往往从定义和公理来构建整个知识体系。也就是说，无论受到现实的启发，还是受到现有知识体系中的问题（例如发现，应该由某个定理，但是目前这个定理还没有得到证明，逻辑链条还不通）的挑战，还是受到来自于自身想象力的启发，人们会提出一些定义，提出一些公理也就是假设成立的命题，然后，人们企图从这些定义和公理出发来得到定理。

[8]再次强调，事实性检验只能否定命题，不能真的证实命题。见后续逻辑和科学的关系的讨论。

我们学习的时候也一样，首先，至少要能够自己来从定义和公理来推导出来定理。其次，更高的要求是，我们也能够受到相应的启发自己去尝试“独立地再一次”提出定义或者公理，猜测出来定理，然后再来证明这些定理。只有这样自己来发展和构建起来这个公理化的知识体系，我们将来在遇到问题的时候，就会有方向和有意识地运用合适的定义、公理和定理来提出和解决问题，甚至运用那些启发来提出和解决问题，而不是去凑去试，看哪一个定义、公理和定理正好可以解决某个问题。更进一步，如果我们有受到启发的经验，我们将来遇到新场景可能也会再次受到启发，从而提出真正的新概念、新公理，猜测和证明新定理，甚至发现新的启发。自己构建起来和体会到知识体系，自己受到了和体会到了促使发展新概念、新公理和新定理的启发，从而具有了自己来面对新场景、再次受启发从而提出和解决问题的可能性，甚至形成这样的习惯和随时准备好这样做的意愿，就是理解型学习的目标 [2]，就掌握了高层知识生成器。

下面我们来看几个把演绎逻辑用于批判性思维的例子。

实际上，基于集合和命题的整个知识体系，有可能会有问题——并不是所有的命题都可以判断其是否成立。我们在下面这个小知识中来给大家做一个小小的展示。

小知识 8.1 (说谎者悖论和理发师悖论). 有一个人经常说谎，然后，他宣称“我在说谎”。我们来看看这句话是否成立。首先，如果这句话成立，那么“我在谎言”本身是谎话，谎言的意思是这句话不成立。产生了矛盾。如果这句话不成立，那么，说谎的反面就是说真话，也就是“我在说真话”。只要这句话是真话，我们又回到了第一种情况。在此产生了矛盾。

在这里，我们用到了一个逻辑的基本性质，任何一个命题其要么成立要么比成立，两者必须而且只能居其一。按照集合的基本要求，给定任何一个集合，任何一个元素要么属于这个集合要么不属于这个集合。我们如果把所有的成立的命题构成一个集合 $\mathbb{T}_p$。那么任何一个命题 l，根据集合元素的确定性，要么属于这个集合要么不属于这个集合。

于是，上面这个说谎者悖论表明了：“任何一个命题其要么成立要么不成立，两者必须而且只能居其一”必然导致“有的命题是否成立不能被判断”，或者“有的命题的真值不是成立和不成立两者必须而且只能取其一的”。

我们再来看另一个例子：小镇上有一个理发师，他宣称他不给自己理发的人理发，但是同时他给所有不是自己理发的人理发。这个规则在不牵涉到他自己的时候是逻辑上完全没有问题的。但是，我们来看看，他怎么把这个规则用于他自己。如果他决定给他自己理发，则他就是那个给自己理发的人，按照其规则，他不能给自己理发的人——在这里就是他自己——理发，因此，产生了矛盾。如果他决定不给自己理发，则按照其规则，他就必须给不给自己理发的人——在这里就是他自己——理发，因此，有产生了矛盾。因此，这条规则是自我矛盾的。

Russell 给这个理发师悖论构建了一个集合的语言来描述。首先，定义一个集合 $\mathbb{S}$ 为“所有的不包含自身的集合”，也就是 $\mathbb{S} = \{x | x \notin x\}$。这个定义中用到了对于一个集合自身是否包含在这个集合之中的判断。按照集合的元素的确定性——给定一个集合，对于任何一个元素，这个元素在属于和不属于这个集合之间必须而且只能有一项成立，这个判断确实可以被执行。

当然，我们之间遇到的集合，都不会去判断这个集合本身是不是这个集合的元素。例如，自然数集合的元素是一个个的自然数，从来不会说，自然数集合本身是否属于自然数集合。但是，这个问题仍然允许被问。

接着 *Russell* 问，那么，$\mathbb{S}$ 是否属于 $\mathbb{S}$ 呢？假设 $\mathbb{S} \in \mathbb{S}$，那么按照 $\mathbb{S}$ 的定义 $\mathbb{S} = \{x | x \notin x\}$，刚好违反了这条定义，因此，$\mathbb{S} \notin \mathbb{S}$。但是，只要 $\mathbb{S} \notin \mathbb{S}$，则按照 $\mathbb{S}$ 的定义，这个 $\mathbb{S}$ 就刚好符合这个定义，因此，$\mathbb{S} \in \mathbb{S}$。矛盾。

看起来，所有的矛盾的根源都在集合元素的确定性——也就是给定一个集合，对于任何一个元素，这个元素在属于和不属于这个集合之间必须而且只能有一项成立——这一条上。

那这个问题到底怎么解决呢？留给你去思考，或者进一步学习。

顺便，从批判性思维的角度，这其实是运用逻辑进行了对集合论和逻辑的本身的批判性思维。

8.6 逻辑和数学、逻辑和科学

Karl popper 的可证伪性。

数学的语言为讨论逻辑学问题提供基础，逻辑学的基础问题成为数学

的研究对象。

8.7 推荐阅读材料

8.8 作业

8.9 本章小结

第九章　一元函数

当一个映射是建立在两个关于数的集合之间的时候，我们就称这个映射为函数。在这一章中，我们来学习什么是一元函数以及几个一元函数的例子。

9.1　一元函数当作一个吞吐机

一元函数是把一个数变成另一个数的一个映射。例如，$f(x)=x+1$，是一个从实数集合到实数集合的映射。我们可以验证其满足映射的要求：对于任意一个实数 $x\in\mathbb{R}$，存在着唯一的 $f(x)=x+1\in\mathbb{R}$。顺便，为了熟悉一下满射和单射的概念，我们还可以验证这是映射是满射和单射。

习题 9.1 ($f(x)=x+1$ 是满射)**.** 证明 $f(x)=x+1$ 是满射，也就是对于任何一个 $y\in\mathbb{R}$，可以找到原像 $x\in R$ 使得 $f(x)=y$。

习题 9.2 ($f(x)=x+1$ 是单射)**.** 证明 $f(x)=x+1$ 是单射，也就是对于任何一个 $x_1\neq x_2\in\mathbb{R}$，$f(x_1)\neq f(x_2)\in\mathbb{R}$。

注意，当我们把 $f(x)=x+1$ 看作是从非负整数集合 $\mathbb{Z}^*$ 到 $\mathbb{Z}^*$ 的映射的时候，就不再是满射了。因此，一个映射是不是满射和单射是和这个映射的定义域和陪域相关的。有的时候，人们也定义一个叫做值域的集合，也就是所有的像的集合 $f_X=\{y|y=f(x),\forall x\in X\}$。

习题 9.3 ($f(x)=x+1$ 是满射)**.** 证明 $f(x)=x+1:\mathbb{Z}^*\mapsto\mathbb{Z}^*$ 不是满射，并且找出来这个函数的值域。

这个一元函数收到或者说吞进去一个数，吐出来一个数，特别像一个吞吐机，特别像一个盒子。很可能正是由于这一点，函数被成为函数："函"（古代字形为）的意思是一个盒子[1]，"数"的意思是这个盒子吞进去和吐出来的都是数。顺便，大多数数学和科学概念的名称，和其含义，是具有相当程度的对应关系的。因此，学习的时候，概念的名称也不仅仅是一个任意选的记号，而往往是可以理解的。

9.2 一元函数当作一个关系

函数，或者说映射，还有另一种理解方式，那就是关系。朴素地说，两个东西之间的关系就是一个东西总是伴随着另一个东西出现。例如，在一元函数这里，就是 $(x, f(x))$ 总是一起出现，给一个 x 后面就有一个 $f(x)$ 跟着。更一般地，关系可以定义为集合的直积集合的子集。

为此，我们先来定义两个集合的直积集合。

定义 9.1 (集合的直积). 给定一个集合 X 和一个集合 Y（可以是同一个集合），我们把集合 $X \times Y \triangleq \{(x, y) \,|\forall x \in X, \forall y \in Y\}$ 称作集合 X 和 Y 直积集合，也称为 *Cartesian* 积（卡氏积）[2]。

关系指的是 $X \times Y$ 的一个子集。回到朴素的语言，来自于 X 集合的 x 和集合 Y 的 y 的关系指的就是一个有序对 (x, y)。于是，一元函数就是一个特定的关系，通过函数 $y = f(x)$ 构建的特定的关系有序对 $(x, f(x))$。

9.3 平面直角坐标系

前面我们已经学习过数轴——实数和数轴上的点一一对应[3]，现在，我们用集合的直积来构建一个平面直角坐标系。平面直角坐标系的作用是让

[1] "函"甲金文由倒"矢"（后讹变为"水"）、一个圈（现写为"凵"）和上面可以手提的小环（现写为"了"）组成，像收藏箭矢的袋子或盒子，本义是箭袋、箭匣。见"汉字理解型学习"网站，https://www.learnm.org/。转引自见"汉语多功能字库"网站，https://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/。

[2] 这个名字来自于笛卡尔（Descartes），这个第一次用数轴的卡氏积来得到平面的人。

[3] 尽管这个命题我们还没有证明。这个证明留给大学阶段的数学来完成。

数对 (x,y) 和平面上的点一一对应。

它是这样来建立这个一一对应关系的。首先，我们已经有了实数集合 $\mathbb{R}$ 和 $\mathbb{R}$ 的卡氏积，也就是有了数对 (x,y) 的集合。接着，我们把 x 轴画成对平方向的一条直线，并且在 $x=0$ 处画一条垂直于 x 轴的直线，称为 y 轴。同时，这个交叉点也当作 y 轴的零点。这样我们就有了坐标轴——每一个坐标轴都是一条在自己的方向上的数轴。由于这两条坐标轴成直角关系，构成了一个平面，所以称为平面直角坐标系。

现在，我们来看看，为什么叫做“坐标系”——也就是给定任何一个数对 (x,y)，如何找到一个在上面的平面上的对应的点。我们这样来找：首先，在 x 轴上，找到数对 (x,y) 中的相应的 x 值，过这个点做一条平行于 y 轴的直线；接着，在 y 轴上，找到数对 (x,y) 中的相应的 y 值，过这个点做一条平行于 x 轴的直线；两条直线的交点就是数对 (x,y) 在平面中所对应的点。根据几何知识，这个交点必然存在，并且唯一。

插入平面直角坐标系，标注一个点 $(1,1)$ 当例子。

两条坐标轴把整个平面分成了四个区域，我们也给四个区域各自一个名字：$x>0,y>0$ 的区域被称为第一象限，$x<0,y>0$ 的区域被称为第二象限，$x<0,y<0$ 的区域被称为第三象限，$x>0,y<0$ 的区域被称为第四象限。由于是基于卡氏积来定义的，平面直角坐标系也经常被称为笛卡尔坐标系。

这里我们需要用到一些几何知识，例如如何做垂线、平行线，两条不相同的直线相交要么只有一个交点，要么没有交点。我们假设这部分知识都已经在小学阶段的几何知识里面学习过。我们也会将来在中学阶段的几何知识部分中心来学习，更加严格地有系统性地来学习。

9.4 一元函数当作一个图形

我们还可以把一元函数看作一个图形，只需要借用上面提到的两个实数集合通过卡氏积构成的实数对平面，然后把每一个关系对 $(x,f(x))$ 都画到实数对平面上去。画的方法就是在 x 轴（通常选水平方向的那个实数轴）上找到 $(x,f(x))$ 中的 x 对应的点，在 y 轴（通常选垂直于水平方向的那个实数轴）上找到 $(x,f(x))$ 中的 $f(x)$ 对应的点，然后做过 x 点的平行于

y 轴的直线同时做过 $f(x)$ 点的平行于 x 轴的直线，两直线相交的点就对应着 $(x, f(x))$。把所有的这样的 $(x, f(x))$ 都画上去就得到了对应着函数 $y = f(x)$ 的图形。

这样，我们就从函数得到了函数的图形。现在，我们反过来，从函数图形试试来得到函数。

所谓得到一个函数，就是，对于任意的 $x \in X$，我们都得找到一个而且是唯一的一个 $y \in Y$。这里的 X, Y 是 $\mathbb{R}$。我们这样来完成这个任务：对于任何一个 x，我们先找到 x 轴上对应的点，然后，过这个点做平行于 y 轴的直线。只要函数在这个点有定义，则这个直线必定和图像交于一点[4]。由于函数是映射，映射要求给定一个 x 有且只有一个 $y = f(x)$，这个函数的图形，在任何一个 x 点，做平行于 y 轴的直线，这条直线和函数图形只能有一个交点。找到这个交点，过这个交点做平行于 x 的直线。这个直线和 y 轴的交点就得到了函数值，也就是 x 的像，$y = y = f(x)$。

那为什么我们要把一元函数看作图形，而且是一个“在任何一个 x 点，做平行于 y 轴的直线，这条直线和图形只能有一个交点”的图形呢？因为从函数的表达式来学习函数，往往给我们造成一个错觉，好像只有能够写出来函数表达式的才叫做函数，例如 $f(x) = x + 1$，$f(x) = x^2$ 等等等。但是，通过把函数和“在任何一个 x 点，做平行于 y 轴的直线，这条直线和图形只能有一个交点”的图形联系起来，我们发现，实际上，只要有一个满足这个要求的图形，我们就定义了函数。更进一步，为什么要求这个性质的图形呢？不过就是为了通过 x 找到 $y = f(x)$ 的上面那个作直线找交点的过程能够找到一个并且只有一个 $y = f(x)$ 而已。

9.5 一元一次函数

前面我们已经提到了一元函数看起来很象一个吞吐机，一元函数实际上是一个从数的集合到数的集合的映射，有的时候一元函数可以表达为一个表达式 $y = f(x)$，一元函数实际上就是在平面直角坐标系上的一条满足

[4]如果函数在这个点有定义而且平行于 y 轴的直线和函数图像没有交点，则函数图像必定也是平行于 y 轴的直线。然而，平行于 y 轴的直线不构成映射——同一个 x 对应着无穷多个 y 的值。

特定要求的曲线图形。

在实际使用中，表达式 $y = f(x)$ 形式的定义往往更加方便。因此，下面我们开始从表达式 $y = f(x)$ 形式来讨论一元函数。这个时候，我们把 x 称作自变量，y 称作因变量。意思是 x 可以任意取，自己决定取什么，但是 y 呢随着 x 的确定就确定了。

其中最简单的是常数函数，也就是 $y = f(x) = c$。$c \in R$ 是一个常数，例如 $c = 0$，$c = 1$。这个函数的因变量的取值 y 不依赖于自变量 x。这个函数画成图形就是平行于 x 轴的直线，和 y 轴相交于 y 轴上的 $y = c$ 这个点。

不过常数函数实在太简单了，实在没什么可研究的。我们来看一个稍微复杂一点的函数——一元一次函数，$y = kx + b$（$k \neq 0$，否则就回到了前面的常数函数）。首先，这个函数过点 $(0, b)$，和 y 轴相交。这样的交点只有一个，并且我们称 b 为截距——和 y 相交和 $y = 0$ 点距离为 b 的点。接着，我们发现，如果忽略这个截距，也就是暂时让 $b = 0$，我们得到的是正比例函数 $y = kx$。我们很清楚这是一条直线，并且过原点。反过来，如果我们已经得到了正比例函数 $y = kx$ 的图象，那么我们只需要把这个图像整体加上一个 b——这体现为把整个函数图像上下平移距离 b，就能够得到 $y = kx + b$ 的图象。

因此，现在我们来关注 $y = kx$ 的图象。我们从一个特利 $y = x$ 开始。通过描点画图法，我们可以猜测，$y = x$ 就是平面直角坐标系中 $45°$ 方向的一条直线：直线经过一三两个象限，而且从正中间经过。描点法的具体过程如下：任意选择一些个 x_i，算出来每一个 x_i 对应的 $y_i = f(x_i)$，组成若干点对 $(x_i, f(x_i))$，把这些点对一一画到图上，尽量用光滑的曲线或者折线把这些点按照 x 的大小顺序相连，看看构成了什么图形。

类似地，$y = -x$ 就是平面直角坐标系中 $135°$ 方向的一条直线：直线经过二四两个象限，而且从正中间经过。

对于更一般的 $y = kx$，如果 $k > 1$，则还是在一三象限中的一条直线，但是 y 增加的更快，因此，比 $y = x$ 更加靠近 y 轴；如果 $0 < k < 1$，则还是在一三象限中的一条直线，但是 y 增加的更慢，因此，比 $y = x$ 更加远离 y 轴而靠近 x 轴；如果 $-1 < k < 0$，则还是在二四象限中的一条直线，但是 y 的绝对值增加的更慢，因此，比 $y = -x$ 更加靠近 x 轴；如果 $k < -1$，则还是在二四象限中的一条直线，但是 y 的绝对值增加的更块，因此，比

$y=-x$ 更加靠近 y 轴。

这样，我们就得到了 $y=kx+b$ 的图象。经常我们需要在函数图像和函数表达式之间做转换。

9.6 一元一次方程的几何含义

一元一次方程是一元一次函数反过来，也就是，在函数中我们从给定的 x 来计算 $y=f(x)$，在方程中，我们从给定的 y 来找到合适的 x 使得 $y=f(x)$。我们发现，这正好是逆映射。因此，方程的求解的前提就是映射是可逆的。

我们来看看，一元一次函数对应的映射是不是可逆的。我们先从映射是否可逆开始。首先，实数集合到实数集合上的一元一次函数是一个满射（留作作业），也是一个单射（留作作业），因此，肯定是一个可逆映射。

习题 9.4 ($f(x)=kx+b$ 是满射). 证明 $f(x)=kx+b$ $(k\neq 0)$ 是满射，也就是对于任何一个 $y\in\mathbb{R}$，可以找到原像 $x\in R$ 使得 $f(x)=y$。

习题 9.5 ($f(x)=kx+b$ 是单射). 证明 $f(x)=kx+b$ $(k\neq 0)$ 是单射，也就是对于任何一个 $x_1\neq x_2\in\mathbb{R}$，$f(x_1)\neq f(x_2)\in\mathbb{R}$。

我们再从函数图象开始。也就是，从给定的 y 能不能找到一个函数图象上的唯一的 x。我们这样来找：选 y 轴上的点 y，过 y 点做 x 轴的平行线，因为一元一次函数的图象也是一条直线，因此这个过 y 点做 x 轴的平行线必然和函数图象有且只有一个交点 P。除非 $k=0$，则 $y=b$ 是常数函数，其和 y 点做 x 轴的平行线可能没有交点或者完全重合。不过，在讨论一元一次函数的时候，我们总是先排除 $k=0$ 的情况。过交点 P，做平行于 y 轴的直线，和 x 轴交于点 Q。交点的 Q 的 x 值就是所要找的满足 $y=f(x)$ 的 x。这个交点唯一。因此，给定 y 的对应的 x 也唯一。因此，函数可逆。

我们再来从代数式看一元一次函数是否可逆。已知 y，要找到 x 使得 $y=f(x)=kx+b$ 成立。可以检验，我们只需要选 $x=\frac{y-b}{k}$，肯定满足上面的要求。当 $k\neq 0$ 时，$x=\frac{y-b}{k}$ 存在且唯一。因此，一元一次函数是否可逆。

我们还可以通过把每一步操作反过来来看看一元一次函数是否可逆。我们先看，给定 y 是否可以找到 z 使得 $y=z+b$ 成立。这个时候已知的 y

就是给未知的 z 增加了 b，因此，反过来就是 $z = y - b$。这里我们用到了加法的逆运算是减法。我们再来看，给定 z，是否可以找到 x 使得 $z = kx$ 成立。这个时候，已知的 z 不过就是未知的 z 称以 k，只要 $k \neq 0$ 这个计算就可以反过来，得到 $x = \frac{z}{k}$。这里我们用到了乘法的逆运算是除法。合起来就是 $x = \frac{y-b}{k}$。

将来判断更一般的方程是否有解和求解的时候，也类似，可以从映射是否可逆来考虑，从函数曲线的角度来考虑，从把每一个运算逆过来来考虑，或者直接检验不知道从哪里获得的解。

9.7 一元二次函数和一元二次方程

一元二次函数的表达式是 $y = ax^2 + bx + c$（$a \neq 0$，否则就回到了一元一次函数）。我们也从最简单的一元二次函数 $y = x^2$ 开始。

首先，从实数集合到实数集合上一元二次函数 $y = x^2$ 不是一个满射，因为对于任意的 $x \in \mathbb{R}$，$y = x^2 \geqslant 0$。也就是说，对于 $y < 0$ 的像，我们找不到原像。

其次，$y = x^2$ 也不是一个单射，因为存在着 $x_1 = x$ 和 $x_2 = -x$ 不同（只要 $x \neq 0$），但是 $y_1 = x_1^2 = x^2, y_2 = x_2^2 = (-x)^2 = x^2$，$y_1 = y_2$。

因此，$y = x^2$ 不是一个可逆映射，原则上不能求解一元二次方程。一方面，我们可以把方程的含义稍微扩大一下，从找到逆映射扩大为给定 y 找到所有的 x 使得 $f(x) = y$ 就可以，不需要唯一，不需要用到逆函数。但是，另一方面，如果我们希望能够写下一个方程的来表达式形式的解，注意表达式一般来说是个函数，那么，我们就希望能够用某种方式把一个方程的多个可能的解从形式上分开，给每一个形式一个表达式。

例如，$y = x^2$ 其实可以做到，只需要分开正负两个分支，也就是 $x_1 = \sqrt{y}$ 和 $x_2 = -\sqrt{y}$（$y \geqslant 0$），或者合起来记作 $x_{1,2} = \pm\sqrt{y}$（$y \geqslant 0$）。这里我们就有了两个分支，每个分支有一个函数形式，每一个分支上，函数可逆。

接着，我们再来看 $y = x^2$ 的图象。首先，我们说它的函数图象关于 y 轴对称，也就是把 x 替换为 $-x$ 函数值不变，也就是 $f(-x) = f(x)$。因此，我们只需要关注 $x \geqslant 0$ 这一侧的函数图象就行，另外一侧只需要通过轴对称图形来转化就可以得到。

为了得到 $y = x^2$ 在第一象限内的图象，我们先来试试描点法。插入描点画图过程。

通过描点法，我们发现，这是一个增函数，函数值随着 x 的增加而增加，并且增加的速度也在变大。我们也可以来证明一下这个结论：我们让 x 增加一个固定的量，例如 1，看看 y 的增加量[5]，也就是

$$f(x+1) - f(x) = (x+1)^2 - x^2 = x^2 + 2x + 1 - x^2 = 2x + 1. \tag{9.1}$$

也就是说，给 x 增加 1 个单位在 y 上的效果，在 x 大的时候更大。注意，我们只关注第一象限，也就是 $x > 0$ 的情况。

对于一元一次函数，例如 $y = g(x) = x$（因为在这里一次函数和二次函数需要放在一起比较，为了区别，把一次函数记作 g），这个增加的速度不依赖于 x，也就是

$$g(x+1) - g(x) = (x+1) - x = 1. \tag{9.2}$$

也就是说，给 x 增加 1 个单位在 y 上的效果，不依赖 x 的取值。

于是，我们发现，这个增加的速度要超过 $y = x$。我们已经知道，$y = x$ 是 45° 的直线，直线的斜率也就是增加的速率是一个常数。那么 $y = x^2$ 的每一个点的斜率，如果可以定义一个斜率的话，要不断地变大，而且要大于 $y = x$ 的斜率。也就是说，$y = x^2$ 的增加的速度，也就是曲线往 y 轴接近的速度，要远远大于 $y = x$ 曲线，而且 x 越大这个差异越大。

有了这个发现，再加上描点法，我们就可以大概把这个函数图画出来了。

搞清楚了特殊的 $y = x^2$ 的一切，我们再来看更一般的一元二次函数 $y = ax^2 + bx + c$。我们说更一般的一元二次函数 $y = ax^2 + bx + c$ 可以被转化为特殊的 $y = x^2$ 的情形，也就是

例 9.1 (从 $y = ax^2 + bx + c$ 到 $y = x^2$)**.** 把 $y = ax^2 + bx + c$ 被转化为接近(这里的接近指的是，将来可以通过平移和缩放等操作，来从后者的函数图

[5]将来我们学习了微积分会发现，这就是在计算导数，或者说微分，或者说曲线斜率。

象得到前者）$y = x^2$ 的形式。

$$ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right), \tag{9.3}$$

$$= a\left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a^2} + \frac{c}{a}\right], \tag{9.4}$$

$$= a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a}. \tag{9.5}$$

于是，我们发现，从 $y = x^2$ 的图象开始，我们：第一、先移动一下 x 轴，使得 $x \to x + \frac{b}{2a}$，也就是把 x 轴平移 $-\frac{b}{2a}$，就得到了函数 $y = \left(x + \frac{b}{2a}\right)^2$ 的图象；第二、再把得到的函数整体乘上一个 a，就得到了 $y = a\left(x + \frac{b}{2a}\right)^2$；第三，把前一步得到的图象整体上下移动也就是加上 $\frac{4ac-b^2}{4a}$，就得到了 $y = ax^2 + bx + c$ 的图象。

注意，由于 $y = x^2$ 有一个最小值，因此 $y = a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac-b^2}{4a}$ 也有一个最大值 $(a < 0)$ 或者最小值 $(a > 0)$。这个最值点的坐标是 $\left(-\frac{b}{2a}, \frac{4ac-b^2}{4a}\right)$。同时，$x = -\frac{b}{2a}$ 也是 $y = a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac-b^2}{4a}$ 的函数图象的对称轴。对于一个一元二次函数，对称轴 $x = -\frac{b}{2a}$，开口方向和大小 (a)，最值点 $\left(-\frac{b}{2a}, \frac{4ac-b^2}{4a}\right)$ 是几个最重要的信息。

类似地，对于求解方程 $y = ax^2 + bx + c$，或者说 $ax^2 + bx + (c - y) = 0$（这里我们可以通过重新定义 c 使得方程的右侧等于零），由于 $y = ax^2 + bx + c$ 和 $y = x^2$ 可以通过平移和缩放联系起来而平移和缩放是可逆操作，那么一旦我们会求解 $y = x^2$，我们就可以把 $y = x^2$ 的解通过平移和缩放变成 $y = ax^2 + bx + c$ 的解。

具体来说，第一、我们先求出来 $y = z_1 + \frac{4ac-b^2}{4a}$，也就是 $z_1 = y - \frac{4ac-b^2}{4a}$；第二，我们再求出来 $z_1 = az_2$，也就是 $z_2 = \frac{z_1}{a}$ $(a \neq 0)$；第三，我们然后求出来 $z_2 = z_3^2$，也就是 $z_3 = \pm\sqrt{z_2}$；我们接着求出来 $z_3 = x + \frac{b}{2a}$，也就是 $x = z_3 - \frac{b}{2a}$；最后，整合起来，就是一元二次方程 $y = ax^2 + bx + c$ 的解，

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4a\,(c - y)}}{2a}. \tag{9.6}$$

通过重新定义 c $((c - y) \to c)$ 可以变成方程右侧 $y = 0$ 的标准形式，其解为

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{9.7}$$

在实数范围内，这个表达式有意义的条件是 $\Delta = b^2 - 4ac \geqslant 0$。如果这个条件不满足，那么，在实数范围内方程无解。

从函数图象的角度来说，在实数范围内方程 $y_* = ax^2 + bx + c$ 无解意味着函数的最值点和给定的 y_* 值不匹配，也就是，当 $a < 0$ 时，所给定的 y_* 比函数 $y = ax^2 + bx + c$ 的最大值点的 y 坐标 $\frac{4ac-b^2}{4a}$ 还要大；当 $a > 0$ 时，所给定的 y_* 值比函数 $y = ax^2 + bx + c$ 的最小值点的 y 坐标 $\frac{4ac-b^2}{4a}$ 还要小。因此，函数 $y = ax^2 + bx + c$ 和直线 $y = y_*$ 没有交点。回到一元二次方程的标准形式，相当于让这里的 y_* 等于零，或者说重新定义了 c。

插入函数 $y = ax^2 + bx + c$ 和直线 $y = y_*$ 有和没有交点的示意图。

9.8 负一次方函数的定义

表达式、函数图像

9.9 为什么需要函数？

给重复出现的模式、结构一个名字。将来构成更加复杂的数学结构的基础。

9.10 推荐阅读材料

9.11 作业

9.12 本章小结

第十章　二元一次方程和二元一次不等式

10.1　换元法：化二元一次方程为一元一次方程

10.2　消元法：等式的形式和代数式的计算

10.3　二元一次方程的几何含义

10.4　二元一次方程的应用

10.5　二元一次不等式组的求解

10.6　二元一次不等式组的几何含义

10.7　二元一次不等式组的应用

10.8　推荐阅读材料

10.9　作业

10.10　本章小结

第十一章　一元二次方程和一元二次不等式

11.1　一元二次方程的几何含义

11.2　配方法求解一元二次方程

11.3　因式分解法求解一元二次方程

韦达定理

11.4　一元二次方程的应用

11.5　一元二次不等式的几何含义

11.6　一元二次不等式组的求解

11.7　一元二次不等式组的应用

11.8　推荐阅读材料

11.9　作业

11.10　本章小结

第十二章　直角三角形和锐角三角函数

12.1　复习：勾股定理

从面积关系开始的证明，从量纲开始的证明。

另一个体系：从勾股定理到面积关系。

将来构建完整的平面几何知识体系。

12.2　点和距离的定义

从数轴开始定义点（暂时接受点和“实数”的一一对应关系，未来会有机会进一步来学习、质疑和思考它）和距离，平面直角坐标系。

直线的几何和解析几何的定义。

12.3　直角三角形的等比例扩大

初次接触三角形相似性。

三角形相似性的研究依赖于全等三角形和平行线的性质。其它更复杂图形的研究依赖于三角形和圆（也往往依赖于三角形）的研究。

但是，直角三角形的相似性可以单独拿出来先学习。

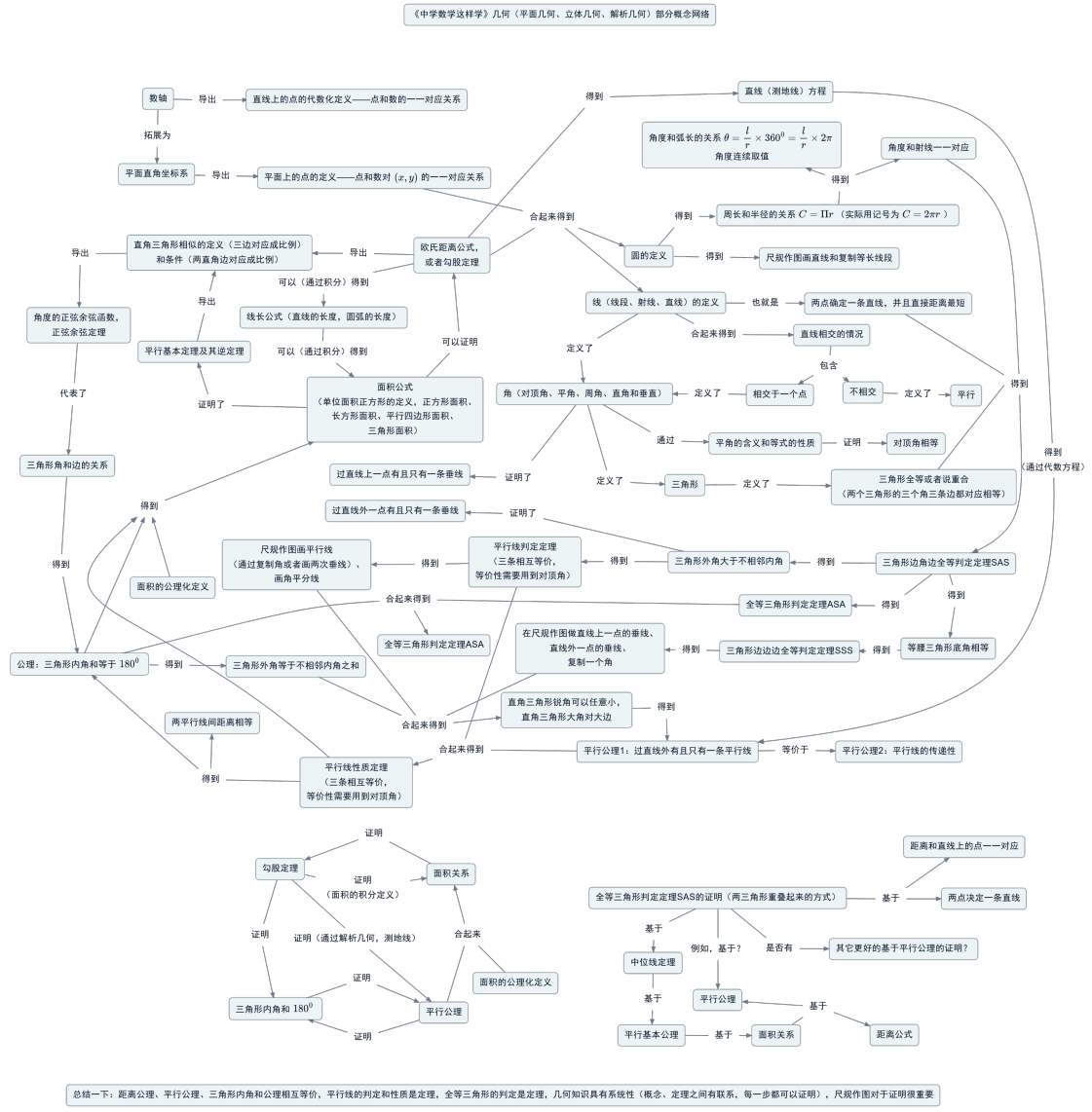

图 12.1: **几何学整体概念地图**。本图展示几何学的三个相互等价和独立的知识起点，以及它们之间的关系。考虑是否印单独一页

12.4 锐角的正弦余弦函数

12.5 三角形中的正弦定理

12.6 三角形中的余弦定理

12.7 三角函数的应用

12.8 圆和角的定义

圆的定义，通过圆弧定义角度，跳过圆周圆弧长度的具体计算到比例。角度和直线的一一对应关系。

12.9 推荐阅读材料

12.10 作业

12.11 本章小结

从数得到了面、点、线、角。

第十三章　平面几何：三角形的全等

13.1　三角形定义

13.2　三角形全等的定义和判定

13.3　为什么首先研究三角形？

13.4　尺规作图

圆规的作用、画直线、复制线段、复制角、平分线段、平分角、做平行线、做垂线

13.5　全等三角形的应用

先用余弦定理证明 SSS，再证明 SAS，ASA，AAS

平行四边形的性质和判定，直角三角形全等的判定

13.6　推荐阅读材料

13.7　作业

13.8　本章小结

第十四章　平面几何：平行

14.1　平行和垂直的定义

14.2　平面几何基本公理

勾股定理、三角形内角和公理、平行公理

14.3　平行线的判定和性质

14.4　推荐阅读材料

14.5　作业

14.6　本章小结

第十五章　平面几何：圆

15.1　圆的定义

15.2　圆中三角形的相关定理

弧、圆周角、圆心角，中垂线，切线
圆心角等于两倍的圆周角，同弧所对的圆周角相等

15.3　共圆定理：把不是圆的问题变成圆的问题

15.4　弧度、弧长和扇形面积

圆的周长和面积

15.5　推荐阅读材料

15.6　作业

15.7　本章小结

第十六章　平面几何：解析方式

The power of mathematics is often to change one thing into another, to change geometry into language.

数学的力量往往在于把一个东西变成另一个东西，例如把几何变成语言。

– Marcus du Sautoy

我们已经把实数和数轴上的点一一对应起来，把平面上的点和二维笛卡尔直角坐标系下的一对对的数联系起来，我们也已经学习过直线可以在二维笛卡尔直角坐标系下用 $y = kx + b$[1]来表示。在这一章中，我们来学习更多的这样的代数表达式和几何图形的对应关系，以及用这个对应关系来解决几何问题，或者偶尔也用这个关系来解决代数问题。

16.1　代数表达式和几何图形

首先，我们把之前已经学会过的代数表达式和几何图形之间的对应关系都复习一边，然后，我们来学习几个新的这样的对应关系。

第一个要复习的是一次函数 $y = kx + b$ 和直线。首先，当 $k = 0$ 的时候，一次函数成了 $y = b$。把二维平面上这样的点 $(x, y = b)$ 都取出来，描成黑体，我们发现，就是平行于 x 轴的直线——把 x 轴向上平移了 b（如果实际上，$b = -|b|$，则向上移动负的，自然就成了向下移动）。接着，按道理，我们应该去关心一下 y 轴的平移。我们发现，y 轴本身是 $(x = 0, y)$ 的所有的点的集合。写成代数表达式就是 $x = 0$。那么，更一般的平行于 y 轴

[1]原则上，为了包含跟 y 轴平行的直线 $x = c$，我们也得允许写成 $x = ly + c$ 的形式，或者更一般地 $kx + ly = c$ 的形式。

的直线，很自然就应该是向右平移 y 轴来得到，也就是 $x=c$。我们也可以通过描点法来把满足 $(x=c,y)$ 这个条件的所有点都找出来看一看。不过，这里有个问题，如何把一般的直线的代数形式 $y=kx+b$ 变成 $x=c$ 的形式呢？如果我们一定要尝试，可以这样，

$$x=\frac{1}{k}y-\frac{b}{k}\overset{?}{\Longrightarrow}x=c \tag{16.1}$$

看起来必须 k 是一个比所有的 y 都要大的数才行，而且 b 呢还跟这个比所有的数都要大的数成比例。将来我们会学习到，这是无穷大。但是，在你们现在的知识范围内这个“无穷大”和“跟无穷大成正比”的概念还解释不清楚。因此，我们只能说，所谓的直线的一般代数表达式 $y=kx+b$ 其实是不能用来表示 $x=c$ 的。为了解决这个问题，我们可以把直线的代数表达式拓展为 $kx+ly=c$。当然，通常情况下，我们也接受“直线的一般代数表达式是 $y=kx+b$”这个不太准确的说法。

下面，我们来看为什么满足 $kx+ly=c$ 的 (x,y) 的集合是一条直线。首先，我们已经看到特例 $y=\frac{c}{l}$（$k=0$）和 $x=\frac{c}{k}$（$l=0$）肯定是直线。现在，我们来处理 $k\neq 0$，$l\neq 0$ 的一般情况。在这个一般情况下，我们回到之前的 $y=kx+b$ 的代数表达式，相当于重新定义了一下这些系数。

习题 16.1（直线的不同代数表达式）. 把代数表达式 $kx+ly=c$ 变形为 $y=kx+b$，找到这样的变换的条件，以及相应的变换，也就是后面代数式的 k,b 和前面的代数式中的 k,l,c 的关系。

我们还发现，$y=kx+b$ 和 $y=kx$ 的差别就是前一条直线是后一条直线向上平移 b。因此，两条直线完全平行。也就是 k 相同的直线平行。b 呢决定了直线和 y 轴的交点，也就是 $y=kx+b$ 永远经过 $(0,b)$ 这个点。因此，k 被称作直线 $y=kx+b$ 的斜率，b 被称作直线 $y=kx+b$ 在 y 轴上的截距。

为此，我们需要回到直线的定义：在直线上，P,Q 两点之间的任何地方取一点 M（于是，P,Q 在边缘，M 在内侧），有距离上的相加关系，也就是（引用直线的定义）

$$|PQ|=|PM|+|MQ|. \tag{16.2}$$

也就是说不可能通过在直线上选择某个点 M 使得加起来的距离 $|PM|+|MQ|$ 大于原来的两点之间的距离 $|PQ|$。也就是说，直线上的点之间的距离

已经是最短的了，只要偏离直线，就可以找到一个点，使得 $|PM|+|MQ|>|PQ|$[2]。

那我们用这个直线的定义来验证一下，是否 $y=kx+b$ 的集合是直线，也就是是否满足 $|PM|+|MQ|=|PQ|$。我们取 $x_P<x_M<x_Q$ 来保证 P,Q 在边缘，M 在内侧。然后，我们来计算这几个距离。这里需要用到距离公式 $|PQ|=\sqrt{(x_Q-x_P)^2+(y_Q-y_P)^2}$。

$$
\begin{aligned}
|PM|+|MQ| &= \sqrt{(x_M-x_P)^2+(y_M-y_P)^2}+\sqrt{(x_Q-x_M)^2+(y_Q-y_M)^2},\\
&= \sqrt{(x_M-x_P)^2+(kx_M-kx_P)^2}+\sqrt{(x_Q-x_M)^2+(kx_Q-kx_M)^2},\\
&= \sqrt{(1+k^2)}\,(x_M-x_P)+\sqrt{(1+k^2)}\,(x_Q-x_M),\\
&= \sqrt{(1+k^2)}\,(x_Q-x_P),\\
&= |PQ|.
\end{aligned}
\tag{16.3}
$$

正好满足直线的要求。

插入 PMQ 直线示意图

这样我们就证明了所有的满足 $kx+ly=c$ 的点的集合是直线。同时，我们来对 $kx+ly=c$ 做一点点简化。当 k,l 其中之一不为零的时候[3]，可以表示为 $\frac{k}{l}x+y=\frac{c}{l}$（相当于 $y=kx+b$）当 $l\neq 0$ 的时候或者 $x+\frac{l}{k}y=\frac{c}{k}$（相当于 $x=ky+b$）当 $k\neq 0$ 的时候。也就是说，$kx+ly=c$ 实际上可以简化为两个参数的数学表达式，尽管看起来有 k,l,c 三个。

[2]原则上，等号不成立还有另一个可能，就是 $|PM|+|MQ|<|PQ|$，不过我们暂时不讨论这个可能。将来学习到非欧几何会遇到这个情形

[3]如果 k,l 都为零，则表达式成了 $0=c$，这就要求 c 也等于零，同时我们发现任意的 (x,y) 都满足 $0x+0y=0$，因此，$0x+0y=0$ 这个表达式对应的点的集合是平面上所有的点。如果 c 不等于零，则没有任何点可以满足这个要求。这样的所有的点都满足或者没有任何点满足表达式的情形，我们不再考虑。因此，我们要求 k,l 其中之一不为零。

有了直线的一般数学表达式，我们就可以来理解“两点确定一条直线”了。也就是，平面上给定任何两个不不同的点，确定了一条并且只有一条直线。

例 16.1. *[两点确定一条直线]* 证明过两点 (x_P, y_P) 和 (x_Q, y_Q) 有且只有一条直线。

我们先尝试用 $y = kx + b$ 来试试通过这两个点，也就是

$$\begin{cases} y_P = kx_P + b, \\ y_Q = kx_Q + b. \end{cases} \tag{16.4}$$

这是二元一次方程组，求解出来 k, b 得到

$$k = \frac{y_Q - y_P}{x_Q - x_P}, b = \frac{y_P x_Q - y_Q x_P}{x_Q - x_P}. \tag{16.5}$$

当 $x_Q - x_P = 0$ 的时候，这个表达式会出问题，分子是有限大小，分母是零。这个时候，实际上，我们就要用 $x = ky + b$ 来表示这个直线了，于是，

$$\begin{cases} x_P = ky_P + b, \\ x_Q = ky_Q + b. \end{cases} \tag{16.6}$$

求解出来 k, b 得到

$$k = \frac{x_Q - x_P}{y_Q - y_P}, b = \frac{x_P y_Q - x_Q y_P}{y_Q - y_P}. \tag{16.7}$$

这个表达式只有当 $y_Q - y_P = 0$ 的时候出问题。但是，$x_Q - x_P = 0$ 和 $y_Q - y_P = 0$ 不可能同时发生，如果同时发生了就意味着 P, Q 实际上是同一个点，而这个和我们要求的两个不同点矛盾。因此，上面求出来的两个直线的表达式，至少总有一个是成立，不可能同时不成立。

下面，我们来证明，如果两者都成立，也就是 $x_Q - x_P \neq 0$ 和 $y_Q - y_P \neq 0$，其实它们是同一个表达式。

$$y = \frac{y_Q - y_P}{x_Q - x_P} x + \frac{y_P x_Q - y_Q x_P}{x_Q - x_P} \tag{16.8}$$

$$\Longrightarrow \frac{x_Q - x_P}{y_Q - y_P} y = x + \frac{x_Q - x_P}{y_Q - y_P} \frac{y_P x_Q - y_Q x_P}{x_Q - x_P} \tag{16.9}$$

$$\Longrightarrow x = \frac{x_Q - x_P}{y_Q - y_P} y + \frac{y_Q x_P - y_P x_Q}{y_Q - y_P}. \tag{16.10}$$

因此，合起来我们证明了，两点确定一条直线。

上面的证明方法是计算形式的，也就是说，我们把直线表达式完全地计算了出来，从给定的两点开始。这个计算过程，未来对于从给定的两点得到直线表达式是非常有用的。此外，通过计算来证明，尤其是通过解方程来证明，往往比通过定理的逻辑演绎操作起来更简单。第一，这说明了计算就是演绎推理；第二，实际上，这也启发了定理的机器证明的研究 [4, 5]。

第二个要复习的是代数表达式 $(x = x_0, y = y_0)$ 和点的关系。通过描点法——在 x 坐标轴上选取对应着 $x = x_0$ 的点，在 y 坐标轴上选取对应着 $y = y_0$ 的点，然后过这两个点分别做垂直于 x 轴而平行于 y 轴的直线和垂直于 y 轴而平行于 x 轴的直线；这两条直线不可能平行于是必然相交于一点，这个点就是对应着代数表达式 $(x = x_0, y = y_0)$ 的点。或者，用另一个思路，把代数表达式 $(x = x_0, y = y_0)$ 看作是同时满足 $x = x_0$ 和 $y = y_0$ 的对象，于是，我们只需要看看这两个条件分别是什么。按照之前学习过的直线，$x = x_0$ 和 $y = y_0$ 正好分别就是两条相互垂直的直线，必然有一个交点，而这个交点刚好同时满足 $x = x_0$ 和 $y = y_0$ 这两个条件。也就是说这个交点刚好就是代数表达式 $(x = x_0, y = y_0)$ 对应的点。

插入描点法找 $(x = x_0, y = y_0)$ 的图

顺便，我们其实也可以先复习点，再复习线的。但是，这里我们由于考虑到找点的时候需要用到线，选择了先复习线，再复习点。

现在我们来学习几个新的代数表达式及其代表的几何图形的关系。首先，我们来学习圆。圆的含义是到某个点 O 距离 $d = r$ 相同的点的集合。不妨我们把这个圆心点 O 就放到坐标原点 $(0, 0)$。按照距离公式，那就是

$$x^2 + y^2 = r^2, \tag{16.11}$$

或者更一般地

$$(x - x_O)^2 + (y - y_O)^2 = r^2, \tag{16.12}$$

对于一般的处于 (x_O, y_O) 的圆心 O。

注意，前面的直线的代数表达式 $y = kx + b$，实际上可以看作一个函数——给定任何的自变量 $x \in \mathbb{R}$ 都可以找到唯一的一个像 $y = kx + b$。但是，这里的 $x^2 + y^2 = r^2$ 就不能看作是函数了：对于一个给定的 $x \in [0, r]$，我们有 $y = \pm\sqrt{r^2 - x^2}$，两个满足要求的 y。这就不满足函数的要求了。当

然，如果我们一定要变成函数，也可以把两者拆开，变成两个函数，也就是 $y=\sqrt{r^2-x^2}$ 和 $y=-\sqrt{r^2-x^2}$。

插入图：上半部分和下半部分两个半圆不同颜色，代表不是函数，分开成两个函数

16.2 通过把几何问题转变成代数表达式来解决几何学问题

我们来用上面的解析几何知识，也就是把几何图形写成代数表达式的知识，来解决一个完全是几何图形的问题。

例 16.2 (三角形中位线定理的解析几何). 三角形中位线平行于底边，并且其长度是底边的一半。

也就是在任意三角形 $\triangle ABC$ 中，D,E 分别是 AB 和 AC 的中点，证明 $DE\parallel BC$，并且长度 $DE=\frac{1}{2}BC$。

有了解析几何，为了证明这两个结论，我们只需要先求出来 D,E 的坐标，再计算出来过 D,E 的直线的代数表达式，然后和直线 BC 的代数表达式去相比较，看是否有相同的斜率。然后，再计算出来 DE 的长度，和 BC 的长度做比较。下面我们来完成这个计算。

我们先来建立一个坐标系：以 B 为坐标原点，BC 方向为 x 轴正方向。于是，我们把 A,B,C 分别表示为，

$$(x_A,y_A),(0,0),(x_C,0). \tag{16.13}$$

用这些信息，我们来计算 D,E 的坐标。我们先来计算直线 AB 的代数表达式，然后在这条直线上，选择一个点 D 使得 $BD=\frac{1}{2}BA$。用前面例 16.1 中的方法，我们得到直线 AB 的代数表达式为

$$y=\frac{y_A}{x_A}x. \tag{16.14}$$

于是 D 的坐标为

$$\left(x_D,\frac{y_A}{x_A}x_D\right). \tag{16.15}$$

下面我们来确定 x_D，其条件是

$$\begin{aligned} BD = \frac{1}{2}BA \Rightarrow \quad & \sqrt{x_D^2 + \left(\frac{y_A}{x_A}x_D\right)^2} = \frac{1}{2}\sqrt{x_A^2 + y_A^2}, \\ \Rightarrow \quad & \frac{x_D}{x_A}\sqrt{x_A^2 + y_A^2} = \frac{1}{2}\sqrt{x_A^2 + y_A^2}, \\ \Rightarrow \quad & x_D = \frac{1}{2}x_A. \end{aligned} \tag{16.16}$$

于是 D 的坐标为

$$\left(\frac{1}{2}x_A, \frac{1}{2}y_A\right). \tag{16.17}$$

同理可得 E 的坐标为，

$$\left(\frac{1}{2}\left(x_A + x_C\right), \frac{1}{2}y_A\right). \tag{16.18}$$

我们再来计算 DE 的代数表达式，会发现，D, E 的 x 坐标不同，但是 y 坐标相同，于是直线为 $y = y_A$，其斜率为零，也就是平行于 x 轴 $y = 0$。我们再来计算 DE 的距离，

$$DE = \sqrt{\left(\frac{1}{2}x_A - \frac{1}{2}\left(x_A + x_C\right)\right)^2 + \left(\frac{1}{2}y_A - \frac{1}{2}y_A\right)^2} = \frac{1}{2}x_C = \frac{1}{2}BC. \tag{16.19}$$

完毕。

前面我们已经看到，不使用解析几何，三角形的中位线定理是一个比较难证明的定理。我们发现，一旦用了解析几何，这个就很简单。

插入图，编一些解析几何证明平面几何的题

16.3 通过把代数表达式转变成几何问题来解决代数问题

我们来用上面的解析几何知识，也就是把代数表达式看作几何图形的知识，来解决一个完全是代数的问题。

例 16.3 (方程的解的几何). 已知如下两个方程无解 $x^2+y^2=4, y=x+b$，求其中的系数 b 的取值。

传统的代数解法如下：

$$\begin{aligned} x^2+(x+b)^2=4 \Rightarrow & \quad 2x^2+2bx+b^2-4=0, \\ \Rightarrow & \quad \Delta=(2b)^2-4\times 2\left(b^2-4\right)<0, \\ \Rightarrow & \quad b\in\left(-\infty,-2\sqrt{2}\right)\cup\left(2\sqrt{2},\infty\right). \end{aligned} \tag{16.20}$$

现在我们来用几何解法。我们注意到 $x^2+y^2=4$ 是一个以原点为圆心半径为 2 的圆，$y=x+b$ 是 45° 的直线。我们化成图。很自然就会发现，随着直线截距的变化，两者的交点会发生变化。两点就是同时满足这两个方程的解。因此，我们只需要找出来直线和圆只有一个交点，也就是直线和圆相切的情形，就能够得到合适的参数 b 的边界值。

于是问题就成了完全一个几何题，给定一个以原点为圆心半径为 2 的圆，求其 45° 方向的切线。插入图 。我们发现，在直角三角形 $\triangle AOB$ 中，OE 是垂足，并且 OE 的长度是 2。按照 45° 的信息，我们发现，$\angle ABO=45^\circ$。于是，$\triangle AOB$ 中是等腰直角三角形，直角边长为 $|b|$。于是，三角形 $\triangle OEB$ 也是一个直角等腰三角形。其直角边为 2。因此，其斜边程度为 $|b|=2\sqrt{2}$。进一步，我们发现，如果我们让 $b<-2\sqrt{2}$，则这里的直线和圆就没有交点了。

同理我们对直角三角形 $\triangle COD$ 做计算，就发现，当 $b>2\sqrt{2}$ 时这里的直线和圆就没有交点了。合起来，满足要求的 b 是，$\left(-\infty,-2\sqrt{2}\right)\cup\left(2\sqrt{2},\infty\right)$。

我们发现两种解法得到的答案一致。你可能会觉得，是不是代数解法更简单呢？可能是的。套用一元二次方程根的判别式就可以。但是，你也会感受到，从几何图形来理解这个问题更清晰：就是通过改变截距来移动一条直线，在这个过程中看直线和圆的交点。未来遇到更加复杂的计算问题，甚至可能到底如何计算不是特别明显的时候，用几何图形来辅助思考，来形成更加深刻的认知，甚至一定程度上的图形认知、直觉认知，对于解决问题也是有意义的。

编一些用几何方法求解代数问题的题

16.4 推荐阅读材料

16.5 作业

16.6 本章小结

在本章中，我们复习和建立了几种几何对象和代数表达式的联系，然后用这些联系来求解了几个例题，包含用代数方法求解几何问题和用几何方法求解代数问题。我们可以把这样的解题方法粗略地称作数形结合。未来，我们会学习到更多几何对象和代数表达式之间的对应关系，于是可以更多地用数形结合来求解更多的更复杂的问题。

但是，其实，这不是本章真正的目的，那仅仅是过程。本章的目的是帮助你进一步理解到：代数式和几何图形本来就是一个东西。在咱们的平面几何体系中，点本来就是通过坐标来定义的，更加复杂的几何对象，也不过就是点的集合，用坐标表达的点的集合。在咱们的平面几何体系中，直线是通过距离来定义，而距离公式被当作了整个几何学知识的基础。你看距离公式，自然就是一个代数表达式。因此，代数表达式和几何图形之间本来就没有区隔。

在这个角度，未来如果距离公式改了，那自然整个几何学的知识就也相当于被改了。如果你未来会学习到非欧几何，其根源就在改变了距离公式。咱们这样的学习，用简单的知识，来体会到高层知识，高视角的对这个学科是什么的理解，可以给未来的学习和创造做更好的铺垫。

第十七章　平面几何：综合

17.1　平行线基本定理

基于正方形面积定义的面积公式、长方形面积公式、平行四边形面积公式、三角形面积公式

平行线基本定理及其逆定理

17.2　三角形的相似

相似三角形的定义和判定

17.3　相似三角形的应用

17.4　数学解题 WHWM 四问和概念地图用于几何证明

17.5　几个平面几何知识体系赏析

17.5.1　Euclid 体系

[6]

17.5.2 Hilbert 体系

[7]

17.5.3 SMSG 体系

[8]

17.5.4 本书的体系

17.6 推荐阅读材料

17.7 作业

17.8 本章小结

第十八章　概率论初步

18.1　古典概型的例子

18.2　事件和事件集合的运算

基元事件、等概率事件、事件的集合运算

18.3　概率：从事件集合到 $[0,1]$ 的映射

18.4　独立事件和互斥事件

18.5　初识条件概率

18.6　概率论的应用

描述和解决概率问题，把非概率的问题转化为概率问题，Monte Carlo

18.7 推荐阅读材料

18.8 作业

18.9 本章小结

第十九章　统计学初步

19.1　描述性统计

19.2　统计推断和人类知识

19.3　从抽样获取整体性质

19.4　条件概率用于统计推断

19.5　统计学的应用

19.6　推荐阅读材料

19.7　作业

19.8　本章小结

第二十章　排列组合

20.1　组合

20.2　排列

20.3　从排列组合到概率

20.4　排列组合的应用

20.5　推荐阅读材料

20.6　作业

20.7　本章小结

第二十一章　逻辑学和数学论证

21.1　命题

21.2　论证

正确的（Sound）非平庸（non-trivial）论证。

21.3　原命题和逆否命题的等价性

把命题成立的对象或者场景当作集合。假设某个全集——称为宇宙——已经给定。

21.4　三段论

三段论的集合论表示。证明过程：每一步，从小前提得到结论，大前提放在括号里面。因为（$\because$）所以（$\therefore$）的记号，得证（　□）记号。

21.5 数学归纳法

21.6 不完全归纳法

条件概率和不完全归纳法

21.7 推荐阅读材料

21.8 作业

21.9 本章小结

逻辑帮助你判断论证的对错，做更严格的论证，但是逻辑不启发论证的思路。后者，靠对知识结构的深刻的认知，尤其是长程联系的建立，靠直觉、想象力、现实的启发。

第二十二章　数学建模

22.1　从数学建模的角度看数学知识的发展

从勾股定理到平方根 $\sqrt{2}$，到无理数

22.2　从把数学知识用于描述现实看真正的数学应用题

22.3　数学建模五步及其应用

22.4　推荐阅读材料

22.5　作业

22.6　本章和本卷小结

数学的发展来源于现实世界的启发和数学研究者对数学知识结构本身的逻辑严密性的追求。其中，在现实启发方面，主要通过双向的数学建模来完成，通过把数学结构用于描述现实来解决现实问题，以及通过把数学结构用来描述现实启发提出新的数学结构。

第二部分

高中篇

第二十三章　函数之幂函数、指数函数、对数函数

23.1　幂函数

23.2　指数函数

23.3　对数函数

23.4　函数和反函数的图形和表达式

23.5　推荐阅读材料

23.6　作业

23.7　本章小结

第二十四章　函数之半圆、半椭圆函数

24.1　半圆函数

24.2　半椭圆函数

24.3　圆

24.4　椭圆

24.5　双曲线

24.6　推荐阅读材料

24.7　作业

24.8　本章小结

第二十五章　函数之三角函数

25.1　余弦函数

25.2　余弦定理

25.3　正弦余弦函数关系

25.4　正切反切函数

25.5　周期函数

25.6　推荐阅读材料

25.7　作业

25.8　本章小结

第二十六章　平面和空间矢量

26.1　直角坐标系和矢量的分量形式

正交归一基矢量

26.2　矢量加法和数乘

26.3　矢量内积

26.4　坐标变换和描述坐标变换的矩阵

26.5　矢量的抽象记号语言

26.6　矢量的应用

26.7　推荐阅读材料

26.8　作业

26.9　本章小结

第二十七章　复数

27.1　负数的开平方

27.2　负数的开三次方

代数基本定理（多项式方程的解的个数和方程的阶数相同）

27.3　复数当作矢量使用

27.4　Euler 公式

复数的模和复角表示，Euler 公式。

27.5　复数的加法和乘法

27.6　复数的应用

27.7　推荐阅读材料

27.8　作业

27.9　本章小结

第二十八章　不等式初步

28.1　几个恒成立不等式

28.2　从恒成立不等式到函数极值

28.3　不等式和极值的应用

28.4　推荐阅读材料

28.5　作业

28.6　本章小结

第二十九章　导数和极值初步

29.1　导数的定义初步

极限和导数，曲线斜率

29.2　简单函数的导数

29.3　函数的导数用于求极值

29.4　导数和极值的应用

29.5　推荐阅读材料

29.6　作业

29.7　本章小结

第三十章　离散变量的概率论和统计学

30.1　从观测数据得到频率再到概率

独立样本，概率极值条件

30.2　从已知离散概率到观测数据的频率

大数定律

30.3 用 Monte Carlo 和 Bootstrap 配合数值计算来沟通频率和概率

30.4 概率论和统计学的应用

30.5 推荐阅读材料

30.6 作业

30.7 本章小结

第三十一章　立体几何

31.1　空间直线的平行和垂直

从三位直角坐标系到平行和垂直

31.2　空间平面的平行和垂直

31.3　推荐阅读材料

31.4　作业

31.5　本章小结

第三十二章　表示论：数学是对操作的研究

32.1　当作对象属性或者过程的数

32.2　数当作加法的表示

32.3　数当作乘法的表示

32.4　数和矩阵当作旋转的表示

32.5　数学是对操作的研究对关系的研究

任何一个操作的集合，再保持某个研究对象的某个特征不变，例如一个皮球还是皮球没有被剪开，的前提下所有的可能的操作的集合。研究这个操作集合的元素之间的关系，从而从已知的具体状态经过操作得到新的状态，或者得到所有新状态的可能，或者研究选择什么样的操作来调控这个对象会满足对这个对象的状态的某个特定要求。这样的操作构成群。

因此，从这个角度可以认为，所有的数学，建立在群论的基础上。不过就是某些群表现出来的结构适合用整数、分数、实数、矩阵等来表示。

32.6 把数学看作是操作的研究的意义

32.7 推荐阅读材料

32.8 作业

32.9 本章小结

第三十三章　数学知识网络和学科大图景

Mathematics is the tool specially suited for dealing with abstract concepts of any kind and there is no limit to its power in this field.

数学是一个尤其适合描述各种抽象概念的工具。在这方面它的能力无限。

– Paul Dirac

33.1　数学一二层知识的系统性

33.2　数学学科大图景，及其和一二层知识的联系

33.3　从数学知识体系到科学知识体系

公理化方案

33.4　高等数学学什么

为什么叫做高等数学。数学知识上的高等，数学思维上的高等，数学知识体系性上的高等。从数学知识的学习，到数学知识的创造性运用和创造。

主要数学分支子学科。

（基本）不再出现数而是出现关系的数学。

33.5 推荐阅读材料

33.6 作业

33.7 本章小结

第三十四章　再论数学建模

34.1　从数学建模的角度看数学知识的发展

从实数到复数，逻辑道路和现实道路

34.2　从把数学知识用于描述现实看真正的数学应用题

34.3　数学建模五步及其应用

34.4　推荐阅读材料

34.5　作业

34.6　本章和本卷小结

数学的发展来源于现实世界的启发和数学研究者对数学知识结构本身的逻辑严密性的追求。其中，在现实启发方面，主要通过双向的数学建模来完成，通过把数学结构用于描述现实来解决现实问题，以及通过把数学结构用来描述现实启发提出新的数学结构。

参考文献

[1] 吴金闪. 小学数学这样学 [M]. 浙江人民出版社, 2022.

[2] 吴金闪. 教的更少, 学得更多 [M]. 科学出版社, 2021.

[3] 吴金闪. 《系统科学导引》第一卷 [M]. 科学出版社, 2018.

[4] 吴文俊. 几何定理机器证明的基本原理 [M]. 科学出版社, 1984.

[5] 吴文俊. 数学机械化 [M]. 科学出版社, 2003.

[6] Euclid, edited by Heath T L. The Thirteen Books of Euclid's Elements, Books 1 and 2 （《几何原本》） [M]. New York, NY, USA: Dover Publications, Inc., 1956.

[7] Hilbert D. The Foundations of Geometry（《希尔伯特几何基础》） [M]. CreateSpace Independent Publishing Platform, 2013.

[8] Moise E E. Elementary Geometry from an Advanced Standpoint [M]. ADDISON-WESLEY PUBLISHING COMPANY, 1990.

名词索引

M

M

理解型学习 Meaningful Learning，人类知识高速公路上以高层知识生成器为目标的理解型学习. 23

人名索引

W

W

吴金闪 北京师范大学系统科学学院教师. 24, 26

插图

1 小学数学整体概念地图 46

12.1 几何学整体概念地图 108

举例目录

0.1 例 (乘法对加法的右侧分配律) 33

9.1 例 (从 $y = ax^2 + bx + c$ 到 $y = x^2$) 100

16.1 例 120

16.2 例 (三角形中位线定理的解析几何) 122
16.3 例 (方程的解的几何) 124

作业目录

6.1 习题 (BMI 的定义) 63
6.2 习题 (理发师悖论) 63
6.3 习题 (集合包含关系和并集一) 66
6.4 习题 (集合包含关系和并集二) 66

7.1 习题 (全部四个映射之间的复合关系) 75
7.2 习题 (Dirac 符号下算赋的验证 1) 76
7.3 习题 (Dirac 符号下算赋的验证 2) 76
7.4 习题 (数独游戏) 79

8.1 习题 (否命题和逆命题) 84

9.1 习题 ($f(x) = x + 1$ 是满射) 93
9.2 习题 ($f(x) = x + 1$ 是单射) 93
9.3 习题 ($f(x) = x + 1$ 是满射) 93
9.4 习题 ($f(x) = kx + b$ 是满射) 98
9.5 习题 ($f(x) = kx + b$ 是单射) 98

16.1 习题 (直线的不同代数表达式) 118