

★ 数学浩瀚如大海

我的数学分析积木

(数学分析学习的思考与经验)



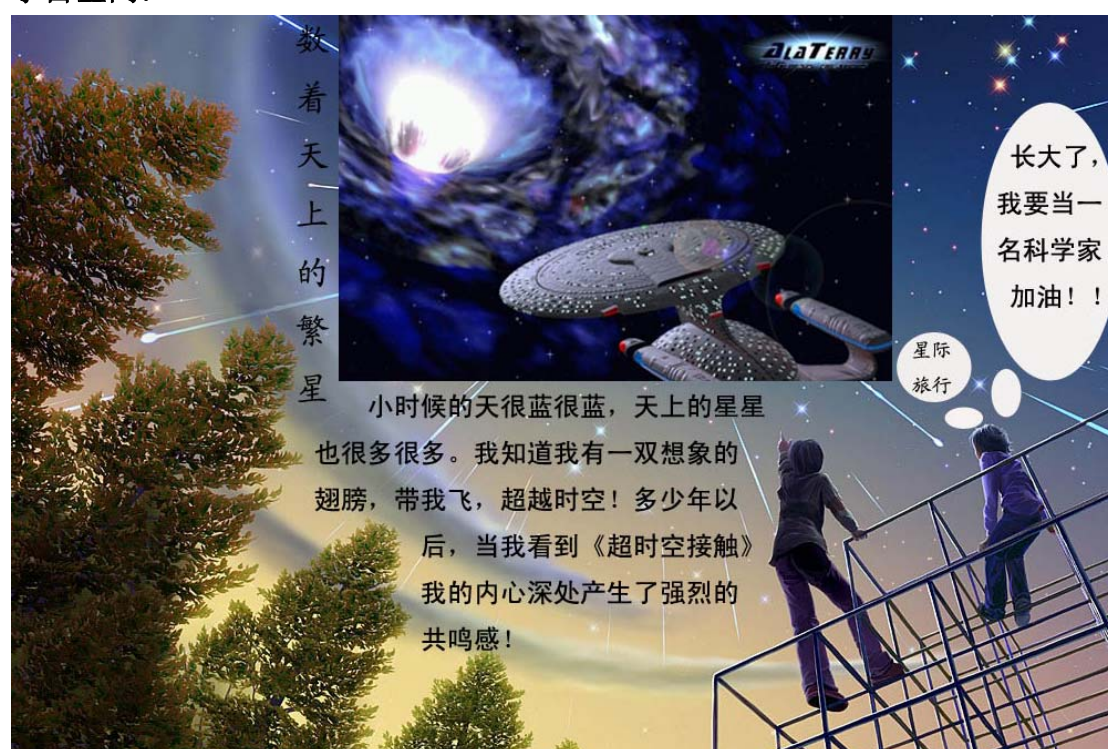
数学探索者号

SCIbird

序 言

本书取名《我的数学分析积木》，原因是本书带有一些微积分回忆录味道，是笔者这些年来对数学的一些感悟所写成的数学笔记。书中观点和内容吸取了不少数学大师的思想精华，同时也包含了自己对数学的一些理解。

为什么写数学书是一个很难回答的问题，肯定不是为了钱，也不是为名，否则就实名制了。SCIbird 这个网络 ID 是“科学鸟”的含义，笔者始终将数学视作科学的组成部分。每个人小时候都有一个梦想，笔者那时候的梦想是探索宇宙空间：



少年时代的梦想

康德有句名言：“有两种东西，我们越是经常、越是执著地思考它们，心中越是充满永远新鲜、有增无减的赞叹和敬畏——我们头上的灿烂星空和心中的道德法则。”

笔者也常常思索为什么“星空的奥秘”那么引人入胜，激励着无数先贤去探索？伽利略、牛顿、欧拉、高斯、庞加莱、爱因斯坦等等，一个个震铄古今的名字铭刻在天文史上。为什么自己也被星空所吸引。

后来笔者领悟到，正是少年时代的好奇心与探索精神驱使着自己去探索。在少年时代，那个没有互联网的时代，笔者已经了解了黑洞、虫洞、相对论等概念，虽然只是不严格的科普知识，但笔者知道，这些概念涉及宇宙最神秘的奥秘。

可惜随着时间的飞逝，好奇心和探索精神磨灭了。小时候那种想一个问题能用一个多月的精神劲没有了。读初中和高中时，自己喜欢上物理。念大学时，喜欢上数学，其程度甚至超过物理。在我看来数学比较自在，想象空间大，当你真正理解数学时，你会发现其数学的奥妙并不亚于星空。特别地，运用数学方法也能认识星空法则。

读研究生时，笔者意识到写不出深刻的文章，更不可能延续科学事业，那时就有写一本数学方法论的想法，算是对这些年自学数学做一个交代。写作上，希望效仿《数学它的内容、方法和意义》这样的科普书，有一定的思维层次和高度，但又不能太高深。想想前苏联那么多大牛院士都肯花时间写出《数学它的内容、方法和意义》这样的好书，但如今科学界的人士太冷漠，对科学普及不热衷，数学界更甚。其结果如何？用卡尔·萨根的话可以概括这样做的后果：“科学激发了人们探求神秘的好奇心，但伪科学也有同样的作用。很少的和落后的科普所放弃的发展空间，很快就会被伪科学和迷信所占领。”如今微博上谣言满天飞，有些谣言稍微有点逻辑和科学知识便能识破，可是……只能说很多人大学白念了。

经常看到微博上某大V就某道数学题发表感慨，说难得多么BT，笔者一看无语了。最近英语要退出高考，变为类四六级模式，引发很多人再次吐槽数学滚出高考。笔者难以接受竟然有那么多人会这么想，无法想象一个不懂数学的民族会发展成什么样（有些人至今仍然以为原命题和逆命题是等价的）。很难想象有比数学更好的形式逻辑学习方式。

联想到卡尔·萨根的话，笔者觉得有必要尽一份力量普及一些数学思想和方法。记得小时候《十万个为什么》风靡一时，同学间互借着看，然后大家在一起讨论，那种风气，远胜当下热议韩剧和美剧，品味也更高。可惜如今风光已不再。当年看《十万个为什么》科普书学会了一种思维方法，称之为“十万个为什么式思维”：

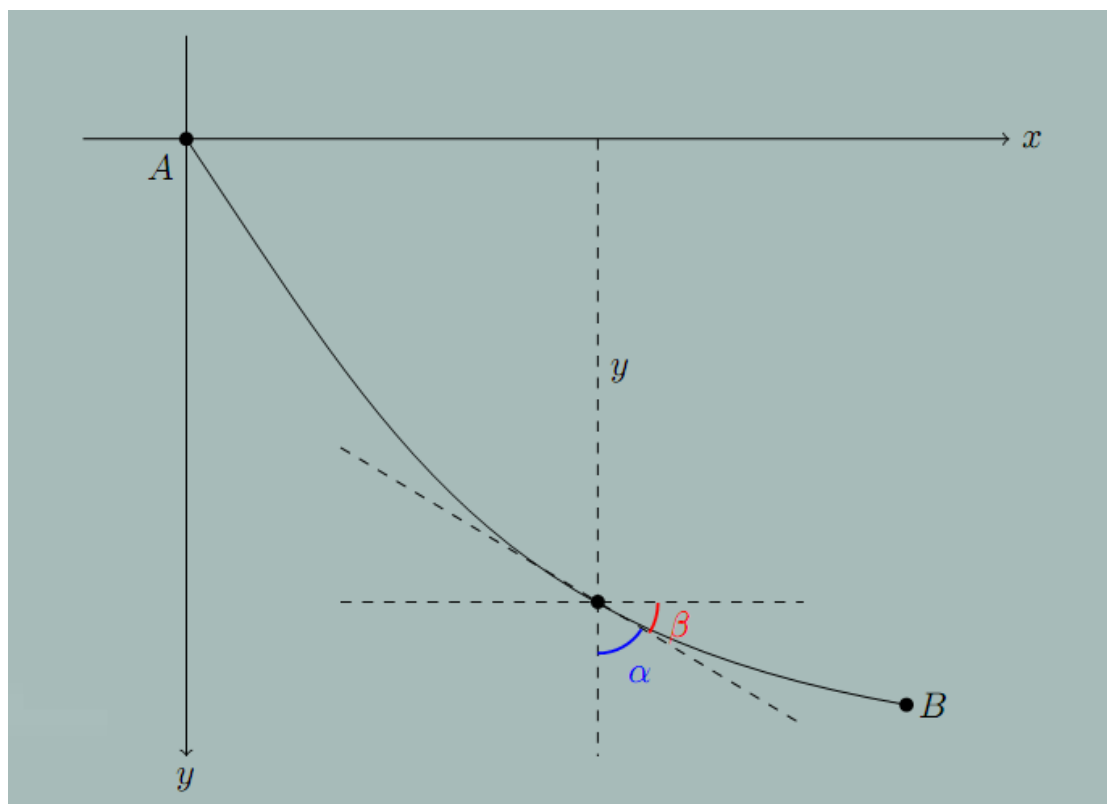
1、是什么？2、为什么？3、怎么办？

就是这平平淡淡的三点，其作用好比思维灯塔，指导了自己这些年的学习，不论是自然科学、数学，经济社会领域问题，受益匪浅。

市面上的数学科普书自己基本上都看过，根据经验，过年低端科普书不少，中高端科普书太少。笔者经历表明存在这样的问题：对于有一定水平的本科生或者不搞数学研究的人，他们没时间和能力去看大部头数学专业系统教材，但看太水的科普书也无聊（这类书的内容很熟悉，没有新意）。有没有难度与《数学它的内容、方法和意义》差不多，或者略难一些的书（但不是教材）。这些人喜欢看一些经典问题和解法，难度不宜太高，要有新思路。涉及问题要尽量高等一些，但也不能太高，举一个例子：**最速降线问题**

一个质点在重力作用下，从一个给定点 A 到不在它垂直下方的另一点 B ，如果不计摩擦力，问沿着什么曲线滑下所需时间最短。

笔者高一时在《形形色色的曲线》里看到这个问题，但书中只给出答案是摆线（旋轮线），没有证明。后来到大一才第一次看到证明。当时第一反应是直线，后来感觉不对，猜想应该是曲线。直观上，开始速度慢，曲线应该比较陡峭；后来速度快了，曲线逐渐平缓。



最速降线问题

这个问题是变分法的先驱性问题，在数学史上有很高地位，试题本身也很有趣。标准解法是解欧拉-拉格朗日方程，但是有一种数学与物理相混合的解法更具有启发性和艺术性，它利用了费马的“光行最速原理”，即光走一条用时间最短的路线。严格从物理角度说，更准确的表述是光走的路线使得光程取驻值。

$$\delta \int_A^B n ds = 0$$

注：光程是一个折合量，可理解为在相同时间内光线在真空中传播的距离。在传播时间相同或相位改变相同的条件下，把光在介质中传播的路程折合为光在真空中传播的相应路程。在数值上，光程等于介质折射率乘以光在介质中传播的路程。

费马的“光行最速原理”导致了物理学中的“最小作用量原理”，由此发展出一种不同于牛顿力学的物理方法。关于变分方法在物理学的应用可参考大学《理论力学》教材或者《分析力学》教材。

回到原题。如上图建立平面直角坐标系，设质点的质量为 m ，由动能定理可知

$$\frac{1}{2}mv^2 = mgy$$

求得质点速度

$$v = \sqrt{2gy}$$

又速度满足

$$v = \frac{ds}{dt} \Rightarrow dt = \frac{ds}{v}$$

所以质点所需时间为

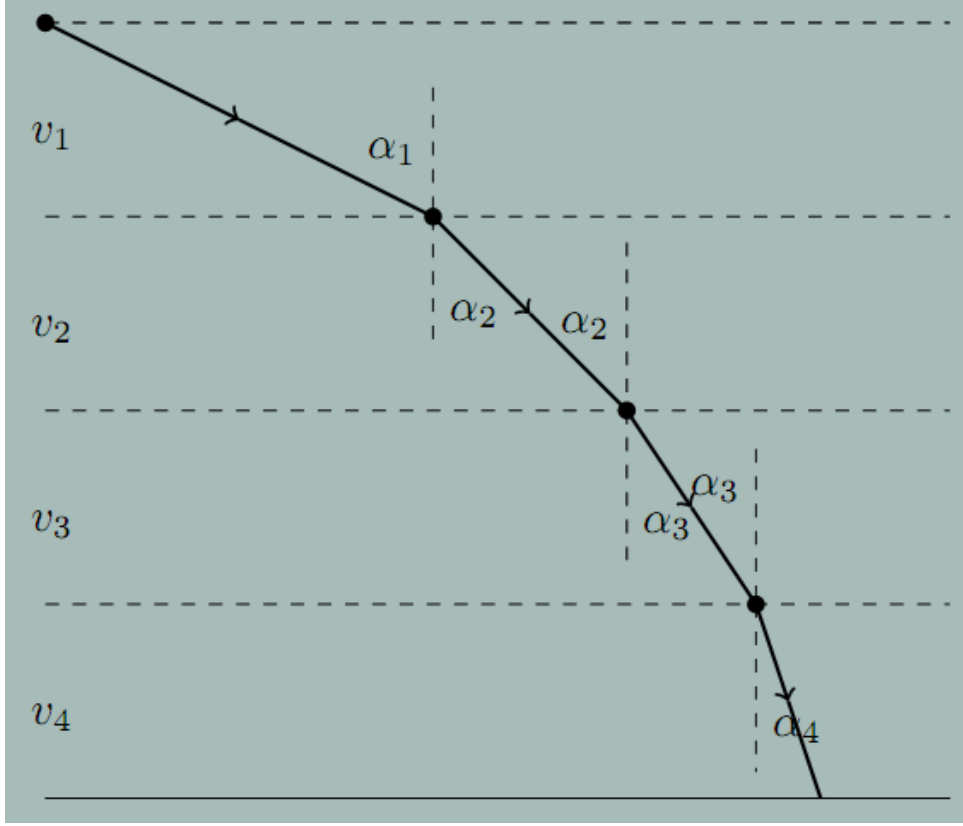
$$T = \int \frac{ds}{v} = \frac{1}{\sqrt{2g}} \int_{x_A}^{x_B} \frac{\sqrt{1+y'(x)^2}}{\sqrt{y}} dx$$

于是原问题等价于寻找一条曲线 $y = y(x)$ 使得时间 $\min T$ 取最小值。

数学系的思维是先考虑存在性，然后证明之，最后求解问题。很多时候，存在性问题非常难证。物理思维则另辟蹊径，在其看来，如果一种物理现象或者物理过程存在，这就说明有对应解。很多时候都是默认有解，然后把注意力集中在解应该具有哪些性质上来，很少把心思放到解的存在性上。工程思维则更激进一些，方法好用就是合理，没工夫在理论方面纠结。

以物理思维看，解决最速降线问题要运用类比思维，构建一个相似的物理模型。

考虑一束光在变介质中的运动情况，先将介质分为 n 层，每一层的折射率都为常数。根据光的折射定律，此时光线为折线：



光线折射示意图

记第 i 层介质内的光速为 v_i ，根据折射定律有

$$\frac{\sin \alpha_1}{v_1} = \frac{\sin \alpha_2}{v_2} = \frac{\sin \alpha_3}{v_3} = \dots$$

将层数无限加细同时利用速度的连续性，可知

$$\frac{\sin \alpha}{v} = \text{const}$$

再由 $v = \sqrt{2gy}$ 以及几何关系

$$\sin \alpha = \cos \beta = \frac{1}{\sqrt{1 + \tan^2 \beta}} = \frac{1}{\sqrt{1 + (y')^2}}$$

可知

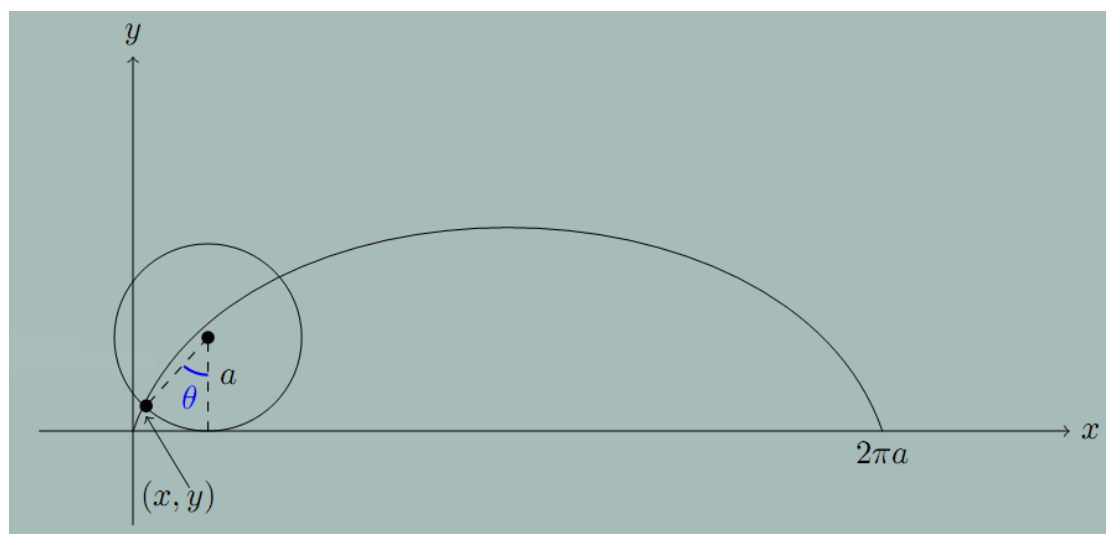
$$\frac{1}{\sqrt{2gy} \sqrt{1 + (y')^2}} = \text{const}$$

由此可知

$$y[1 + (y')^2] = 2a$$

其中， a 为待定常数。

上述微分方程的解为： $x = a(\theta - \sin \theta)$, $y = a(1 - \cos \theta)$ ，这是摆线（旋轮线）的参数方程。



摆线（旋轮线）

一个圆沿一直线纯滚动，则圆上一固定点所经过的轨迹称为摆线。摆线有一个神奇的性质，所谓等时性。将摆线看做是一个滑梯，最低点记为 C 点，则不论初始位置如何，质点沿着光滑摆线下滑到最低点的时间都相同。可以算出，质点下滑到最低点所用时间为

$$T = \pi \sqrt{\frac{a}{g}}$$

注意，这种等时行是精确的，没有物理中单摆的 5 度角以内近似条件。

从纯数学角度讲，也可以用微积分方法求解，一个典型的方法是“参数法”，非常有启发性。先将问题做一般化处理，考虑积分

$$J[y(x)] = \int_{x_1}^{x_2} F(x, y(x), y'(x)) dx$$

这里曲线 $y(x)$ 是连续可微的，且两端点固定。

所有这样的函数 $y(x)$ 所成集合 M 称为容许函数。我们的目标是找到函数 $y(x)$ ，使得 $\forall \bar{y}(x) \in M, J[\bar{y}(x)] \geq J[y(x)]$ 。这类问题如今归到泛函极值问题中，任何一本变分法教材都有详细介绍，下面主要介绍一下参数法德巧妙之处。引入参变量 α ，满足

$$\bar{y}(x) = y(x) + \alpha \eta(x)$$

其中, $\eta(x)$ 连续可微且满足 $\eta(x_1) = \eta(x_2) = 0$.

如此构造出一个函数

$$\varphi(\alpha) = \int_{x_1}^{x_2} F(x, y + \alpha\eta, y' + \alpha\eta') dx$$

依题意, 有 $\varphi(\alpha) \geq \varphi(0)$, 由一元函数极值理论知 $\varphi'(0) = 0$, 即

$$\varphi'(0) = \int_{x_1}^{x_2} [F'_y(x, y, y')\eta + F'_{y'}(x, y, y')\eta'] dx = 0$$

利用分部积分及 $\eta(x_1) = \eta(x_2) = 0$, 可知

$$\int_{x_1}^{x_2} F'_{y'}\eta' dx = - \int_{x_1}^{x_2} \eta \frac{d}{dx} F'_{y'} dx$$

于是

$$\int_{x_1}^{x_2} [F'_y(x, y, y') - \frac{d}{dx} F'_{y'}(x, y, y')]\eta(x) dx = 0$$

根据变分法基本引理, 设 $f(x)$ 连续可微, 若任取连续可微函数 $\eta(x)$, 满足 $\eta(x_1) = \eta(x_2) = 0$ 且

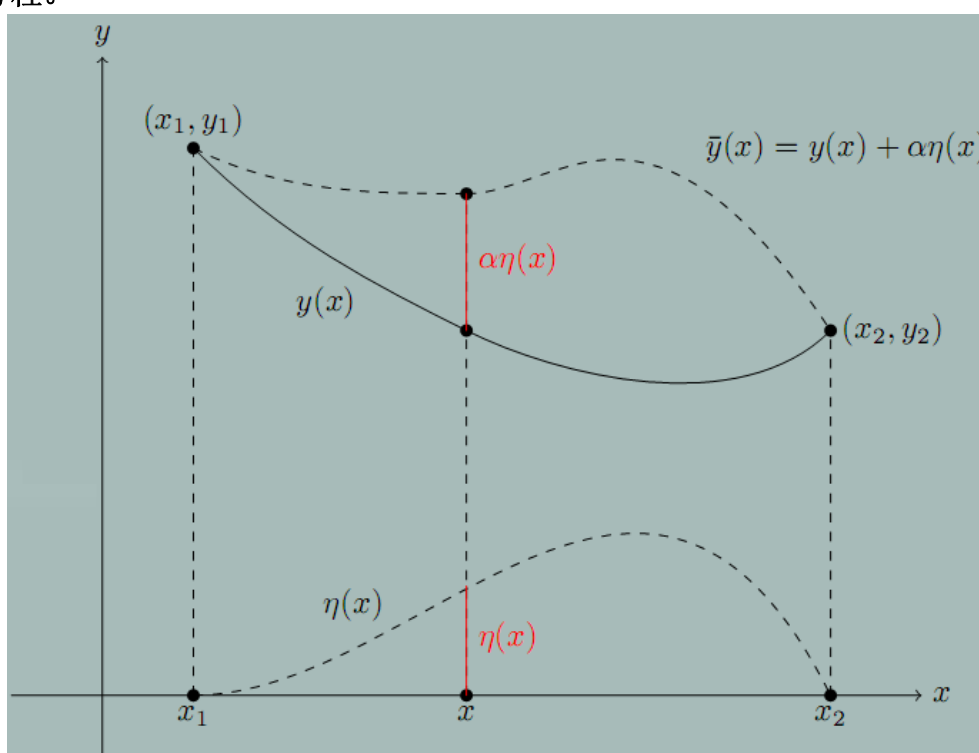
$$\int_{x_1}^{x_2} f(x)\eta(x) dx = 0$$

则 $f(x) \equiv 0$.

因此, 得到欧拉-拉格朗日方程

$$F'_y(x, y, y') - \frac{d}{dx} F'_{y'}(x, y, y') = 0$$

对于最速降线问题, 取 $F(x, y, y') = \sqrt{1 + (y')^2} / \sqrt{2gy}$, 剩下的工作是解常微分方程。



站在物理角度看上面的参数化方法是比较自然的，假设最速降线问题的解是曲线 $y(x)$ ，我们在 $y(x)$ 基础上叠加一个扰动 $\alpha\eta(x)$ ，看看积分值有什么变化。此处参数 α 相当重要，以此为桥梁将一个非常陌生的泛函极值问题转化为熟知的一元函数极值问题，这种转换思维非常实用。

类似的方法还有

$$\varphi(\alpha) = \int_0^{+\infty} e^{-\alpha x} \frac{\sin x}{x} dx$$

以及多元函数泰勒展开技巧

$$\varphi(\alpha) = f(x + \alpha h, y + \alpha k)$$

等等，此间巧妙大家需自己体会。

本书写作风格上尽量遵循上述风格，尽可能揭示一下数学分析中那些典型方法的奥妙之处。正如前面所言，学习时要多问：1、是什么？2、为什么？3、怎么办？由于教材里大部分笔墨都集中在 1 和 3 上（定理+证明），而对 2 言之甚少，本书尽量弥补一下这方面的缺憾，但限于笔者水平，必不完美。

在写作风格上，笔者认为本书继承了张筑生老师《数学分析新讲》（笔者的第一本数学分析教材）风格与精神。本书多次提到新讲关键字，也希望仍同笔者这本小书的人能去读读新讲。当然，也有一些人读过新讲后觉得没有 SCIbird 说得那么好。这也是自然的，原因是期望太高或者新讲不适合你，你应该选择一本适合自己的数分教材并坚持看完。说来奇怪，看过太多的书之后，笔者反倒感觉是书选人，而不是人选书。

就选材来说，笔者本书大部分内容都是常见内容，分散在各类教材中。但时下天朝数分书多如牛毛，更不乏《数学分析习题课讲义》这样的精品，如何体现出价值，如何体现出特色。笔者在动笔前也斟酌了一下，首先定位在数学分析上，这方面最熟悉，运用起来比较自如。其次，在方法的组合上多用心，使得读者读完之后能有“原来这种方法可以这么用”之感。最后是考虑差异性，提供难度层次各不相同的材料，避免产生这种感觉，微积分只能算算极值、面积、体积之类的。

为了提高本书的思维层次性和吸引力，添加了一些著名定理的高级科普文章，如 Atiyah-Singer 指标定理、费马大定理的现代证明、Milnor 的 7 维怪球等等，用现在的网络流行语是高端大气上档次。但其本意是力图说明，尽管数学

方法和技术上有高低难易之分，但在数学思想上是相通的，没有本质差别。

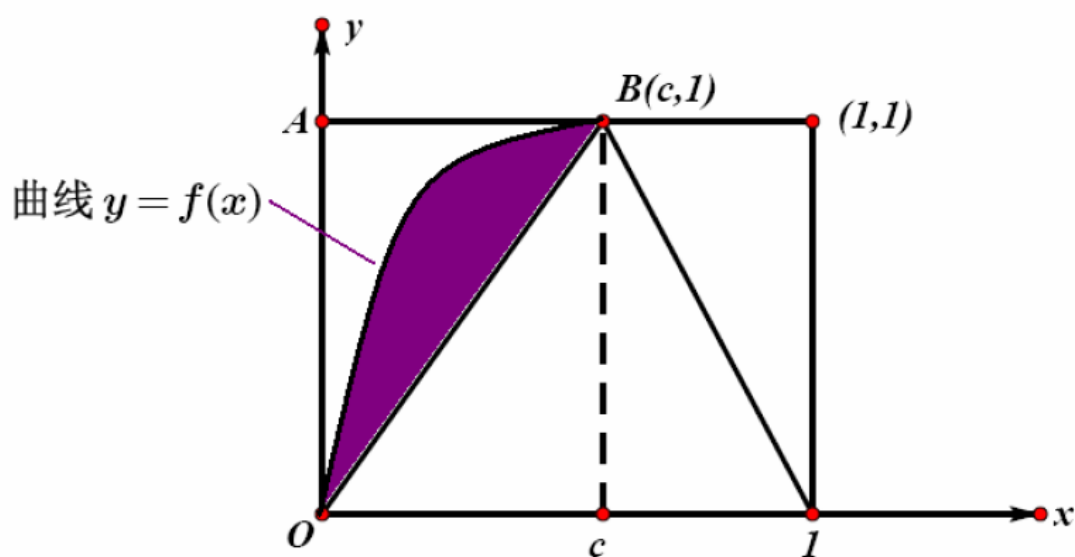
在写作过程中笔者深感数学方法论方面的书不好写，很多书写成了谈数学哲学，干货不多。笔者认为，一个人在 40 岁之前阅历不足，还是少谈哲学（但不是不了解），多务实一些为好，这方面笔者有过教训，也深刻反思过。所以在构思本书内容时，笔者坚持技术含量和水平的底线，不能写成泛泛而谈的股评式兑水书。同时，笔者也极力避免写成一本难题集。为此，笔者加入了很多描述性文字，这些文字是笔者的数学思考心得，这是本书最大的亮点。

如果一种理论或方法不容易说明怎么办？笔者的经验是多用例子，这和数学中多做题是一个道理。很多人相当反对题海战术，不屑做题，这也不对。适当做题还是很有帮助的，就像士兵一样，几十年没打仗了，难道平时就不用射击训练了吗？很多人做的题很多，但效果不佳，是因为没有深层次的思考。关于如何思考，本书正文里给出了不少建议，这里就不赘述了。

在形式上，本书也力图多元化、多角度、多层次地看待问题。特别强调几何直观的重要性。以 2014 年北大数分第 10 题为例：

10. 设 $f: [0, 1] \rightarrow [0, 1]$ 是 C^2 函数, $f(0) = f(1) = 0$, 且 $f''(x) < 0, \forall x \in [0, 1]$. 记曲线 $\{(x, f(x)) | x \in [0, 1]\}$ 的弧长是 L . 证明 $L < 3$.

由题中条件不难猜出曲线 $y = f(x)$ 的图形类似于一个开口向下的抛物线，曲线向上凸。



如上图所示，容易发现问题中的 3 应该是单位正方形的三边之和（除了 x 轴上那一边）。于是自然猜想：

曲线 $y = f(x)$ 在区间 $[0, c]$ 上的弧段长度不大于 $OA + AB$.

不难算出: $OA + AB = f(c) + c = 1 + c$. 其中 $c \in (0, 1), f(c) = \max f(x) = 1$.

另一半的情况是“本质对称”的, 证明类似. 于是, 关键的猜想是

$$\ell = \int_0^c \sqrt{1 + f'(x)^2} dx \leq XXX = c + f(c)$$

这就像一个“黑箱子”, 知道两端情况, 猜中间 XXX 是什么. 经过简单的试算可知

$$\int_0^c (1 + f'(x)) dx = c + f(c) - f(0) = c + f(c)$$

剩下的就是证明不等式

$$\sqrt{a^2 + b^2} \leq a + b, \quad a, b > 0$$

两边平方即可. 再取 $a = 1, b = f'(x)$, 就得到了

$$\ell = \int_0^c \sqrt{1 + f'(x)^2} dx \leq \int_0^c (1 + f'(x)) dx = c + f(c) = 1 + c = OA + AB$$

当然, 具体可说明下, 不等式中的等号取不到.

不仅仅是图形, 带箭头的交换图表, 数字表格等等都可以运用之, 现在的教材特别是分析教材形式太单调, 这不利于启发性. 想成为高手就得像特种兵一样, 只精通一种方法是不够的.

本书是笔者在以 *SCIBird* 这个网络 ID 在博士数学论坛所发表的一些列文章的基础之上, 补充修订整理而成. 修订过程中, 删除了个别例子, 补充了更恰当的例子, 改正了一些表述不准确和笔误的地方. 但限于水平、时间和精力, 不可能像那些负责人的老师一样修改了十几遍.

粗略估计, 本书的数学分析例子有 100 多个, 不可能每个都一一细说. 所以写作上采取有的放矢, 部分例子细说, 部分例子留给读者去思考. 书中大部分例子是关于数学分析的, 原因是写作前的定位是: 写给那些读过数学分析教材, 但想重温提高自身分析水平的读者 (也包括考研的). 因此本书各章在解释笔者的学习经验时, 列举的都是分析例子, 其实别的数学分析以往的例子也是可以拿来主义的.

本书是自己多年学习数学的经验总结, 全书正文 13 章, 框架是效仿《孙子兵法》13 篇, 笔者称各章提到的思维方法为“思维模块”(即我的数学分析积木), 这些思维模块是自己这些年从实践摸索中提炼和总结出来的, 同时也借鉴了前人的数学思想 (很多想法大家不谋而合). 书中介绍的思维模块不仅仅适用于数

学分析，相信同样适用于几何与代数领域。

根据笔者的亲身经验，本书篇幅设定在 200 页左右，如此打印出来厚度比较适中，携带也方便。书中包含了一些插图，其来源是梅加强和徐森林两位老师的《数学分析》教材（PDF 文件截图），也包括一些在 Google 上搜到的英文文件截图，剩下的就是笔者自己用几何画板软件和画图板画的示意图。这里对原图作者的贡献表示感谢，特别是梅加强和徐森林两位老师。这些图主要是为了表达数学思想，重点在于意会，而非精确（达不到海报那样的精致）。写作过程中深感现代的电脑技术还有待发展，自己用手能轻松画出来的图，用电脑制图却相当费劲。无奈，只能从别人的电子文档中，寻找一些图片来表述自己的想法，不得已而为之。

最后，关于本书如何使用需要说明一下。本书不是自给自足的书，需要数学分析教材配套使用。这里推荐教材（网上有 PDF 电子版）

1. 《数学分析新讲》 张筑生 编著
2. 《数学分析》 卓里奇 编著

别的数学分析教材也可以，但笔者写作本书时主要以上面两本教材为蓝本，例子也多出于这两本书。本书是写给已经学过数学分析的读者（含自学的），不懂数学分析的还是先找本数学分析书踏踏实实学完之后再看书。学数学需要经过

简单 → 复杂 → 简单

梅花香自苦寒来，不经过复杂怎能领悟到高层次的简单。高手们经常说“大道至简、返璞归真”之类的话，那都是先经过复杂的探索过程的，是众里寻她千百度，蓦然回首那人却在灯火阑珊处的感悟。

如果某位读者机缘巧合之下看到本书，且非常认同本书内容，那么强烈建议多动手推导一下本书中那些基础例子。数学是干出来的，不是看出来的，有缘加用心，你才能真正获得收获！

本书版权由 SCIBird 所有，严禁用于商业用途。

SCIBird

scibird@sina.cn

2014 年 6 月 1 日

本书献给我亲爱的爸爸妈妈

目 录

1 数学的联系与统一性.....	1
2 数学分析三大基本思想之逼近.....	10
3 数学分析三大基本思想之变换.....	20
4 数学分析三大基本思想之分解.....	31
5 抓住问题的核心.....	41
6 类比是一种重要的数学思维.....	49
7 猜想与形式推导.....	59
8 硬算是门基本功.....	69
9 发掘几何直观.....	82
10 要有逆向思维.....	93
11 数学中的构造方法.....	104
12 多读一些著名定理.....	116
13 重视反例.....	129

附录:

1 微积分基本定理新讲.....	136
2 多项式逼近连续函数的折线逼近证明.....	143
3 布劳威尔不动点定理.....	147
4 处处连续处处不可导函数.....	151
5 误差与正态分布.....	158
6 重积分变量代换公式.....	165
7 素数定理的简单解析证明笔记.....	175

第 1 章 数学的联系与统一性

本章是全书第 1 章，具有纲领作用，因此笔者将重点放在整体思维框架上，不做技术性细节展开。细节内容将在后续 12 章中一一介绍。

本书是自己多年学习数学的经验总结，书中介绍的思维模块不仅仅适用于数学分析，相信同样适用于几何与代数领域。在数学领域中，具体内容可能不同，但思想大多是一致的。当然，既然是笔者的学习经验，肯定主观色彩较多，而且世上没有万能的方法，任何一种方法都有其局限性和适用范围。所以对本书提及的观点和方法也要辩证的看，取其精华。

首先要端正态度，对于一个定理或一个问题，我们不应该用做考试题的应试态度来对待，而应该用研究数学问题的态度来对待，尽量挖掘出新的东西，而不局限于问题中的结论本身。

笔者看到一个数学定理或问题，会下意识地提出三个问题：

1. 定理证明本身是构造性的，还是存在性的，证明关键地方是什么？
2. 定理的应用是什么，有没有几何直观或熟悉的例子可解释或验证？
3. 定理能否推广（一般化）？或者一般情况用反例否定掉。

比如连续函数介值定理，如果使用闭区间套方法证明，则属于构造性证明；如果使用上确界方法证明，则属于存在性证明。构造性方法更适用于应用问题，现实问题更重视可操作性。存在性证明通常可以避开某些技术性困难，而且结论往往更有普遍性。难说哪个方法更好，两者各有特色，需要依具体问题而定。比如，你需要解决某个实际应用问题，肯定希望用构造性方法。

把握证明的关键或本质非常重要，可以加深对数学问题的理解。比如

$$\frac{1}{2} + \cos x + \cos 2x + \cdots + \cos nx = \frac{\sin(n + \frac{1}{2})x}{2 \sin \frac{1}{2}x}$$

这个等式称为狄利克雷核（函数），在研究 Fourier 级数收敛性时要用到，证明也不难。粗看起来三角函数能求和还是很吃惊的，但如果考虑欧拉公式

$$e^{inx} = \cos nx + i \sin nx$$

再令 $q = e^{ix}$ ，由此发现上述三角函数求和本质是等比数列求和，于是豁然开朗。

有些人学习数学，只重视抽象的结论，不太重视例子，更别提应用了，这不是一个好习惯。有一个故事是关于代数几何大牛 Deligne 的，据说常常当人家兴致勃勃写了一黑板的高深发现时，Deligne 不慌不忙地站起来说：“讲得很精彩，不过您的结论是错的！”弄了几次大家不禁觉得 Deligne 实在是神仙下凡，毕竟他再怎么强也不能刚刚接触人家的理论半个小时就比人家钻研了好几年还要更明白啊。Deligne 后来透露了自己的秘密，他在听人家讲座时脑子里面准备好几个例子，看到定理推论等等都先用例子验证一番，有时候还真能发现问题。

Deligne 这个案例说明了多考虑例子的重要性，同时也能培养自己的数学直觉。就数学分析来说，导数可以看做的函数切线的斜率或者速度，积分可以看做面积。牛顿-莱布尼茨公式可以看做是距离-速度关系式等等。

$$S(t_2) - S(t_1) \approx \sum_i v(t_i) \Delta t_i \rightarrow \int_{t_1}^{t_2} v(t) dt$$

不难理解 $S'(t) = v(t)$ 关系，即

$$v(t) := \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} = S'(t)$$

有人觉得这些例子太俗了，层次太低。其实，了解一点数学史就知道，微积分的起源就是这些例子。大牛也是从菜鸟慢慢成长起来的。

几何直观，也是笔者注重推荐的。比如理解函数乘积的微分公式

$$d(uv) = u dv + v du$$

可以抽象出下面的面积关系，一目了然：

dv	$u dv$	$du dv$
v	uv	$v du$
	u	du

据科学调查，人们对几何图形的印象，要强于对公式和文字的印象。对于分析爱好者来说，几何直观也可以活跃思维，避免陷入 $\varepsilon - \delta$ 的泥潭当中。

上面的 1 和 2 属于数学中的联系性，而 3 属于统一性。以 Riemann-Lebesgue

引理为例:

$$\lim_{\lambda \rightarrow \infty} \int_a^b f(x) \sin \lambda x \, dx = \lim_{\lambda \rightarrow \infty} \int_a^b f(x) \cos \lambda x \, dx = 0$$

在考察了教材中的证明后,发现证明中没有用到 $\sin x$ 的可导性质(甚至连续性),这时候可猜想对于一般的 **Riemann** 可积周期函数 $g(x+T)=g(x)$, 是否有类似的结论呢? 答案是肯定的:

广义 **Riemann-Lebesgue** 引理

$$\lim_{\lambda \rightarrow \infty} \int_a^b f(x) g(\lambda x) \, dx = \frac{1}{T} \int_0^T g(x) \, dx \int_a^b f(x) \, dx$$

其中, 函数 $f(x)$ 在 $[a, b]$ 上 **Riemann** 可积。

这个例子就是将结论一般化的代表。类似的推广还有将结论从低维推广到高维情形, 如重积分变量代换公式

$$\int_{\varphi(E)} f(x) \, dx = \int_E f(\varphi(t)) |\det D\varphi(t)| \, dt$$

不过, 一般化的推广并不总是正确的, 比如

$$\int \frac{1}{\sqrt{1-x^2}} \, dx = \arcsin x + C$$

是一个初等积分, 但是如果把 x^2 替换成 x^4 后, 则

$$\int \frac{1}{\sqrt{1-x^4}} \, dx = \text{非初等函数}$$

第二个积分就不是初等积分了, 称为椭圆积分。实际上将 x^2 换成 x^n , $n \geq 3$, 不定积分都不是初等函数, 这种一般化推广行不通。

低维情况不能推广到高维情况的例子中, 有一个著名代表来自复变函数。我们知道若 Ω 是复平面 \mathbb{C} 中的区域, $a \in \mathbb{C} - \Omega$, 则存在 Ω 上的解析函数 f , 不能解析延拓到 a 点, 比如

$$f(z) = \frac{1}{z-a}$$

这个结论看起来很自然。但是对高维空间 \mathbb{C}^n ($n \geq 2$) 中的情形, 这个结论并不成立。实际上根据 **Hartogs** 定理, 若 $\Omega \subset \mathbb{C}^n$ 为区域, K 是 Ω 中的紧致子集 (有界闭集), 且 $\Omega - K$ 是连通的。任取 $\Omega - K$ 上的解析函数 f , 则存在一个 Ω 上的解析函数 F , 使得 F 在 $\Omega - K$ 上等于 f 。

这个例子说明了多复变函数理论与单复变函数理论有很大的不同, 不能简

单的类比推广。正是这些不同，使得多复变函数方面的研究非常活跃。

上面举了好几个例子。当然，并不是所有数学定理都能同时满足以上 1、2、3 这三点，但在很多情况下还是可以满足的。若能带着这三个问题思考数学定理，学同样的定理，但你的思维层次加深了。传统的数学分析因为使用的数学工具有限，得到的深刻结论较少，但若将实变函数和复变函数的一些基本工具引入后，必将大大提升认识层次。实际中，可能因为自身能力有限，问题难度可能超出自己的能力圈了，比如对某个定理的特例（如 $n=2$ 的情况），我们可以证明。但对一般的情况 n ，可能原来的方法就不行了。对这一点不用气馁，我们可以请教牛人，或者多看看其它书。

笔者曾经被一些牛人指点过，受益匪浅。但是，绝大多数时候还是得靠自己思考，你不可能靠复制别人的路线而成功，关键还是要有自己的想法。对于上面提到的三个问题，笔者的经验之一是多读书，多读文章，同时注意搜集和分析不同的证明，很多时候对比不同的证明，你往往会有新的感悟（相对于只看一个证明来说）。

上面总结的 1、2、3 点思考只是投石问路，数学的联系和统一性肯定不只这三点。就联系方面，一个自然的思路是考虑逆命题是否成立，比如连续函数介值定理。我们可以提问，如果函数不连续，介值定理是否成立？其实，后面学到导数介值定理表明，介值性不是连续函数特有的。

类似的道理，我们研究完一种运算后，自然应该看看其逆运算。比如研究完微分运算后，可以考虑其逆运算，在某些情况下（如连续函数），发现了微积分基本定理，

$$\int_a^b f'(x) dx = f(b) - f(a)$$

从而揭示了微分与积分之间的互逆运算关系。

注意，这种互逆运算关系不是没条件的成立。由导数介值定理可知被积函数 $f'(x)$ 不能有第一类间断点，这是必要条件。

互逆运算的类比拓展是考虑两种运算的可交换性，这是非常重要的思路。比如，数学分析中很重要的积分号下求极限问题，通常不能交换

$$\lim \circ \int_a^b \neq \int_a^b \circ \lim$$

由此引出了**控制收敛定理**。实际应用中常见的积分号下求导运算就属于这类问题。其实我们在线性代数中已经接触到了非交换问题，如矩阵乘法（通常）

$$AB \neq BA$$

物理学中著名的海森堡测不准原理，其数学本质就是算子的非交换性。

在数学分析中，联系的普遍性还体现在很多定理具有“对偶”特点，特别体现为“连续”VS“离散”，比如

$$\lim_{x \rightarrow a+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a+} \frac{f'(x)}{g'(x)} = A \quad \text{VS} \quad \lim_{n \rightarrow \infty} \frac{b_n}{a_n} = \lim_{n \rightarrow \infty} \frac{b_n - b_{n+1}}{a_n - a_{n+1}} = A$$

罗比达法则 VS Stolz 定理

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad \& \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega t} d\omega$$

VS

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

Fourier 变换与逆变换 VS Fourier 级数

这两个例子不但形式上特别像，同时也有内在联系，具体讨论见本书第 6 章。

数学分析中的统一性也很明显，如微分学方面的柯西中值定理

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(\xi)}{g'(\xi)}$$

该定理将拉格朗日中值定理与罗尔中值定理统一起来。

积分学方面的代表，则是 Stokes 公式

$$\int_M d\omega = \int_{\partial M} \omega$$

借助外微分工具，Stokes 公式将牛顿-莱布尼茨公式、格林公式和高斯公式统一起来。

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

$$\iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \int_{\partial D} P dx + Q dy$$

$$\iiint_{\Omega} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dx dy dz = \iint_{\partial \Omega} P dy dz + Q dz dx + R dx dy$$

级数方面也可以进行统一，我们在数学分析中学到了泰勒级数

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

这实际上只是复变函数中洛朗级数的特例

$$F(z) = \sum_{n=-\infty}^{\infty} c_n (z - z_0)^n$$

从统一性角度进行思考，不免要进行抽象，使用的数学工具也越来越深。但所得结果也更深刻了（比如 **Riemann** 积分仅仅是一种特殊的线性泛函）。因此若想使得自身数学分析水平有质的飞跃，最终必须要跳出数学分析这个框架，向实变函数、复变函数和泛函分析等领域进军。因此，本书也尝试了引入一些标准数学分析教材以外的例子，如椭圆函数和自守函数（不过读者不必过于担心，这几个例子主要讲述想法，是科普性质的）。数学的联系性和统一性特点分散本书的 100 多个例子中。

全书构架上采用递进式结构。首先，笔者归纳出数学分析三大基本思想，即逼近、变换和分解，这是基本功。其次，抓问题核心、类比思维、形式推导与硬算，这四项属于中阶技能。最后，几何直观、逆向思维、构造方法、多读著名定理和反例，这五项属于高阶技能，作用是如虎添翼。总的来说，这 12 个思维模块基本代表了笔者目前对分析的理解水平。相信读者读过本书，必有收获。

评注：跨数学分支联系和统一性是 20 世纪数学的特征，这方面最杰出代表是著名的 **Atiyah-Singer** 指标定理，这里简单科普一下。

考虑线性算子 $D: V \rightarrow W$ ，其中 V, W 是线性空间。指标定理源于研究方程 $Df = g$ ，其中 $g \in W$ 给定，而 $f \in V$ 未知。记核空间

$$\ker D = \{f \in V \mid Df = 0\},$$

相应的余核空间（商空间）记为

$$\text{coker } D = W / DV$$

如果 $\dim \ker D$ 和 $\dim \text{coker } D$ 均是有限的，定义其维数差

$$\text{Ind}(D) = \dim \ker D - \dim \text{coker } D$$

为 D -指标。

考虑椭圆微分方程 $Df = g$ （此时 D 称为椭圆微分算子，是一类特殊的线性算子），数学上证明了对椭圆微分算子来说， $\dim \ker D$ 和 $\dim \text{coker } D$ 均是有限维的。此时

$$\text{Ind}(D) = \dim \ker D - \dim \text{coker } D$$

称为解析指标。

注：核空间和余核空间是研究解的椭圆方程惟一性和存在性的重要路径。

椭圆微分算子 D 可以推广到更一般的空间上，如紧致、可定向、光滑流形 M 上，在流形上考虑数学问题也是现代数学的一大特色。前苏联大数学家盖尔芳德发现椭圆微分算子 D 的解析指标是同伦不变的，于是他猜想解析指标 $\text{Ind}(D)$ 应该是一个拓扑不变量，但他不知道是哪些拓扑不变量，因而没能证明。简而言之，指标定理问题可以概括为（盖尔芳德猜想）：

如何用流形 M 本身的拓扑不变量表示解析指标 $\text{Ind}(D)$ ？

经过数学家 Atiyah 和 Singer 经过努力证明了下面：

Atiyah-Singer 指标定理（紧支上同调形式）：

$$\dim \ker D - \dim \text{coker } D = \int_{T^*M} \text{Ch}[\sigma(D)] \cdot \text{Td}(TM \otimes \mathbb{C})$$

其中， T^*M 表示余切丛， $TM \otimes \mathbb{C}$ 表示复化切丛， Ch 称为陈特征（纪念几何大师陈省身陈老的杰出贡献）， $\sigma(D)$ 称为主象征， Td 称为 Todd 类。

上面指标定理等式右边非常复杂，难怪数学大师盖尔芳德没想出来。右边这些量虽然复杂，但对于拓扑学家来说，都是很熟悉的拓扑不变量（这里用 de Rham 上同调形式表示）。

因为陈类是流形上的拓扑不变量，由 de Rham 定理知陈类可由微分形式表示，而陈特征和 Todd 类都可以表示成陈类的多项式，用 $\{\text{Chern-Class}\}$ 表示与陈类有关的多项式，所以指标定理可写成下面的底流形上“科普形式指标定

理”（这种科普写法来源于张伟平院士）：

$$\dimker D - \dimcoker D = \int_M \{Chern - Class\}$$

注：读者可以把 $\{Chern - Class\}$ 想象为一大堆复杂的微分形式，上述公式表述方法比较直观，又不失数学严密性，还突出了陈省身（陈老）的数学贡献。网上有张伟平院士关于《Atiyah-Singer 指标定理》的科普讲座（第31讲，见下方链接），感兴趣的读者可以去看看。

http://www.nim.nankai.edu.cn/activites/lecture_2007.htm

这里不可能去解释上面一大堆抽象的数学概念，不过可以科普下指标定理的证明思想，想法比较有代表性。

我们定义“拓扑指标” $t-ind$ 为

$$t-ind = \int_{T^*M} \text{Ch}[\sigma(D)] \cdot \text{Td}(TM \otimes C)$$

Atiyah-Singer 指标定理说明我们用两种截然不同的方法得到的指标是相等的，即解析指标=拓扑指标，或 $\text{Ind}(D) = t-ind$ 。注意到前者来自于分析领域，后者来自于拓扑领域，指标定理把分析和拓扑两大数学领域联系到一起。

证明方法采用的是“公理化”思路，先建立两条“指标公理（拓扑思路）”，然后证明满足这两条公理的指标若存在必惟一。接下来验证拓扑指标 $t-ind$ 满足公理，这就证明了满足公理的指标存在性。最后，只需要验证解析指标 $\text{Ind}(D)$ 分别满足这两条公理（技术上最大的困难所在）即可，由惟一性

$$\text{Ind}(D) = t-ind$$

这种公理化的证明思路值得借鉴。

指标定理不仅体现了数学的联系性，也体现了数学的统一性。在此之前，人们已经证明了 Gauss-Bonnet 定理、Riemann-Roch 定理、符号差定理，这都是数学中大名鼎鼎的定理。而指标定理的问世，直接将上述定理变为了特例，这就是统一性的威力。

最后科普一段故事，讲述的是指标定理是如何问世的。Atiyah 是一个拓扑学家，最初发现所谓 Spin 流形的 A-亏格是整数，始终不解。于是他向分析学家 Singer 请教，Singer 起初也想不明白，无法解释。后来有一天 Singer 来了灵感，发现 A-亏格是 Dirac 算子（粗略地说，可以看成 Laplace 算子的平方根）的解

析指标，因而 Atiyah 和 Singer 开始钻研 Dirac 算子。后来 Smale（证明 4 维以上 Poincare 猜想那位大师）路过牛津，与 Atiyah 聊起彼此工作时，Atiyah 提到了那个问题。Smale 说前几天他看到盖尔芳德的一篇关于椭圆微分算子指标的文章，建议 Atiyah 去看看。Atiyah 后来感慨到，需要突破 Dirac 算子的局限而考虑更广泛的椭圆算子，这才有了 Atiyah-Singer 指标定理。

最初指标定理是利用配边理论证明的，但 Atiyah 觉得方法不够自然，而且技巧性太强。Atiyah 冥思苦想，终于发展了所谓拓扑 K -理论，有了这门工具作为铺垫之后，指标定理的证明就水到渠成了。对于余切丛 T^*M ，存在一个 K -群，记为 $K(T^*M)$ ，由其中可以自然生成拓扑指标 $t-ind$ ，于是指标定理的整体证明思路可浓缩概括为下列交换图表：

$$\begin{array}{ccc}
 \{\textit{elliptic} - D\} & \xrightarrow{\textit{Ind}} & \mathbf{Z} \\
 \sigma(D) \downarrow & & \updownarrow \\
 K(T^*M) & \xrightarrow{t-ind} & \mathbf{Z} \\
 \downarrow & & \\
 \textit{Ind}(D) = t-ind(\sigma(D)) & &
 \end{array}$$

需要指出，拓扑 K -理论的应用范围不仅仅是指标定理，指标定理只是该理论的副产品。

第2章 数学分析三大基本思想之逼近

从本章开始，计划用三篇文章来介绍数学分析中的三大基本思想，即**逼近**、**变换**和**分解**。

对数学分析普遍的看法是，上手容易，难以精通，易学难精。除了数学分析科目本身博大精深外（想想数分后续科目复变、实变、ODE 和 PDE），造成这种现象的原因之一是数学分析中的技巧特别多，初步统计包括逼近（放缩和夹逼）、变换（等价无穷小和不定积分换元）、分解（级数展开和累次积分）、反演变换、递推、RMI 原理、对称、引入参变量（收敛因子）、极端法、归纳法、构造法和计算两次法等等。这些方法技巧，本身也是宝贵的数学思想，它们分散在数学分析的若干定理证明和经典习题当中，单独拿出来都能写一篇文章。这方面详细内容可参考谢惠民等编著的《数学分析习题课讲义》一书，该书把上述知识结构梳理的很全面，笔者强烈推荐。

但凡事大都有主要矛盾，学数学分析也应该抓主要思想。根据笔者这些年对数学分析的体会，感觉有三大基本思想可以作为数学分析的核心，即**逼近**、**变换**和**分解**。围绕微分、积分和级数这三大主题，展开上述三大基本思想，构成了数学分析的主干。不论是初学者，还是重温者，抓住上面的主干，就有了大方向。本章先介绍数学分析第一大基本思想：**逼近**。

数学分析处理的对象涉及无穷，因而必须考虑极限。那么如何严格描述极限过程呢？为此，引入了 $\varepsilon-\delta$ 语言（相当于公理），所讨论的极限，即 $\varepsilon-\delta$ 意义下的极限。本文所讨论的逼近，也是 $\varepsilon-\delta$ 意义下的逼近。出于对几何直观的偏爱，笔者更喜欢用 $x_n \rightarrow x_0$ （箭头）表示逼近过程，而不是用 $|x_n - x_0| < \varepsilon$ 。

谈及分析中的逼近，免不了涉及不等式，这也是分析难学的原因之一。毕竟，不等式比等式难处理多了，而且经常要分段估计（如三段法）。逼近是一种重要的思想，用简单的近似复杂的，用熟悉的近似陌生的，这种思想绝非数学独有。

回想极限的定义、连续的定义、定积分的定义，都体现了数学中的逼近思想。当然，极限和连续可以用拓扑的方法定义，但失去了直观性。学数学分析还是尽量采用直观的方法为好，至少在水平达到一定程度之前如此。

逼近的常见技巧是放缩和夹逼，方法的框架很简单，但具体应用时需要与具体问题相结合。本文重点介绍逼近的思想在数学分析中的体现，下面用两个例子来说明。

第一个例子：闭区间上的连续函数可由多项式一致逼近。

这是魏尔斯特拉斯的多项式逼近连续函数定理，学过数分的人都知道。多数教材对这个定理的证明采用伯恩斯坦多项式的构造方法。这里我们不去探讨证明过程，重点说明一个有趣的数学证明思路，想法来自概率论。

考虑独立重复的伯努利实验，它有两个结果 A, B ，出现结果 A 的概率为 x ，出现结果 B 的概率为 $1-x$ 。则 n 次独立重复实验中，恰好出现 k 次的结果 A 的概率为

$$P_n(k) = C_n^k x^k (1-x)^{n-k}$$

显然， $\sum P_n(i) = 1$ 。这种概率分布也叫二项分布。固定 n 和 x 不动，将 $P_n(k)$ 视作变量 k 的函数。这个函数在 nx （不一定是整数）附近有一尖峰，在两侧迅速递减。于是直观上可以设想，对这样的 $nx \approx k$ （以下默认）， $P_n(k)$ 比较接近于 1（在和式 $\sum P_n(i) = 1$ 中权重较大）。因此，下面的加权逼近想法就自然了：

$$\sum_{i=1}^n P_n(i) f\left(\frac{i}{n}\right) \longrightarrow f\left(\frac{k}{n}\right) \approx f(x)$$

记多项式 $B_n(f, x) = \sum_{i=0}^n f\left(\frac{i}{n}\right) C_n^i x^i (1-x)^{n-i}$ ，利用初等不等式放缩技巧可以证

明：当 $n \rightarrow \infty$ 时，多项式序列 $B_n(f, x)$ 一致逼近连续函数 $f(x)$ 。

注：证明细节可参考张筑生老师的《数学分析新讲》第三册。

也有教材采用高级工具来证明，如核函数方法。细节见卓里奇的《数学分析》第二册，或者《新讲》第三册含参数积分章节。笔者偶然在网上看到一位台湾数学老师介绍多项式逼近连续函数的好文章，重点就是介绍用“核函数”法来构造逼近多项式（推荐大家读一读）：

Weierstrass 逼近定理

本质上，这种核函数法利用了广义函数中的狄拉克函数序列性质，详细内容可参考广义函数或数理方程的相关书籍。这方面一个著名的结论是

$$f(x) = \int_{\mathbb{R}} f(t) \delta(x-t) dt$$

除了核函数外，还涉及到卷积工具，所构造的“核函数序列”的极限具有狄拉

克函数 δ 性质。方法的巧妙之处在于将卷积（积分）分成两段 $|x| \geq \varepsilon$ 和 $|x| < \varepsilon$ 进行分段估计，这从证明过程中可以看出。

核函数： 设 $\Delta_n(x)$ 是定义在 $(-\infty, +\infty)$ 上以 n 为参数的函数，并且满足：

1. 非负性，即 $\Delta_n(x) \geq 0$ ；
2. 广义可积性（正则性），即

$$\int_{-\infty}^{+\infty} \Delta_n(x) dx = 1$$

3. 狄拉克序列性质，即 $\forall \delta > 0$ ，有

$$\lim_{n \rightarrow \infty} \int_{|x| < \delta} \Delta_n(x) dx = 1$$

注：这里的 $\Delta_n(x)$ 是正核，实际应用中也会遇到非正核，如 Fourier 级数中的狄利克雷核。另外，参数也不一定是 n ，可以是连续变量。

这种核函数图形有一个尖点，两旁迅速下降，其极限情况就是狄拉克 δ -函数。

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases} \text{ 以及 } \int_{-\infty}^{+\infty} \delta(x) dx = 1$$

如果这个核函数比较光滑，且能展开成一致收敛的幂级数，那么截取前面有限部分就构造出了逼近多项式，这也是魏尔斯特拉斯本人原来证明的思路。魏尔斯特拉斯最初的证明，使用的核函数是“热核”（正态核）

$$\Delta_h(x) = \frac{1}{h\sqrt{\pi}} e^{-\frac{x^2}{h^2}}$$

其中参数 h 是连续变量。上面这个函数也常见于正态分布。

然后，魏尔斯特拉斯证明了当 $h \rightarrow 0+$ 时，卷积

$$f * \Delta_h(x) = \frac{1}{h\sqrt{\pi}} \int_{-\infty}^{+\infty} f(u) e^{-\left(\frac{u-x}{h}\right)^2} du$$

一致收敛于 $f(x)$ 。

最后，魏尔斯特拉斯将问题限制在闭区间 $[a, b]$ 上，并将热核函数 $\Delta_h(x)$ 进行幂级数展开，再截取前边若干有限项截断，就得到了一个多项式 $Q(x)$ ，用它代替 $\Delta_h(x)$ ，再利用卷积构造出了多项式 $P(x)$ ，这样就证明了多项式逼近连续函数定理。

通常教材上选取的核函数是 Landau 核，这个核函数本身就是多项式，因此相比原证明减少了一步（即避免了 $Q(x)$ 近似 $\Delta_h(x)$ ）这一步，但本质没有改变。

Landau 的证明用到了核函数:

$$\Delta_n(x) = c_n(1-x^2)^n, -1 \leq x \leq 1$$

然后构造多项式序列

$$P_n(x) = \int_0^1 f(u) \Delta_n(u-x) dx$$

以上核函数方法均出现了卷积, 卷积的好处是使一个函数“光滑化”, 常用来构造一些光滑函数, 这种思想在含参变量积分中很普遍。

根据概率论知识, 发现二项分布的极限分布(中心极限定理)是正态分布(也就是热核), 不知道伯恩斯坦本人是否受此启发, 想到了巧妙的伯恩斯坦多项式 $B_n(f, x)$. 细心观察不难发现, 伯恩斯坦多项式证明法相当于离散核。

不过笔者感到最直观方法, 还是勒贝格的“折线逼近法”, 大体思路是: 闭区间上的连续函数可以用折线逼近(比较直观)。而折线函数可以用函数 x 和 $|x|$ 进行线性组合而得到。考虑到 $|x| = \sqrt{x^2}$, 经过一些代换, 利用二项级数展开, 得到了一致收敛的幂级数, 这说明 $|x|$ 可以用多项式一致逼近。

考虑二项级数展开

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$$

当 $\alpha > 0$ 时, 上述级数在闭区间 $[-1, 1]$ 上一致收敛于函数 $(1+x)^\alpha$.

于是在闭区间 $[0, 2]$ 上 \sqrt{x} 可以展开成一致收敛的幂级数

$$\sqrt{x} = [1 + (x-1)]^{\frac{1}{2}} = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} (x-1)^n$$

这样我们证明了:

引理 1: 连续函数 $\varphi(x) = \sqrt{x}$ 在闭区间 $[0, 1]$ 上可以用多项式一致逼近。

再定义函数

$$\lambda(x) = \frac{|x|+x}{2} = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$$

不难证明, 连续函数 $\lambda(x)$ 可在任意任意闭区间 $[-c, c]$ 上用多项式一致逼近。

引理 2: 定义在闭区间 $[a, b]$ 上的任意折线函数 $\Lambda(x)$ 都可以表示成

$$\Lambda(x) = c + c_0 \lambda(x-x_0) + \cdots + c_m \lambda(x-x_m)$$

因而定义在闭区间上的任意折线函数都可以用多项式一致逼近。

有了上面的多项式逼近折线函数定理后，可利用函数的一致连续性证明：

闭区间 $[a, b]$ 上的连续函数可由折线函数一致逼近。

上述思路可概括为：多项式 \rightarrow 折线函数 \rightarrow 连续函数，具体证明请参考本书附录。

闭区间上多项式逼近连续函数定理不能推广到 \mathbb{R} 上。对此，可进行逆向思考，如果可以，则多项式序列 $p_n(x)$ 是柯西列。但 $m \neq n$ 时，多项式 $p_n(x) - p_m(x)$ 是非常数多项式，不可能是柯西列。进一步可以证明，若 \mathbb{R} 上的连续函数 $f(x)$ 可由多项式一致逼近，则 $f(x)$ 必是多项式。

多项式逼近连续函数定理，也可以推广到三角多项式（如 Fourier 级数中的费耶和），甚至更一般的抽象空间，即斯通-魏尔斯特拉斯定理（见《新讲》第三册）。另一个有意思的推广是所谓 Muntz-Szasz 定理：

将多项式函数基底 $1, x, x^2, x^3, \dots$ 替换成基底 $1, x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \dots$ ，之后，魏尔斯特拉斯的逼近定理仍然成立。即

$$\overline{\text{span}\{1, x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \dots\}} = C[a, b]$$

这里 $0 < \lambda_1 < \lambda_2 < \lambda_3 < \dots$ ，且满足

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} = +\infty$$

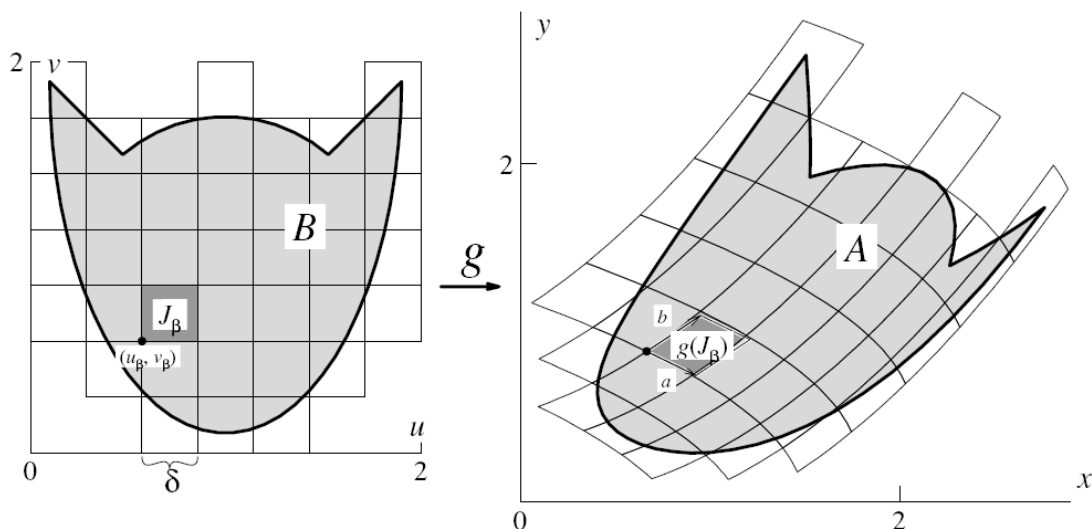
注：此时基底 $1, x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \dots$ 的线性组合不一定是多项式。Muntz-Szasz 定理的证明需要复变函数和泛函分析知识，读者可参考 Rudin 的《实分析与复分析》或者 P.Lax 的《泛函分析》。

第二个例子：重积分变量换元公式。

由于积分的 Riemann 和定义可以看做是“体积线性组合”（然后再求极限），所以解决在变换 φ 下的体积变化关系是解决问题的关键。于是，考察一个小立方体 C 在变换 φ 作用下的像集 $\varphi(C)$ 的体积就是自然思路了。但是，由于 φ 一般是非线性变换，所以像集 $\varphi(C)$ 多半要发生扭曲变形。直观上，如果立方体 C 充分小，变换 φ 又比较光滑，那么 $\varphi(C)$ 扭曲程度就不那么厉害。

由微分思想可知，一个映射 φ 的作用在一点 x_0 邻域范围内可以近似看做是一个局部线性映射 $L(x) = \varphi(x_0) + \varphi'(x_0)(x - x_0)$ 。稍微精确一些描述，取定小立方体 C 的中心点 x_0 ，那么当 C 充分小时，线性变换 $L(x) = \varphi(x_0) + \varphi'(x_0)(x - x_0)$ 与变换 $\varphi(x)$ 相差高阶无穷小量。也就说处理问题时， $\varphi(x)$ 局部上可由线性变换

近似代替，微分思想的精华也在于此。



借用上图说明一下（这里 $g = \varphi$ ）

根据线性代数中的一个结论，对可逆线性变换 A ，有体积关系

$$\mu(A(C)) = |\det A| \mu(C)$$

这里用符号 μ 表示若尔当体积。

令线性变换 $A = L$ ，则 $\det A = \det \varphi'(x_0)$ 或写成 $\det A = \det D\varphi(x_0)$ 。

由此得到体积近似关系： $\mu(\varphi(C)) \approx |\det \varphi'(x_0)| \mu(C)$ 。进而

$$f(x)\mu(\varphi(C)) \approx f(\varphi(t))|\det \varphi'(t_0)|\mu(C)$$

其中 $x = \varphi(t)$ 。根据积分的 Riemann 和定义，不难猜出重积分变量代换公式

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt$$

此处 x, t 都是向量， $D\varphi(t) = \varphi'(t)$ 表示雅可比矩阵（导数的推广）。上述想法表述为下面的重积分变量代换公式：

设 $\Omega \subset \mathbf{R}^m$ 是一个开集，

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射， $E \subset \Omega$ 是一个闭若尔当可测集。如果

- (1) 雅可比行列式 $\det D\varphi(t) \neq 0, \forall t \in \text{int } E$;
- (2) φ 在 $\text{int } E$ 中是单射。

那么 $\varphi(E)$ 也是一个闭若尔当可测集，并且对于任何在 $\varphi(E)$ 上连续的函数 $f(x)$ 都有

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt.$$

这个定理的证明有些复杂, 细节见《数学分析新讲》第二册和本书的附录部分, 基本想法是“采用把局部微分同胚分解成简单微分同胚的复合, 然后再转化为累次积分来证明”。

《新讲》中的证明方法, 将在“变换”一章中细说。这里, 说说该定理的 **Riemann** 和证明法, 引自下面的《美国数学月刊》文章。这里将该文章的核心部分翻译过来, 大家可以感受一下逼近法的巧妙证明。

对于连续函数 $f(x)$ 可分离其正部和负部以分别讨论积分, 而不影响最终的结果。所以, 以下不妨设 $f(x) \geq 0$ 恒成立。

THE FORMULA FOR CHANGE IN VARIABLES IN A MULTIPLE INTEGRAL

J. SCHWARTZ, Yale University

A well-known theorem of elementary analysis states:

THEOREM. *Let D_1 and D_2 be open sets in Euclidean n -space E^n , and let $h: D_1 \rightarrow D_2$ be a 1-1 mapping of D_1 onto D_2 such that h and its inverse h^{-1} are continuous and have continuous derivatives. Let*

$$J(x) = \left| \det \left(\frac{\partial h^i(x)}{\partial x^j} \right) \right|$$

be the absolute value of the Jacobian determinant of the transformation h . Then a function $f(x)$ is integrable over the domain D if and only if the function $f(h(x))J(x)$ is integrable over D_1 , and we have

$$(1) \quad \int_{D_2} f(x) dx = \int_{D_1} f[h(x)]J(x) dx.*$$

限于篇幅, 本文只给出上述文章的最核心内容。首先, 设 D_1, D_2 是欧氏空间 \mathbf{R}^n 中的开集, $h: D_1 \rightarrow D_2 = h(D_1)$ 是一个双射, 且映射 h 与逆映射 h^{-1} 是连续可微的。即 $h: D_1 \rightarrow D_2$ 是微分同胚。

(重积分变量代换定理) 记符号

$$J(x) = \left| \det \left(\frac{\partial h^i(x)}{\partial x_j} \right) \right|$$

表示映射 h 对应的雅可比矩阵“行列式的绝对值”。则函数 $f(x)$ 在 D_2 上可积, 当且仅当 $f(h(u))J(u)$ 在 D_1 上可积, 以及

$$\int_{D_2=h(D_1)} f(x) dx = \int_{D_1} f[h(u)]J(u) du$$

证明: 对 $x \in \mathbf{R}^n$, 引入 **max** 范数 $|x| = \max_{1 \leq i \leq n} |x_i|$, 其中 x_i 是向量 x 的第 i 个坐标分

量。于是一个中心为点 p , 棱长为 $2s$ 的立方体可以表示成 $|x - p| \leq s$ 。

如果 A 是一个线性变换, 对应矩阵为 $(a_{ij})_{n \times n}$, 则有

$$A(x) = A[x_1, \dots, x_n] = \left[\sum_{j=1}^n a_{1j} x_j, \dots, \sum_{j=1}^n a_{nj} x_j \right]$$

再定义矩阵范数

$$|A| = \max_{1 \leq i \leq n} \left| \sum_{j=1}^n a_{ij} \right|$$

由此得到 $|A(x)| \leq |A| |x|$ 。

对映射 $h(x) = [h_1(x), \dots, h_n(x)]$, 记其雅可比矩阵 $j(x) = j_{ik}(x)$ 为

$$j_{ik}(x) = \frac{\partial h_i(x)}{\partial x_k}$$

集合 $S \subset \mathbf{R}^n$ 对应的体积(测度)记为 $\mu(S)$ 。则一个中心为点 p , 棱长为 $2s$ 的立方体 $C \subset D_1$ 可以表示成 $|x - p| \leq s$, 其对应体积 $\mu(C) = (2s)^n$, 由此可得到下面的拉格朗日中值定理

$$h_i(x) - h_i(p) = \sum_{k=1}^n j_{ik}[p + \theta_i(x)(x - p)](x_k - p_k)$$

其中 $0 \leq \theta_i(x) \leq 1$ 。所以, 得到

$$|h(x) - h(p)| \leq s \max_{y \in C} |j(y)|$$

换句话说, 像集 $h(C)$ 包含在立方体

$$|z - h(p)| \leq s \max_{y \in C} |j(y)|$$

之中。由此得到体积估计

$$\mu(h(C)) \leq \left\{ \max_{y \in C} |j(y)| \right\}^n \mu(C)$$

将上述结论应用于映射 $A^{-1}h$, 这里 A 是可逆线性映射。同时利用线性代数中的一个重要结论: $\mu(A^{-1}(S)) = |\det A^{-1}| \mu(S)$, 其中 S 为闭集。令 $S = h(C)$, 因为 C 是有界闭集, 映射 h 是连续可微的, 所以

$$\mu(A^{-1}h(C)) = |\det A^{-1}| \mu(h(C)) \leq \left\{ \max_{y \in C} |A^{-1}j(y)| \right\}^n \mu(C)$$

移项得到

$$\mu(h(C)) \leq |\det A| \left\{ \max_{y \in C} |A^{-1}j(y)| \right\}^n \mu(C)$$

这一不等式的妙用将在下面的 **Riemann** 和中得到体现。将立方体 C 分割成有限个内部互不相交的小立方体 C_1, C_2, \dots, C_M , 对应中心分别为 x_1, x_2, \dots, x_M . 记 $\delta > 0$ 为这样的正数, 它比上述 M 个小立方体的棱长都要小。

对立方体 C_i , 取 $A = j(x_i)$, 再应用上面的不等式以及体积的可加性, 得到

$$\mu(h(C)) \leq \sum_{i=1}^M |\det j(x_i)| \left\{ \max_{y \in C_i} |j(x_i)^{-1}j(y)| \right\}^n \mu(C_i)$$

因为矩阵函数 $j(x)$ 是连续函数, 所以当 $z \rightarrow y$ 时,

$$j^{-1}(z)j(y) \rightarrow I \quad (\text{单位矩阵})$$

由此得到不等式估计

$$\left\{ \max_{y \in C_i} |j(x_i)^{-1}j(y)| \right\}^n \leq 1 + \eta(\delta)$$

其中正数 $\eta(\delta)$ 满足: $\delta \rightarrow 0$ 时, $\eta(\delta) \rightarrow 0$.

进而得到核心不等式估计

$$\mu(h(C)) \leq [1 + \eta(\delta)] \sum_{i=1}^M |\det j(x_i)| \mu(C_i)$$

注意到上面不等式右边实际是体积的 **Riemann** 和, 令 $\delta \rightarrow 0$ 得到

$$\mu(h(C)) \leq \int_C J(x) dx$$

对于连续函数 $f(x)$ 可分离其正部和负部分别讨论积分, 而不影响最终的结果。所以这里不妨设 $f(x) \geq 0$ 恒成立。根据积分的 **Riemann** 和定义, 得到

$$\int_{D_2=h(D_1)} f(x) dx \leq \int_{D_1} f[h(u)]J(u) du$$

注: 为避免符号混淆, 这里对 D_1 上的积分, 采用积分变量 u 来表示。

到此为止, 我们得到的是一个不等式, 下面巧妙地利用逆变换 h^{-1} 得到反向不等式 (对称性体现), 进而得到等式。

将变换 $u = h^{-1}(x)$ 代入上面积分不等式, 得到

$$\int_{D_2} f(x) dx \leq \int_{D_1} f[h(u)]J(u) du \leq \int_{D_2} f[h \circ h^{-1}(x)]J(h^{-1}(x))J'(x) dx = \int_{D_2} f(x) dx$$

其中, $J'(x)$ 是映射 $h^{-1}: D_2 \rightarrow D_1$ 对应的雅可比行列式的绝对值。且满足

$$J(h^{-1}(x))J'(x) = 1$$

所以, 我们证明了重积分变量代换定理。

评注: 对比重积分变量代换定理的证明,《新讲》中给出的是变换形式的证明,而上面是 **Riemann** 和的不等式估计证明。就证明本身来说,后者更基本,技术上也简单一些。但是,《新讲》中的证明方法却是现代数学中的标准方法,思想上更丰富。必须纠正一个误区,与考试不同,学数学最忌讳以证明方法的难度和简洁与否来评估。比如,后来人找到了素数定理的初等证明,但这个初等证明并没有带来新的东西。初中里,传统的平面几何有很多优美巧妙的简洁证明方法,但就深刻性与普适性来说,解析几何却更胜一筹。好的方法应该包含深刻的数学思想,同时打开一片新的天地。搞数学不排斥“杀鸡用牛刀”的方法,只要你能先证明出来,方法没有限制。

第3章 数学分析三大基本思想之变换

本章介绍数学分析中的三大基本思想之变换。广义的变换应该作为一种思想来理解，即对某个数学对象进行操作，转化为另一个对象，要求后者相对容易处理。当然，很多问题还要加上逆变换。

我们在中学已经接触到一些简单的变换，如三角变换和万能公式。进入大学学习数学分析或微积分，接触到求极限方法的等价无穷小代换也是变换的一种。像微分理论中的罗比达法则，积分理论中的第一、二类换元法，级数理论中的算术平均求和法（即 $C-1$ 求和法，例如 Fourier 理论中的费耶和），常微分方程中的分离变量法、常数变易法和积分因子法等等，都是变换思想的具体表现。

从实用角度看，最常见的是线性变换，这方面内容在线性代数课程中有详细论述。后来人们进一步抽象出线性算子，并把它推广到无穷维线性空间，这些内容可以在泛函分析课程中学到。需要指出，非线性变换也是存在的，而且更普遍，处理起来也更困难。数学分析作为基础课程，主要以处理线性变换为主（包括逆变换），微分学的精髓就是函数增量的线性主部，站在算子角度，微分是线性变换的一种。另外，不少非线性变换的处理方法是从线性变换中衍生出来的，这就更体现出线性变换的重要性。

还是用几个具体的“变换例子”来说明吧。首先说说不定积分，这方面技巧非常多，很多教材都单独列出一节讲一些经典的不定积分技巧，比如有理函数如何积分，某些根式如何积分等等。其中使用得最多的是分部积分法和第一、二类换元积分法。

利用微分公式 $d(uv) = vdu + u dv$ ，经过移项变换，得到

$$u dv = d(uv) - v du$$

再求逆变换（积分）的结果

$$\int u dv = uv - \int v du$$

这就是分部积分法，该变换的指导精神是不定积分 $\int v du$ 比较容易求出。

第一类不定积分换元法的思想是凑微分

$$\int f(u(x))u'(x) dx = F(u(x)) + C, \quad F'(t) = f(t)$$

关于第二类不定积分换元法，实际上是逆变换法。通常涉及反函数，很多

书中给的条件并不一致。《数学分析习题课讲义》中收录了下面的定理：

设 $f(x)$ 有原函数，函数 $x = x(t)$ 可微且有函数 $t = t(x)$ 满足 $x(t(x)) \equiv x$ ，若

$$\int f(x(t))x'(t) dt = F(t) + C$$

则有

$$\int f(x) dx = F(t(x)) + C$$

证明：已知 $f(x)$ 有原函数，记为 $U(x)$ ，则有 $U'(x) = f(x)$ 。又已知

$$F'(t) = f(x(t))x'(t)$$

由复合函数求导法则得到

$$\frac{dU(x(t))}{dt} = U'(x(t))x'(t) = f(x(t))x'(t) = F'(t)$$

因此 $U(x(t))$ 与 $F(t)$ 只相差一个常值函数

$$U(x(t)) = F(t) + C$$

将 $t = t(x)$ 代入，且利用恒等式 $x(t(x)) \equiv x$ ，于是就有

$$U(x(t(x))) = U(x) = F(t(x)) + C$$

对上面两边求导可知

$$\frac{dF(t(x))}{dx} = U'(x) = f(x)$$

于是我们证明了

$$\int f(x) dx = F(t(x)) + C$$

一般教材上使用第二换元积分法时要求：可微函数 $x = x(t)$ 存在“反函数 $t = t(x)$ ”关系。显然，若反函数存在，则自然满足恒等式 $x(t(x)) \equiv x$ 。但反过来满足恒等式 $x(t(x)) \equiv x$ 并不意味着反函数存在。上面的定理实际上说明：只需要函数 $x = x(t)$ 可微且有函数 $t = t(x)$ 满足恒等式 $x(t(x)) \equiv x$ ，即可使用第二换元积分法。无须反函数存在或单调性假设。

第一、二类换元积分法主要是“一阶微分形式不变性”的体现，目的是简化被积函数的结构，使得容易求解。这两类换元积分法可以推广到一元定积分理论中，这里不细说了，详细内容请参考《数学分析新讲》。重积分中也有换元积分公式，但根据经验，高维积分换元主要是为了将积分域变得简单，一维积分换元主要是为了将被积函数变得简单，这一点区别要注意。

尽管不定积分的技巧非常多，但必须指出，大部分初等函数是积不出来的，

即原函数是非初等函数。比如下列积分

$$\int \frac{\sin x}{x} dx, \int e^{-x^2} dx, \int \frac{1}{\ln x} dx, \int \sin x^2 dx, \int \cos x^2 dx$$

即便是不涉及超越函数的第一类椭圆积分

$$\int \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

其原函数也是非初等的。原函数是非初等积分也称不可积。

据笔者查阅到的文献，目前还没有判定一个连续函数的不定积分是否为非初等积分的“万能准则”（充要条件），大部分判定定理只能针对某一类函数成立。比如下面的定理

（定理）设 f, g 为有理函数， g 不是常值函数，如果 $\int f(x)e^{g(x)} dx$ 是初等函数，则存在有理函数 h ，使得

$$\int f(x)e^{g(x)} dx = h(x)e^{g(x)} + C.$$

这个定理的证明涉及微分代数，有兴趣的读者可自行查阅相关文献。

在积分理论中，有一类非常重要的变换，即含参数变换。

$$\varphi(t) = \int_{\mathbf{R}} f(x)K(x, t) dx$$

这里 t 是参变量， $K(x, t)$ 称为核函数。这种方法是定义非初等函数的一种常见方法，比如数学分析里非常重要的 Γ 函数和 B 函数。参变量核函数方法的典型例子有两个，一个是多项式逼近连续函数定理中的 Landau 方法

$$P_n(t) = \int_0^1 f(x) \Delta_n(t-x) dx$$

其中核函数为

$$K(x, t) = \Delta_n(t-x) = c_n(1-(t-x)^2)^n$$

另一个例子是，Fourier 求和理论中狄利克雷积分

$$S_m(x, f) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \frac{\sin(m + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt$$

这里不仅涉及核函数，也涉及到卷积运算。关于卷积运算的进一步内容可参考卓里奇的《数学分析》下册。

细细品味积分，发现其是一种加权求和的极限过程，含有平均的意味。很多函数经过积分处理（比如卷积），得到的新函数性质往往比较好，可以进行求导、积分等光滑运算。其中有一种方法令笔者印象非常深刻，称之为收敛因子法。下面用例子来说明，考虑广义积分

$$\int_0^{+\infty} \frac{\sin x}{x} dx = ?$$

我们引入一个收敛因子 e^{-tx} ，然后定义一个新函数

$$\varphi(t) = \int_0^{+\infty} e^{-tx} \frac{\sin x}{x} dx$$

数学分析中证明了新函数 $\varphi(t)$ 有很好的性质，可以积分号下直接求导，于是

$$\varphi'(t) = - \int_0^{+\infty} e^{-tx} \sin x dx = - \frac{1}{1+t^2}$$

等式右边的函数可以求出原函数，再利用放缩和无穷过程，不难得到

$$\int_0^{+\infty} \frac{\sin x}{x} dx = \varphi(0) = C = \frac{\pi}{2}$$

能够熟练地构造收敛因子，并进行微分和积分计算，这可是硬功夫和绝技。物理大师费曼就特别擅长此法，根据其在自传中回忆，他一直不会用复变函数中的围道积分方法，所以一直用含参数积分方法（积分号下求导等），并且解决了很多数学系棘手的问题。

微分中比较典型的代表是凑全微分的积分因子法，考虑下面的微分形式

$$P(x, y)dx + Q(x, y)dy$$

如果能找到一个可微函数 $U(x, y)$ ，使得存在函数 $M(x, y)$ ，满足

$$dU(x, y) = M(x, y)P(x, y)dx + M(x, y)Q(x, y)dy$$

则称 $M(x, y)$ 为积分因子。

这种凑全微分的积分因子法很像高中时的凑不等式方法，技巧性和经验性较强。若能成功找到积分因子，解决问题还是相当简洁的。

应用例子：考虑微分方程

$$\frac{dy}{dx} + P(x)y = Q(x)$$

换一种表示： $dy + P(x)ydx = Q(x)dx$

取积分因子 $\mu = e^{\int P(x)dx}$ ，则

$$d\left(ye^{\int P(x)dx}\right) = Q(x)e^{\int P(x)dx}dx$$

等式两边直接积分可以求出 y 。这里

$$U(x, y) = ye^{\int P(x)dx}, \quad M(x, y) = \mu$$

级数理论中也有变换的代表，如著名的阿贝尔变换（分部求和）和算术平均求和法。这里主要说说后者，记

$$S_n = \sum_{k=1}^n a_k$$

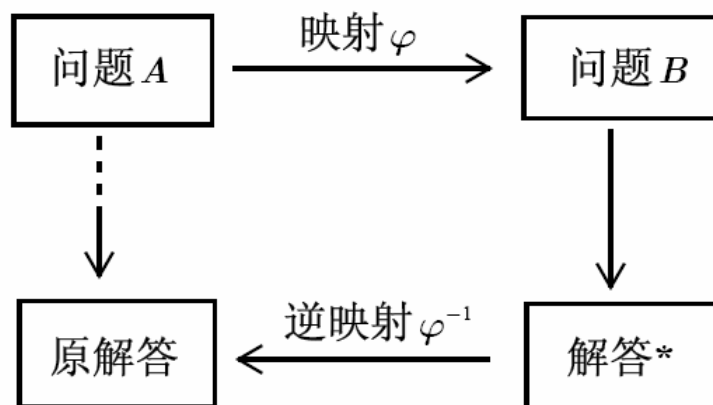
算术平均求和法即考虑

$$\frac{S_1 + S_2 + \cdots + S_m}{m} \rightarrow ?$$

如果极限 $\lim S_n = S$ 存在，则算术平均和的极限也存在，同样等于 S ；反过来，如果极限 $\lim S_n = S$ 不存在，算术平均和的极限也可能存在。有些时候，根据具体问题，可能需要算术平均和的极限相关结论。比如，连续函数的 Fourier 级数不一定处处收敛，但其部分和的算术平均和（费耶和）却是一致收敛的。

接下来说一种十分重要和实用的变换思想，所谓 **RMI 原理**。即关系 (Relation)、映射 (Mapping)、反演 (Inversion) 原理。它是经过建立一种映射，把所研究的对象从一个系统结构 A 中映射到另一个系统结构 B 中去，利用新的系统中的知识，研究问题的解，然后再通过反演，得到原来问题的解答。

通俗地说，RMI 原理可以表示成下图



目标原解答包含在问题 A 中，但直接求解问题 A 非常困难。于是设法构造映射 φ ，将问题 A 转化为问题 B，后者相对容易处理。解答出问题 B 后再通过逆映射 φ^{-1} （反演变换），得到原解答。

在实际应用中，最大的问题往往在于构造映射 φ ，需要开动脑筋，挖掘创造力才行。RMI 原理最具代表性的例子是 Fourier 变换和 Laplace 变换。

以 Fourier 变换求解常微积分方程 $p(D)f = g$ 为例，这里 p 是一个常系数多项式， D 是微分算子，函数 $g = g(x)$ 是已知的，求未知函数 $f = f(x)$ 。

Fourier 变换的优点之一是把微分运算变为代数运算，后者在数学上更容易处理。将 Fourier 变换（构造映射 φ ）

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)e^{-i\omega x} dx$$

作用在常微分方程 $p(D)f = g$ 两边，得到

$$q(\omega)\hat{f}(\omega) = \hat{g}(\omega)$$

于是可以直接求解

$$\hat{f}(\omega) = \hat{g}(\omega) / q(\omega)$$

最后，等式两边作反演变换（Fourier 逆变换），得到 $f = f(x)$ 。

上面说得比较抽象，想具体了解该方法还是要找几个例子算一算，这方面读者可参考《数学物理方法》方面的教材。

最后说的这个例子是关于重积分变量代换公式的，内容摘自《数学分析新讲》第二册。这里重点说一下新讲中的证明思想，它将逼近、变换与分解这三大数学分析思想融为一体。

（重积分变量代换定理）设 $\Omega \subset \mathbf{R}^m$ 是一个开集，

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射， $E \subset \Omega$ 是一个闭若尔当可测集。如果

（1）雅可比行列式 $\det D\varphi(t) \neq 0, \forall t \in \text{int } E$ ；

（2） φ 在 $\text{int } E$ 中是单射。

那么 $\varphi(E)$ 也是一个闭若尔当可测集，并且对于任何在 $\varphi(E)$ 上连续的函数 $f(x)$ 都有

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt.$$

这个定理的证明有些复杂，张筑生老师把它拆成若干引理，每个引理相对

容易证明一些。前面几个引理主要关于可积性条件讨论的，不是本质部分，真正核心部分是从简单图形逼近开始，想法是“采用把局部微分同胚分解成简单微分同胚的复合，然后再转化为累次积分来证明”。

如果 \mathbf{R}^n 的子集 S 可以表示为有限个两两无公共内点的闭方块的并集，那么我们就说 S 是一个简单图形。任何简单图形当然都是闭若尔当可测集。新讲中证明了：为证明重积分变量代换公式，只需要考虑积分区域最简单的情形。即如下结论成立：

设 $\Omega \subset \mathbf{R}^m$ 是一个开集， $E \subset \Omega$ 是一个闭若尔当可测集，而

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射，且满足

- (1) 雅可比行列式 $\det D\varphi(t) \neq 0, \forall t \in \text{int } E$;
- (2) φ 在 $\text{int } E$ 中是单射。

如果函数 f 在集合 $\varphi(E)$ 上是连续，并且对任意闭方块 $\Pi \subset \text{int } E$ ，都有

$$\int_{\varphi(\Pi)} f(x) dx = \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt$$

那么就有

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt$$

上述定理是分析中“逼近思想”的充分体现。

接下来讨论所谓“简单变换”

设 $\varphi^h(t^1, \dots, t^m): \Omega \rightarrow \mathbf{R}$ 是一个连续可微函数，我们把如下形状的变换称为简单变换：

$$\begin{cases} x^i = t^i, i \neq h \\ x^h = \varphi^h(t^1, \dots, t^m) \end{cases}$$

换句话说，简单变换是这样一种连续可微映射

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

它（至多）只改变 $t = (t^1, \dots, t^m) \in \Omega$ 的一个坐标。

我们先对简单变换证明重积分变量代换公式成立。

定理： 设 $\Omega \subset \mathbf{R}^m$ 是一个开集， $E \subset \Omega$ 是一个闭若尔当可测集，而

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个简单变换，且满足

- (1) 雅可比行列式 $\det D\varphi(t) \neq 0, \forall t \in \text{int } E$;
- (2) φ 在 $\text{int } E$ 中是单射。

如果函数 f 在集合 $\varphi(E)$ 上是连续, 那么

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt$$

证明: 必要时给变量重新编号, 可设简单变换 $x = \varphi(t)$ 具有这样的形式

$$\begin{cases} x^i = t^i, & i = 1, 2, \dots, m-1 \\ x^m = \psi(t^1, \dots, t^{m-1}) \end{cases}$$

根据引理 5, 只须对任意闭方块 $\Pi \subset \text{int } E$, 证明公式

$$\int_{\varphi(\Pi)} f(x) dx = \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt.$$

不妨设在 Π 上有

$$\det D\varphi(t) = \frac{\partial \psi}{\partial t^m} > 0$$

(另一种情形可类似讨论)。把 Π 写成

$$\Pi = \Pi' \times \Pi'' \subset \mathbf{R}^{m-1} \times \mathbf{R}$$

这里

$$\Pi' = [\alpha^1, \beta^1] \times \dots \times [\alpha^{m-1}, \beta^{m-1}]$$

$$\Pi'' = [\alpha, \beta]$$

于是, 我们有

$$\varphi(\Pi) = \left\{ (x', x^m) \left| \begin{array}{l} x' \in \Pi' \\ \psi(x', \alpha) \leq x^m \leq \psi(x', \beta) \end{array} \right. \right\}$$

因而由累次积分公式, 得到

$$\begin{aligned} \int_{\varphi(\Pi)} f(x) dx &= \int_{\Pi'} dx' \int_{\psi(x', \alpha)}^{\psi(x', \beta)} f(x', x^m) dx^m \\ &= \int_{\Pi'} dx' \int_{\alpha}^{\beta} f(x', \psi(x', t^m)) \frac{\partial \psi(x', t^m)}{\partial t^m} dt^m \\ &= \int_{\Pi'} dt' \int_{\alpha}^{\beta} f(t', \psi(t', t^m)) \frac{\partial \psi(t', t^m)}{\partial t^m} dt^m \\ &= \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt \end{aligned}$$

于是我们对简单变换证明了重积分变量代换公式成立。

从上面可以看出简单积分区域和简单变换的妙处了吧, 即方便利用累次积分, 上述证明是变换与分解思想的结合。对于一般的连续可微变换, 我们把它分解成一些列简单变换的复合。需要注意的是, 这种分解通常只在局部情形成立, 整体分解一般不存在。这里采用的思想就是“把局部微分同胚分解成简单

微分同胚的复合”，这种方法体现了“分解思想”。

定理：设 $\Omega \subset \mathbf{R}^m$ 是一个开集，

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射， $\tau \in \Omega$ 。如果 $\det D\varphi(\tau) \neq 0$ ，那么存在 $\delta > 0$ ，使得重积分变量代换公式对于包含在邻域 $U_\rho(\tau, \delta)$ 之中的任何闭若尔当可测集成立。这就是说，对任何闭若尔当可测集 $E \subset U_\rho(\tau, \delta)$ 和任何在 $\varphi(E)$ 上连续的函数 f ，都有

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt.$$

证明：因为

$$\det D\varphi(\tau) = \frac{\partial(\varphi^1, \dots, \varphi^m)}{\partial(t^1, \dots, t^m)}(\tau) \neq 0$$

所以这行列式前 $m-1$ 行至少有一个不等于 0 的 $m-1$ 阶主子式。我们可以给变量 t^1, \dots, t^m 重新编号，使得 $\det D\varphi(\tau)$ 的 $m-1$ 阶主子式

$$\frac{\partial(\varphi^1, \dots, \varphi^{m-1})}{\partial(t^1, \dots, t^{m-1})}(\tau) \neq 0$$

仿此，用数学归纳法可以证明：适当地给变量 t^1, \dots, t^m 编号，可以使得 $\det D\varphi(\tau)$ 的各阶顺序主子式都不等于 0。在下面的讨论中，假定变量 t^1, \dots, t^m 已经按照这样的要求排列妥当。

我们定义 m 个变换（分解思想），

$$\theta_k: \begin{cases} x^i = \varphi^i(t^1, \dots, t^m), i \leq k \\ x^j = t^j, j > k \end{cases}$$

这里 $k = 1, 2, \dots, m$ 。

容易看出， $\det D\theta_k(\tau)$ 与 $\det D\varphi(\tau)$ 的第 k 个顺序主子式相等，因而

$$\det D\theta_k(\tau) \neq 0, k = 1, 2, \dots, m$$

我们可以选取 $\delta > 0$ 充分小，使得这 m 个变换 $\theta_1, \dots, \theta_m$ 在开集 $U_\rho(\tau, \delta)$ 之上都是微分同胚（这里用到了逆映射定理）。再令

$$\begin{cases} \psi_1 = \theta_1 \\ \psi_k = \theta_k \circ \theta_{k-1}^{-1}, k = 2, \dots, m \end{cases}$$

由此可知， $\psi_1 = \theta_1$ 是定义在开集 $U_\rho(\tau, \delta)$ 之上的简单变换， $\psi_k = \theta_k \circ \theta_{k-1}^{-1}$ 是

定义在 $\theta_{k-1}(U_\rho(\tau, \delta))$ 之上的简单变换, 并且在 $U_\rho(\tau, \delta)$ 之上有

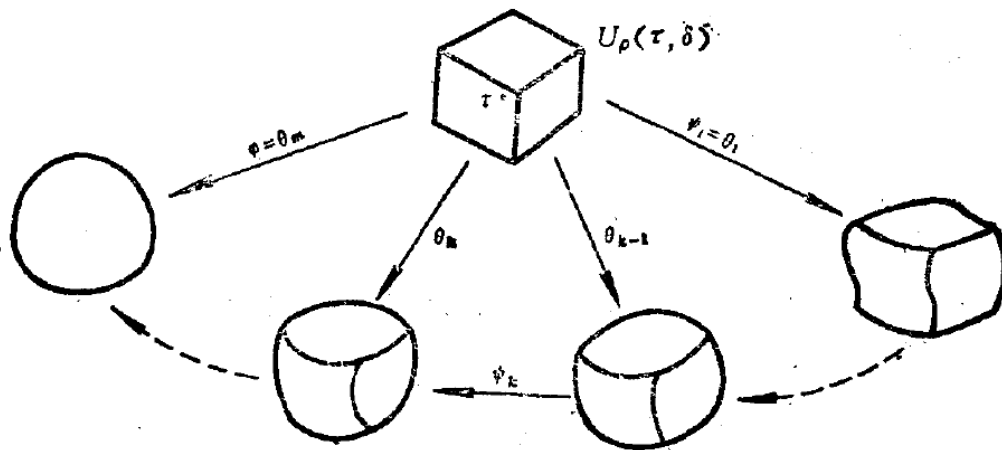
$$\varphi = \psi_m \circ \psi_{m-1} \circ \cdots \circ \psi_1$$

这样我们已将 φ 局部地分解为简单变换的复合, 在此基础上逐次应用引理 6 得到

$$\begin{aligned} \int_{\varphi(E)} f(x) dx &= \int_{\psi_m \circ \psi_{m-1} \circ \cdots \circ \psi_1(E)} f(x) dx \\ &= \int_{\psi_{m-1} \circ \cdots \circ \psi_1(E)} f(\psi_m(u)) |\det D\psi_m(u)| du \\ &= \cdots \\ &= \int_E f(\varphi(t)) |\det D\varphi(t)| dt \end{aligned}$$

这样我们就证明了结论。

一般地, 整体微分同胚是很强的条件, 通常难以满足, 而局部微分同胚相对容易满足。这里我们证明了重积分变量代换公式是局部成立的, 为了得到整体性结论, 只需利用紧致性 (有限覆盖) 与积分的有限可加性即可, 证明细节见本书附录。



简单变换分解 $\varphi = \psi_m \circ \psi_{m-1} \circ \cdots \circ \psi_1$ 示意图

(逆映射定理) 设 $\Omega \subset \mathbf{R}^m$ 是一个开集, $\varphi: \Omega \rightarrow \mathbf{R}^m$ 是一个连续可微映射, $a \in \Omega$. 如果映射 φ 在 a 点对应的雅可比矩阵行列式满足

$$\det D\varphi(a) \neq 0$$

那么 φ 在 a 点是局部微分同胚的。也就是存在包含 a 点的邻域 U 和包含 $b = \varphi(a)$ 点的邻域 V , 使得映射 $\varphi: U \rightarrow V$ 是微分同胚。

评注：最后笔者说一点题外话，为什么一再强调要读一些大定理（比如重积分变量代换定理）。大定理通常比较难，证明也比较长，通常考试不会考，课堂上可能也没时间讲。但站在学习的角度看，大定理的证明往往涉及多方面知识的综合运用，就像本文举了一堆例子，但还是重积分变量代换定理的证明最能体现逼近、变换与分解三大思想的巧妙融合。一般的定理或习题，很难体现知识的交叉运用，而交叉综合运用恰恰是数学分析思想的精华所在。在科研中，局部战是少见的，打综合战和整体战确实很常见的，而且有时候甚至事先不清楚到底哪种方法或工具能解决问题。很多导师鼓励学生数学阅读面尽可能广泛一些，也是有良苦用心的，眼光开阔一些总是有好处的。

第4章 数学分析三大基本思想之分解

本章介绍数学分析中的三大基本思想之分解。需要强调的是，逼近、变换和分解这三大分析基本思想是统一的，在处理数学问题时常常是综合运用。

笔者将分解看作这样一种数学思想：将一个复杂的结构或问题，分解成若干子结构，使得这些子结构尽可能简单。若按照广义理解，从一个复杂问题中分离出主要矛盾，这也是一种分解思想。

前者的例子可以考虑幂级数分解，如考虑超越函数

$$e^z = 1 + z + \frac{1}{2!}z^2 + \cdots + \frac{1}{n!}z^n + \cdots$$

如此将一个复杂的超越函数 e^z 分解成幂级数形式，而幂级数可以看做多项式的推广，而我们对子结构 z^n 非常熟悉（可类比整数的加减乘除运算）。实际上，幂级数是研究超越函数的有力工具。当然我们也可以使用 Fourier 级数，或者其它函数级数。

后者的例子可以考虑微分，考虑函数增量 $\Delta y = f(x + \Delta x) - f(x)$ ，数学上证明了对可导函数，有

$$\Delta y = f'(x)\Delta x + o(\Delta x)$$

Δy 的线性主部记为 $dy = f'(x)\Delta x$ （微分）， dy 对 Δy 的贡献较大，属于主要矛盾。如此，我们从一个非线性增量中，分离出线性部分，达到简化问题的目的。这个例子很平凡，但却很实用，特别在物理分析中（如微元法）。很多宏观物理过程在短时间内或小尺度范围内变化很小，可以近似为线性过程。

数学上已经证明了非初等函数是大多数，而研究非初等函数的常见途径有三个，即积分法（如含参数积分）、微分方程法和级数法（特别是幂级数法）。从实用角度看，级数法更方便，比如利用幂级数进行数值计算。

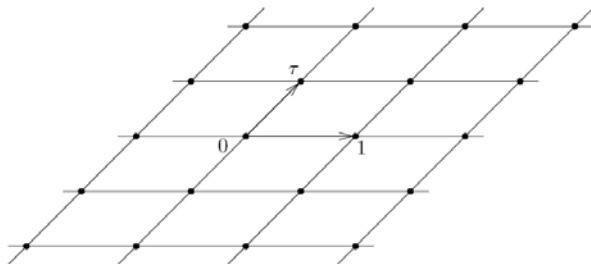
用级数研究非初等函数的著名例子是椭圆函数。当年高斯和阿贝尔提出了研究椭圆积分的反函数，后称椭圆函数，这种函数有两个不共线的复周期 ω_1, ω_2 满足 $f(z + \omega_1) = f(z + \omega_2) = f(z)$ 。与历史不同，现在教材上直接定义具有不共线的复周期 ω_1, ω_2 的亚纯函数为椭圆函数。

那么能不能不利用积分反演，直接构造双周期的椭圆函数。直观上不难想象，

椭圆函数的所有周期是复平面上一些格子（周期格）

$$\Lambda = \{m\omega_1 + n\omega_2 \mid m, n \in \mathbf{Z}\}$$

其中复数 ω_1 和 ω_2 是实线性无关的。



那么形式上，如下函数满足双周期性：

$$\sum_{\omega \in \Lambda} \varphi(z - \omega) \text{ 或 } \prod_{\omega \in \Lambda} \varphi(z - \omega)$$

这里函数 φ 待定。

之所以强调是形式上，是因为要使得上面的函数有意义，还要考虑收敛性问题。椭圆函数中最有名的要数魏尔斯特拉斯椭圆函数：

$$\wp(z) = \frac{1}{z^2} + \sum_{0 \neq \omega \in \Lambda} \left[\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right]$$

据说，这种形式的构造类比了余切级数展开（欧拉等式）

$$\pi \cot \pi z = \frac{1}{z} + \sum_{n=1}^{\infty} \left(\frac{1}{z+n} + \frac{1}{z-n} \right)$$

注：数学上可以证明非常数椭圆函数在一个周期四边形内必有奇点，且阶数至少是二（证明见《特殊函数概论》）， $\wp(z)$ 的构造选择了一个二阶极点的情形（不妨设 $z=0$ 为二阶极点）。于是根据椭圆函数周期性要求，可知 $z=\omega$ 为二重极点。再由上面形式级数的思想，考察下列函数就自然了：

$$\sum_{0 \in \Lambda} \frac{1}{(z - \omega)^2}$$

但是这个级数不收敛，所以为满足收敛性还得再减去一项 $1/\omega^2$ 。当 $\omega=0$ 这种情况下， $1/\omega^2$ 发散，所以应该剔除，同时把 $1/z^2$ 单独写出来。于是级数

$$\wp(z) = \frac{1}{z^2} + \sum_{0 \neq \omega \in \Lambda} \left[\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right]$$

是收敛的，具有二阶极点，且满足 $\wp(z + \omega_1) = \wp(z + \omega_2) = \wp(z)$ 。

$\wp(z)$ 的导数 $\wp'(z)$ 也是椭圆函数，它们之间满足上面的微分方程

$$\wp'(z)^2 = 4\wp(z)^3 + g_2\wp(z) + g_3$$

$\wp(z)$ 和 $\wp'(z)$ 也所以非常重要, 因为它们不仅形式简单, 还可以作为“基函数”, 这种基函数分解思想 (当然不是线性的) 可以由下述定理描述:

设 $\wp(z)$ 的周期为 ω_1, ω_2 , 则所有以 ω_1, ω_2 为双周期的椭圆函数

$$f(z + \omega_1) = f(z + \omega_2) = f(z)$$

都可以表示成下列形式

$$f(z) = H(\wp(z)) + \wp'(z)Q(\wp(z))$$

其中, $H(z), Q(z)$ 为有理函数, 与 $f(z)$ 有关。

有了这个定理, 研究椭圆函数的重点可以放在研究基函数 $\wp(z)$ 和 $\wp'(z)$ 上, 这是利用分解思想, 缩小问题范围的经典例子。

分解思想在数学中非常普遍 (不仅仅是数学分析), 我们在证明一个复杂数学定理时, 常常分成若干引理 (如前面章节多次提到的重积分变量代换定理), 这其实应用了分解思想。学会如何把一个复杂问题简化, 拆成若干简单问题, 这是一门真功夫, 需要能力和经验的积累。

分解思想的重要性还体现在可操作性上。数学上很多定义逻辑上很抽象, 也很完备, 但单靠定义是无法操作的。比如重积分的 Riemann 和定义, 如果靠重积分的 Riemann 和定义来计算, 甚至证明数学定理, 恐怕举步维艰。累次积分就应运而生

$$\iint_{X \times Y} f(x, y) dx dy = \int_X dx \int_Y f(x, y) dy$$

累次积分的妙处还在于“给无序的定义引入一种有序的计算”。在无序中引入顺序或偏序, 这种思想的妙用不仅仅在数学中存在, 大家需要好好体会。本书第三章证明重积分变量代换定理的关键地方就利用了累次积分:

$$\begin{aligned} \int_{\varphi(\Pi)} f(x) dx &= \int_{\Pi'} dx' \int_{\psi(x', \alpha)}^{\psi(x', \beta)} f(x', x^m) dx^m \\ &= \int_{\Pi'} dx' \int_{\alpha}^{\beta} f(x', \psi(x', t^m)) \frac{\partial \psi(x', t^m)}{\partial t^m} dt^m \\ &= \int_{\Pi'} dt' \int_{\alpha}^{\beta} f(t', \psi(t', t^m)) \frac{\partial \psi(t', t^m)}{\partial t^m} dt^m \\ &= \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt \end{aligned}$$

广义上理解, 递归关系或者递推关系 $a_{n+1} = f(a_n)$ 也可以看做是一种分解, 即在同一法则 f 下形成一种逐级发展的“自相似结构”。比如我们计算积分

$$a_n = \int_0^{\frac{\pi}{2}} \sin^n x \, dx$$

时, 就设法建立 a_n 与 a_{n-2} 的递归关系。利用分部积分

$$\begin{aligned} a_n &= \int_0^{\frac{\pi}{2}} \sin^n x \, dx = \int_0^{\frac{\pi}{2}} \sin^{n-1} x \, d(-\cos x) = -\cos x \sin^{n-1} x \Big|_0^{\frac{\pi}{2}} + \int_0^{\frac{\pi}{2}} \cos x \, d(\sin^{n-1} x) \\ &= (n-1) \int_0^{\frac{\pi}{2}} \sin^{n-2} x \cos^2 x \, dx = (n-1)a_{n-2} - (n-1)a_n \end{aligned}$$

可知递推关系

$$a_n = \frac{n-1}{n} a_{n-2}$$

容易算出 $a_0 = \frac{\pi}{2}$, $a_1 = 1$, 于是

$$a_n = \begin{cases} \frac{(n-1)!!}{n!!}, & n = 2k \\ \frac{(n-1)!!}{n!!} \cdot \frac{\pi}{2}, & n = 2k+1 \end{cases}$$

建立递归关系是一种可操作性非常强的分解思想, 不论是在解题或者是科研中, 非常实用, 大家应该熟练掌握。

说到递归关系, 有一个非常经典的例子, 将逼近与递归联系到一起。这就是求解非线性方程的经典解法---牛顿迭代法。

非线性问题非常复杂, 直接求解基本不可能, 于是人们发展了一些实用性较高的近似方法来求解非线性问题。如压缩映射不动点法和牛顿迭代法(又称切线法), 我们主要谈谈后者。

设函数 $f(x)$ 在 $[a, b]$ 上连续可微, 并且满足

$$f(a)f(b) < 0, \quad f'(x) \neq 0, \quad \forall x \in [a, b]$$

根据函数零点定理(介值定理), 可知 $f(x) = 0$ 在区间 (a, b) 内至少有一根 ξ 。根据 $f'(x) \neq 0$ 可知导数在区间内不变号(根据导数介值定理), 因而是严格单调的。所以可知 ξ 是惟一的。

那么如何计算零点 ξ ? 牛顿想到了一种切线方法, 该方法的具有非常强的几何直观。切线法的思想: 设曲线 $y = f(x)$ 在一点 x_0 处的切线方程为

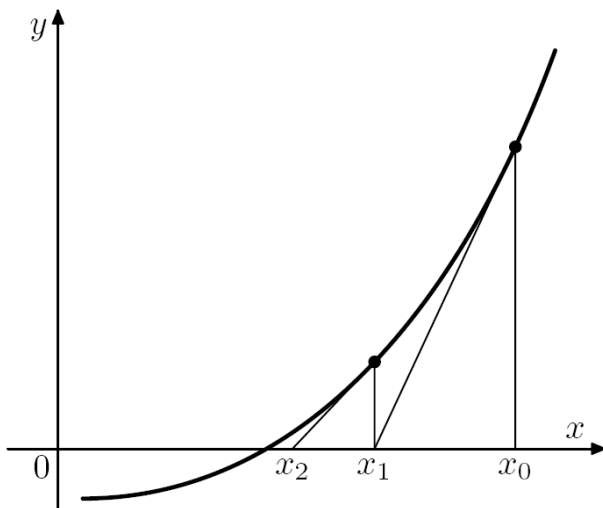
$$y = f(x_0) + f'(x_0)(x - x_0)$$

为了近似求解方程 $f(x) = 0$ ，这里用上述切线与 x 轴的交点近似代替零点 ξ 。

换句话说，就是用方程

$$f(x_0) + f'(x_0)(x - x_0) = 0$$

的解近似代替方程 $f(x) = 0$ 的解。



牛顿迭代法示意图

上面我们用了逼近思想和变换思想，即局部上用切线近似代替曲线，用线性变换代替非线性变换。

解方程 $f(x_0) + f'(x_0)(x - x_0) = 0$ ，得到

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

以 x_1 代替 x_0 点，交 x 轴于点 x_2 ，于是

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

如此反复迭代，得到一个迭代序列 $\{x_n\}$ （建立递归关系）

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}, \quad n = 1, 2, \dots$$

如果此序列存在极限 $x_n \rightarrow c$ ，对上式两边取极限，得到

$$c = c - \frac{f(c)}{f'(c)}$$

由此得到 $f(c) = 0$ ，再由惟一性可知 $\xi = c$ 。

剩下的问题是如何保证此递归序列存在极限 $x_n \rightarrow c$ ，通常要加上一些附加

条件（不唯一），这方面一个比较实用的定理如下：

（定理）设函数 $f(x)$ 在 $[a, b]$ 上二阶连续可微，并且满足条件

$$f(a)f(b) < 0, f'(x)f''(x) \neq 0, \forall x \in [a, b]$$

如果可以选取初值 x_0 满足

$$f(x_0)f''(x_0) > 0$$

那么由迭代序列

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}, \quad n = 1, 2, \dots$$

产生的数列 $\{x_n\}$ 单调收敛于方程 $f(x) = 0$ 在区间 (a, b) 内的惟一解 ξ 。

注：上述定理的证明请参考《数学分析新讲》第二册。

关于利用分解思想研究复杂问题的例子，还有一种比上面分析味道更浓的例子，就是用幂级数法解微分方程。这方面有一个很实用的定理：

设二阶常微分方程

$$\frac{d^2 y}{dx^2} + p(x) \frac{dy}{dx} + q(x)y = 0$$

其系数函数 $p(x)$ 和 $q(x)$ 在邻域 $|x - x_0| < r$ 内是单值解析的，则在 $|x - x_0| < r$ 内存在惟一的一个解 $y = y(x)$ 满足： $y(0) = a_0, y'(0) = a_1$ 且

$$y(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n$$

作为应用，考虑下面的例子： $y'' - 2xy' - 4y = 0, y(0) = 0, y'(0) = 1$ 。
在 $x_0 = 0$ 点展开成幂级数，令

$$y(x) = \sum_{n=0}^{\infty} a_n x^n$$

其收敛域为 $(-\infty, +\infty)$ 。将幂级数解逐项求导并代回原方程，得到

$$\sum_{n=0}^{\infty} (n+1)(n+2)a_{n+2}x^n - \sum_{n=1}^{\infty} 2na_n x^n - \sum_{n=0}^{\infty} 4a_n x^n = 0$$

即

$$(2a_2 - 4a_0) + \sum_{n=1}^{\infty} [(n+2)(n+1)a_{n+2} - 2(n+2)a_n] x^n = 0$$

上面的幂级数恒等于 0，于是幂级数系数均为 0，由此得到递推关系

$$\begin{cases} 2a_2 - 4a_0 = 0 \\ (n+2)(n+1)a_{n+2} - 2(n+2)a_n = 0 \end{cases}$$

其中 $n=1, 2, 3, \dots$

从而得到递推数列 $a_{n+2} = \frac{2}{n+1}a_n$, $n=1, 2, 3, \dots$

显然, 我们只要求出数列 a_n 的通项公式即可。

因为 $y(0) = a_0 = 0$, $y'(0) = a_1 = 1$, 所以利用递推关系得到

$$a_2 = a_4 = \dots = a_{2n} = 0, \quad n=1, 2, 3, \dots$$

于是只需要考虑 a_{2n+1} 的通项公式。此时递推方程为

$$a_{2n+1} = \frac{1}{n}a_{2n-1}, \quad n=1, 2, 3, \dots$$

利用递推公式依次计算得到

$$\begin{aligned} a_{2n+1} &= \frac{1}{n}a_{2n-1} \\ a_{2n-1} &= \frac{1}{n-1}a_{2n-3} \\ &\dots \quad \dots \\ a_5 &= \frac{1}{2}a_3 \\ a_3 &= \frac{1}{1}a_1 \end{aligned}$$

以上各等式相乘, 并利用 $a_1 = 1$, 得到

$$a_{2n+1} = \frac{1}{n!}, \quad n=0, 1, 2, 3, \dots$$

将 a_n 通项公式代入幂级数, 得到

$$y = \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_{2n+1} x^{2n+1} = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{n!} = x \sum_{n=0}^{\infty} \frac{(x^2)^n}{n!} = x e^{x^2}, \quad x \in (-\infty, +\infty)$$

不难验证其满足原方程及初始条件, 所以 $y(x) = x e^{x^2}$ 为所求函数。

一般来说, 分析给人的印象是涉及连续数学的, 而级数兼具连续和离散两种特征, 上面的例子就是把一个连续问题, 转化为离散问题。反过来也可以用连续数学的方法求解离散数学问题, 代表方法是组合数学中的“母函数法”(也称为生成函数法)。

其想法是, 对于一个数列 a_0, a_1, a_2, \dots , 构造一个形式幂级数

$$S(x) = a_0 + a_1x + a_2x^2 + \cdots$$

数列 a_0, a_1, a_2, \cdots 之间通常满足某些关系 (如递推关系), 以此建立 $S(x)$ 的关系式, 然后再做泰勒级数展开, 其系数就是 a_n .

还是用例子来说明: 考虑传说中的斐波那契数列

$$1, 1, 2, 3, 5, 8, \cdots$$

写成递推公式: $a_0 = a_1 = 1, a_{n+2} = a_{n+1} + a_n$. 这问题标准的解法是特征方程, 不过借助分析方法更具特色.

设 $S(x) = a_0 + a_1x + a_2x^2 + \cdots$, 再利用 $a_0 = a_1 = 1, a_{n+2} = a_{n+1} + a_n$ 可以建立级数之间的等式 $(1 - x - x^2)S(x) = 1$ (建议读者动手算一算), 于是

$$S(x) = \frac{1}{1 - x - x^2}$$

把上面的有理函数展开成泰勒级数, 求出其系数

$$a_n = \frac{1}{\sqrt{5}} \left\{ \left(\frac{1 + \sqrt{5}}{2} \right)^{n+1} - \left(\frac{1 - \sqrt{5}}{2} \right)^{n+1} \right\}, \quad n = 0, 1, 2, \cdots$$

这就是斐波那契额通项公式. 想进一步了解母函数方法的读者可参考组合数学方面的教材.

除了幂级数之外, **Fourier** 级数分解也能解决一些棘手的问题. 设 $f(x)$ 是 $[-\pi, \pi]$ 上的连续可微函数, 满足 $f(\pi) = f(-\pi)$. 如果

$$\int_{-\pi}^{\pi} f(x) dx = 0$$

证明不等式

$$\int_{-\pi}^{\pi} f^2(x) dx \leq \int_{-\pi}^{\pi} [f'(x)]^2 dx$$

等号成立当且仅当 $f(x) = a \sin x + b \cos x$.

证明方法是先将 $f(x)$ 进行周期延拓, 再考虑 **Fourier** 级数分解

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

根据题意, 可知 $a_0 = 0$. 导函数 $f'(x)$ 的 **Fourier** 级数为

$$f'(x) \sim \sum_{n=1}^{\infty} n b_n \cos nx - n a_n \sin nx$$

由帕塞瓦尔等式, 可知

$$\int_{-\pi}^{\pi} f^2(x) dx = \pi \sum_{n=1}^{\infty} (a_n^2 + b_n^2) \leq \pi \sum_{n=1}^{\infty} (n^2 a_n^2 + n^2 b_n^2) = \int_{-\pi}^{\pi} [f'(x)]^2 dx$$

等号成立当且仅当 $n \geq 2$ 时, $a_n = b_n = 0$, 即 $f(x) = a \sin x + b \cos x$. 这个不等式叫做 Wirtinger 不等式。

最后再次说说《数学分析新讲》第二册给出的重积分变量代换公式

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt$$

的证明。之所以反复说, 是因为这个例子太经典了! 我们知道, 大多数坐标变换 φ 都是非线性的, 新讲中的处理方法是考虑“简单变换”:

$$\begin{cases} x^i = t^i, i \neq h \\ x^h = \varphi^h(t^1, \dots, t^m) \end{cases}$$

换句话说, 简单变换是这样一种连续可微映射

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

它(至多)只改变 $t = (t^1, \dots, t^m) \in \Omega$ 的一个坐标。

简单变换至少比一般的变换 φ 要简单。再考虑下面的变换 m 个变换(分解思想),

$$\theta_k: \begin{cases} x^i = \varphi^i(t^1, \dots, t^m), i \leq k \\ x^j = t^j, j > k \end{cases}$$

这里 $k = 1, 2, \dots, m$. 使得这 m 个变换 $\theta_1, \dots, \theta_m$ 在小的开集 $U_\rho(\tau, \delta)$ 之上都是微分同胚(这里用到了逆映射定理)。再令

$$\begin{cases} \psi_1 = \theta_1 \\ \psi_k = \theta_k \circ \theta_{k-1}^{-1}, k = 2, \dots, m \end{cases}$$

由此可知, $\psi_1 = \theta_1$ 是定义在开集 $U_\rho(\tau, \delta)$ 之上的简单变换, $\psi_k = \theta_k \circ \theta_{k-1}^{-1}$ 是定义在 $\theta_{k-1}(U_\rho(\tau, \delta))$ 之上的简单变换, 并且在 $U_\rho(\tau, \delta)$ 之上有

$$\varphi = \psi_m \circ \psi_{m-1} \circ \dots \circ \psi_1$$

这种分解可以类比线性代数中, 将一个非奇异方阵表示成若干初等矩阵的乘积思想。都是将一个复杂对象, 分解成若干简单对象的复合, 这也是人类认识世界的一种自然方法。

评注: 微分、积分和级数三大领域是数学分析的基石, 逼近和变换这两章的大部分内容都涉及微分和积分, 为了不厚此薄彼, 本章大部分内容与级数有关。级数是一门强大的数学工具, 不论是牛顿、高斯、雅可比、魏尔斯特拉斯, 还

是庞加莱，都是运用级数的大师。级数理论的综合运用是“分解”思想的最佳例子，同时最好结合一些线性代数的分解思想。

说到分解思想，笔者想起了高中时一件小事。那时笔者已经初步掌握了单元微分法和积分法，偶然在一篇文章中看到“**矢量也可以积分**”的内容，感到无比诧异。习惯了代数思维的自己，很难想象矢量如何积分。后来，隔了一年多才想明白，其实也简单。矢量可以投影到三维坐标轴上，得到三个分量。然后用矢量分量的坐标表示，从而将一个矢量积分转化为三个代数积分。多年后回想这件事情仍然很有意思，困惑不在于技术上，而在于理念或想法的转变上。这个例子本质上是将“**几何思维转化为代数思维**”，这种转化远不像字面上那样容易。

第5章 抓住问题的核心

本章原本打算以“抓住问题的本质”为标题的，但再三斟酌之后觉得不妥，还是换成“抓住问题的核心”这个标题吧。先看一个问题：

$$\frac{1}{2} + \cos x + \cos 2x + \cdots + \cos nx = ?$$

这个问题常见于高一的三角函数章节。课本上的证明是一个小技巧，即用 $\sin \frac{1}{2}x$ 乘以上式左侧，然后利用积化和差公式，得到类似

$$a + (b-a) + (c-b) + \cdots$$

这样的关系，最终剩下 $\frac{1}{2}\sin(n+\frac{1}{2})x$ 这一项，所以

$$\frac{1}{2} + \cos x + \cos 2x + \cdots + \cos nx = \frac{\sin(n+\frac{1}{2})x}{2\sin\frac{1}{2}x}$$

这个问题可比 $1+2+\cdots+100=5050$ 难多了，当年自己没做出来，所以印象深刻。后来问数学老师，这么巧妙的乘以 $\sin \frac{1}{2}x$ 裂项相消技巧是如何想出的，得到的回复是显然。当时自己觉得数学直觉比较差，经常是大家都认为显然，自己却感觉不那么显然。这个三角函数求和问题一直在我心里还是纳闷：到底是巧合，还是有更深刻的原因呢。

从证明来说，乘以 $\sin \frac{1}{2}x$ 这个因子是关键，或者是本题的核心。但自己总觉得其背后应该有更深刻的原因。后来，笔者在一本课外书上看到一个提示，

$$e^{inx} = \cos nx + i \sin nx$$

这是欧拉公式，自己高一时就知晓。但关键的地方是书中提示令 $q = e^{ix}$ ，则

$$q + q^2 + \cdots + q^n = \frac{q(1-q^n)}{1-q}$$

是可以求和。自己当时忽然意识到，将 $q^n = e^{inx} = \cos nx + i \sin nx$ 代入和式右侧，再分离出实部，不就是求出来 $\cos x + \cos 2x + \cdots + \cos nx$ 了吗。当初的困惑豁然开朗，原来三角函数求和的本质是等比级数求和啊（借助欧拉公式）！意识到这一点，才抓住了这个问题的真正核心！在数学分析中，研究 Fourier 级数的收敛性时，自然遇到狄利克雷积分，到时再次遇到这个求和问题。

上面的例子是一个引子，意在说明问题的真正核心不是乘以 $\sin \frac{1}{2}x$ 这个因子，而是等比级数求和（借助欧拉公式）。认识到这一点才算对原问题有一个质的飞跃。

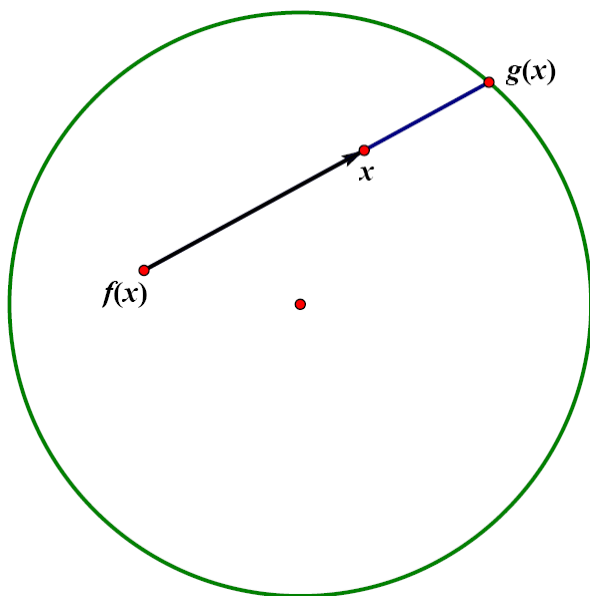
当然，本书主要讨论数学分析内容，所以不能泛泛而谈本质论。比如，虽然可以用非常漂亮的分析方法来证明布劳威尔不动点定理，但是要想讨论不动点定理的本质，肯定要将讨论框架扩展到拓扑领域。尽管如此，这个漂亮分析证明仍然值得推荐（引自《数学分析新讲》第三册）。

布劳威尔不动点定理：

设 D^n 是 n 维单位闭球， $f: D^n \rightarrow D^n$ 是连续映射，则存在 $x_0 \in D^n$ ，满足 $f(x_0) = x_0$ 。

在布劳威尔不动点定理的条件中，单位闭球 D^n 的凸性非常重要。比如把 D^n 换成单位球面 S^n ，则结论不再成立。反例是映射 $f(x) = -x$ ，显然映射 $f: S^n \rightarrow S^n$ 是连续映射，但没有不动点。

《新讲》给出的布劳威尔不动点定理的解析证明核心思想在于下面这张图



如果布劳威尔不动点定理不成立，那么可以做出一个映射 $g: D^n \rightarrow S^{n-1}$ 满足 $g(x) = x$ ， $x \in S^{n-1}$ 。这样的连续映射 g 称为收缩映射， S^{n-1} 称为 D^n 的收缩核。但是，可以证明这样的收缩核实际上不存在，从而不动点定理为真。

为利用分析工具，先假设 f 为光滑映射。反证假设 f 没有不动点，于是构造出光滑的收缩映射 $g: D^n \rightarrow S^{n-1}$ 。接着利用外微分中的 Stokes 公式证明这个光滑收缩映射 g 是不存在的。证明方法是计算两次思想，即对同一个对象用两种不同的方法计算，结果应该是一样的。否则，即产生矛盾。

设 $g = (g_1, g_2, \dots, g_n)$ ，构造微分形式 $\omega = g_1 dg_2 \wedge \dots \wedge dg_n$ 。
一方面， $g(x) = x \in S^{n-1} = \partial D^n$ ，所以 $\|g(x)\| = 1$ ， $\forall x \in S^{n-1}$ ，从而

$$\begin{aligned}\int_{\partial D^n} \omega &= \int_{D^n} d\omega = \int_{D^n} d(g_1 dg_2 \wedge \cdots \wedge dg_n) \\ &= \int_{D^n} \frac{\partial(g_1, \cdots, g_n)}{\partial(x_1, \cdots, x_n)} dx_1 \wedge \cdots \wedge dx_n = 0\end{aligned}$$

另一方面, 由 $g(x) = x$, $x \in S^{n-1}$, 直接代入, 得到

$$\int_{\partial D^n} \omega = \int_{D^n} d\omega = \int_{D^n} dx_1 \wedge \cdots \wedge dx_n = \text{vol}(D^n) > 0$$

这就证明了不存在光滑的收缩映射 $g: D^n \rightarrow S^{n-1}$.

于是证明了当 f 是光滑映射时, 存在不动点. 对于连续映射, 可以选择一个光滑映射序列来一致逼近. 假设连续映射 $f: D^n \rightarrow D^n$ 没有不动点.

因为单位闭球是有界闭集 (紧致性), 所以连续函数 $\|f(x) - x\|$ 在 D^n 上一定取得正的最小值 $\mu = \min \|f(x) - x\| > 0$. 取定 $\varepsilon > 0$ 满足

$$0 < 3\varepsilon < \mu = \min \|f(x) - x\|$$

根据魏尔斯特拉斯逼近定理, 存在多项式向量函数 $p(x)$, 使得

$$\|f(x) - p(x)\| < \varepsilon$$

不过对于 $x \in D^n$ 可能有 $p(x) \notin D^n$ (映射到单位闭球外面). 但可以判定

$$\|p(x)\| \leq \|p(x) - f(x)\| + \|f(x)\| < 1 + \varepsilon$$

于是令

$$h(x) = \frac{1}{1+\varepsilon} p(x)$$

则得到光滑映射 $h: D^n \rightarrow D^n$.

再次利用含绝对值不等式, 得到

$$\begin{aligned}\|h(x) - f(x)\| &= \left\| \frac{1}{1+\varepsilon} p(x) - f(x) \right\| = \frac{1}{1+\varepsilon} \|p(x) - f(x) - \varepsilon f(x)\| \\ &\leq \frac{1}{1+\varepsilon} \|p(x) - f(x)\| + \frac{\varepsilon}{1+\varepsilon} \|f(x)\| < \frac{2\varepsilon}{1+\varepsilon} < 2\varepsilon\end{aligned}$$

由此得到估计

$$\|h(x) - x\| \geq \|x - f(x)\| - \|h(x) - f(x)\| \geq 3\varepsilon - 2\varepsilon = \varepsilon > 0$$

但是根据前面的结论, 光滑映射 h 具有不动点 $h(x) = x$, 矛盾! 定理成立.

注: 《新讲》的证明是非常严格的, 必须证明利用射线做出的映射 $g: D^n \rightarrow S^{n-1}$ 确实是光滑映射.

根据经验, 数学分析中很多结论往往要在后续高级课程(如复变和实变)中才能看得清问题关键所在。比如 **Riemann** 积分中的控制收敛定理(也称阿尔泽拉定理)

设闭区间 $[a, b]$ 上的 \mathbf{R} 可积函数序列 $\{f_n(x)\}$ 点点收敛于某个 \mathbf{R} 可积函数 $f(x)$, 即 $f_n(x) \rightarrow f(x)$ 。如果函数序列 $f_n(x)$ 在 $[a, b]$ 上一致有界, 即存在 $M > 0$, 使得

$$|f_n(x)| \leq M, \forall n, \forall x \in [a, b]$$

恒成立。则有积分关系(积分运算与极限运算可交换)

$$\lim_{n \rightarrow \infty} \int_a^b f_n(x) dx = \int_a^b f(x) dx \quad (\text{即 } \lim_{n \rightarrow \infty} \circ \int_a^b = \int_a^b \circ \lim_{n \rightarrow \infty})$$

根据数学大师 **Lebesgue** 的工作, 这一定理可以推广到 \mathbf{L} 积分中去, 一般称为 **Lebesgue** 控制收敛定理, 是实变函数课程中最基本的定理了。

(**Lebesgue**) 设有限可测集 E 上的可测函数序列 $f_n(x)$ 满足

$$|f_n(x)| \leq M, \forall x \in E, \forall n \in \mathbf{N}$$

其中 $M > 0$ 为常数。如果 $\lim f_n(x) = f(x)$ 几乎处处成立, 那么 $f(x)$ 也是 \mathbf{L} 可积的, 且可以在积分号下求极限:

$$\lim_{n \rightarrow \infty} \int_E f_n(x) dx = \int_E f(x) dx$$

因为可测函数的极限仍然可测, 所以不必像阿尔泽拉定理那样额外要求极限函数是可积的。控制收敛定理的关键在于“有限可测集 E ”的结构和性态特征, 这在数学分析课程中, 我们难以发现。但是, 在实变函数课程中有如下深刻的叶果洛夫定理:

设可测集 E 满足 $m(E) < \infty$, 可测函数序列 f_n 在 E 上几乎处处收敛于一个有限的可测函数 f , 则 $\forall \delta > 0$, 存在子集 $E_\delta \subset E$ 满足 $m(E_\delta) < \delta$, 且函数序列 f_n 在集合 $E - E_\delta$ 上一致收敛于可测函数 f 。

有限可测集上的叶果洛夫定理深刻揭示了“点点收敛与一致收敛相差不多”的特征, 站在这个角度看控制收敛定理就自然了。我们可选取选取可测集 $A \subset E$ 使得 $m(E - A)$ 充分小, 且 $f_n(x)$ 在 A 上是一致收敛的。于是

$$\begin{aligned} \left| \int_E f_n(x) dx - \int_E f(x) dx \right| &\leq \int_E |f_n(x) - f(x)| dx, \\ &\leq \int_A |f_n(x) - f(x)| dx + \int_{E-A} |f_n(x) - f(x)| dx \\ &\leq \int_A |f_n(x) - f(x)| dx + \int_{E-A} 2M dx < \varepsilon \end{aligned}$$

这里不等式右边第一项利用了一致收敛性，第二项利用了测度 $m(E-A)$ 可以任意小，从而得到了放缩关系。上面的放缩过程可以严格写成 $\varepsilon-\delta$ 语言，于是由叶果洛夫定理证明了控制收敛定理。

初学数分的人必须严格套用 $\varepsilon-\delta$ 语言格式很长时间，但对高手来说，还纠缠于 $\varepsilon-\delta$ 语言的话，那就什么也不用干了。这就好比一个人学写律诗，初学必须严格按照律诗格式来写，一丝不苟地套用，甚至有些机械。但到了一定境界后，自由发挥的空间就大了，甚至挥洒自如。新手喜欢全面与具体，高手喜欢简洁与核心，这是不同的思维层次造成的结果。每一个学数学的人，最终都要走到抓问题核心的道路上来（特别是当导师的）。

其实我们在逼近、变换和分解这三章中也主动点出了问题的核心。比如利用核函数方法证明魏尔斯特拉斯多项式逼近连续函数定理，其关键在于核函数的狄拉克序列性质，而《新讲》中给出的重积分变量代换公式证明的关键在于“把局部微分同胚分解成简单微分同胚的复合，然后再转化为累次积分来证明”等等。

最后，我们再来谈谈一个难一点的例子，它是关于形式微分算子法的。考虑常微分方程： $y' = f(x), y(0) = 0$ 。直接利用牛顿莱布尼茨积分公式，得

$$y(x) = \int_0^x f(t)dt$$

英国工程师海维塞德思考上述方法后，提出了一个形式微分算子法，定义算子 $D = \frac{d}{dx}$ ，则微分方程可写成 $Dy = f(x)$ ，于是移项得： $y = \frac{1}{D}f(x)$

对比上面的积分过程可知 $\frac{1}{D} = \int_0^x$ （积分算子），以及 $\frac{1}{D^2} = \frac{1}{D} \frac{1}{D} = \int_0^x \int_0^x$ 等。

海维塞德将这个思想应用到一般的常系数微分方程中去，考虑方程

$$P(D)y = f(x), y(0) = y'(0) = y''(0) = \cdots = y^{(n-1)}(0) = 0$$

这里 $P(x)$ 是一个 n 次多项式。于是得到形式解

$$y = \frac{1}{P(D)}f(x)$$

海维塞德认为如果 $1/P(D)$ 能展开成关于 $1/D$ 的幂级数，即

$$\frac{1}{P(D)} = \sum_{k=1}^{\infty} \frac{a_k}{D^k}$$

则原方程的解即为

$$y = \frac{1}{P(D)} f(x) = \sum_{k=1}^{\infty} a_k \left(\frac{1}{D^k} f(x) \right)$$

以一个具体的例子来说明上述方法可能更容易些。

考虑微分方程: $y' - y = 1$, $y(0) = 0$, 写成算子形式 $(D-1)y = 1$, 这里约定数字 1 为恒等算子。则按上面的形式算子方法得到形式解 (等比级数展开)

$$y = \frac{1}{D-1} \cdot 1 = \frac{1}{D} \left(\frac{1}{1-\frac{1}{D}} \right) = \frac{1}{D} \left(1 + \frac{1}{D} + \frac{1}{D^2} + \cdots \right) \cdot 1$$

根据定义有

$$\frac{1}{D} \cdot 1 = \int_0^x 1 dt = x, \quad \frac{1}{D^2} \cdot 1 = \frac{1}{D} \left(\frac{1}{D} \cdot 1 \right) = \int_0^x t dt = \frac{x^2}{2}, \quad \cdots, \quad \frac{1}{D^n} \cdot 1 = \frac{x^n}{n!}$$

$$\text{所以 } y(x) = \frac{1}{D} \left(1 + x + \frac{x^2}{2} + \cdots \right) = \frac{1}{D} e^x = \int_0^x e^t dt = e^x - 1$$

代回原微分方程验证, 它确实是一个解。

受过严格数学训练的人难以接受这种魔术般的形式方法, 但工程师不受此约束, 只要这个方法的结果是对的即可, 至于数学上的严格性不在考虑范围之内。此所谓“实用即是合理”原则。

实际上对于某些问题, 这种形式微分算子方法确实比传统方法效率高。当然, 本文的关键问题不是方法简洁与否, 而是如此奇怪的形式算子方法为什么是对的, 是巧合吗? 许多数学家不相信有那么多巧合, 其背后必定有深刻的数学逻辑。如果不是巧合, 那么很可能打开一片新的天地。

历史确实如此, 人们经过深入研究发现, 上面的形式算子解法本质是一种积分变换, 称为 Laplace-变换 (一般在数学物理方法课程中讲授)。

L-变换一般定义为:

$$L[f(t)] = \int_0^{+\infty} f(t) e^{-pt} dt = F(p)$$

这属于含参数积分问题, 为了保证被积函数的收敛性需要对 $f(t)$ 和正数 p (实际可以取复数) 附加一些控制条件, 实际应用中基本能满足。这里避开一些琐碎数学细节而直奔问题核心。

考虑微分方程:

$$P(D)y = f(x), y(0) = y'(0) = y''(0) = \cdots = y^{(n-1)}(0) = 0$$

则根据 L-变换性质, 有

$$L[y(x)] = Y(p), L[y^{(m)}(x)] = p^m Y(p), L[f(x)] = F(p)$$

对原方程两边做 L-变换得到: $P(p)Y(p) = F(p)$ ---把微分方程变换成代数方程 (对比下即知道 $P(p)$ 相当于形式微分算子里的 $P(D)$, 这说明海维塞德的形式微分算子方法本质是 L-变换)。

为对比海维塞德的方法, 默认可展开成幂级数

$$\text{求解代数方程得到: } Y(p) = \frac{1}{P(p)} F(p) = \sum_{k=1}^{\infty} \frac{a_k}{p^k} F(p) = \sum_{k=1}^{\infty} a_k \left(\frac{1}{p^k} F(p) \right)$$

L-变换存在逆变换 L^{-1} , L^{-1} 也是线性变换。则依照定义

$$y(x) = L^{-1}[Y(p)] = L^{-1} \left[\sum_{k=1}^{\infty} a_k \left(\frac{1}{p^k} F(p) \right) \right] = \sum_{k=1}^{\infty} a_k \left(L^{-1} \left[\frac{1}{p^k} F(p) \right] \right)$$

根据 L-变换的定义, 直接验证不难得到

$$L \left[\int_0^x f(t) dt \right] = \frac{1}{p} F(p)$$

反复应用上述结论, 得 k 重积分对应的 L-变换为 $\frac{1}{p^k} F(p)$, 于是求逆变换得到

$$L^{-1} \left[\frac{1}{p^k} F(p) \right] = \int_0^x dt \cdots \int_0^x f(t) dt$$

这正是海维塞德的形式微分算子法

$$\frac{1}{D^k} f(x) = L^{-1} \left[\frac{1}{p^k} F(p) \right] = \int_0^x dt \cdots \int_0^x f(t) dt$$

注意到 L-变换的每一个数学性质是有严格的数学证明的, 而不再是一种形式方法。这就揭示了形式微分算子的数学核心或本质是 Laplace 变换。

求逆变换 L^{-1} 有一般的数学公式, 不过要涉及复平面上的积分。比较实用的方法是对常见的函数建立 L-变换表, 应用时再查表 (像求积分中原函数), 这很像建立原函数表方法 (本质上导数表)。

评注：抓问题的核心能力需要多阅读、多思考、多积累，养成一种数学思考习惯，慢慢的就能提高了。所谓核心，也不一定非得是文字或者数学公式，有时候一张图，或者一张带箭头的图表，也能表明问题核心，要活学活用！

现在各行各业，很多人都爱谈抓住问题的本质（笔者也如此），但细细琢磨，发现其实说的是抓住问题的关键地方、核心或主要矛盾。问题的本质岂是那么容易发现的？除非本质这个词被滥用了。不过入乡随俗，大家都这么说，自己也把本质概念作广义理解，本文也如此。

笔者在高中时代做文科和理科试题（特别是选择题）时有一个习惯，即读题后，在试卷上用笔圈出“关键字”。特别是文科试题，一段话 100 多个字，真正有助于解题（或提示）的可能就三个词，而正确选项往往与这三个词相对应。这种圈定“关键字”方法使得自己那时的客观题准确率非常高。当初是为了应试才发现这个方法，后来在读书时也效仿这个方法，找出一篇文章最核心，或者最精华的观点。

上面这个方法在数学中也有用。什么叫做数学能力？不好定义，但有一个标准：假定一个数学定理的证明有 20 行，你若能找到证明中最关键的 3 行，这就是一种数学能力。为什么高手之间讨论常常话不多，但帮助非常大，就是能直击问题的核心。

第6章 类比是一种重要的数学思维

人们认识事物的一个常用的方法是找熟悉的东西来模仿，一般要求已知的对象与未知的对象有某些相似性，由此推断两者之间具有相似的特征或结构。这种方法其实就是类比方法了，是一种重要的数学思维。

进入大学以后，数学体系已经开始庞杂了，即便是数学分析这门基础课，其延伸分支可以包括复变函数、实变函数、ODE、PDE、微分几何，甚至泛函分析等等。知识容量大，记忆内容也多，难以找到头绪。面对这种当下现状，类比思维就很重要了，至少可以在一定程度上降低进入门槛。

就数学分析而言，细心一点会发现，很多概念和定理具有相似性，如收敛与一致收敛、连续与一致连续、多项式与幂级数、罗比达法则与 Stolz 定理、Dini 定理与 Levi 定理、Fourier 级数与 Fourier 变换等等。如果读者学习的时候留意这些相似性，那么在探索和记忆过程中，就能减少一些困难。

不过，类比是一种主观的不充分的似真推理，在数学中经常与猜想、形式推导等方法相混合。所以类比得到的结论不总是正确的，这就需要严格的数学证明来跟进补充。应该说，类比思维需要一定想象力和经验积累，所谓厚积薄发。

下面用几个具有代表性的例子来说明一下类比思维的运用，读者要用心体会。

1. 罗比达法则与 Stolz 定理

笔者是工科出身，因而先接触到罗比达法则，后自学《新讲》接触到 Stolz 定理，而数学系的教学过程通常是先学 Stolz 定理，再学罗比达法则。

笔者发现两者很有意思，一个是关于连续的，一个是关于离散的，形式上也挺像。以 $\frac{0}{0}$ 型极限问题为例，罗比达法则不难理解，因为有柯西中值定理。以此为桥梁，很容易过渡到导数之比上（假定导数之比的极限存在）

$$\lim_{x \rightarrow a+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a+} \frac{f'(x)}{g'(x)} = A$$

笔者一般记忆成：导数之比极限存在推出函数之比极限存在。

Stolz 定理的问题提法很自然，即两个无穷小数列极限之比的情况，相当于

离散罗比达法则。笔者采取类比思路，设法从罗比达法则中诱导出一个离散版本，想法是在可导函数上取一个无穷点列：

在 (a, b) 中选择一个严格单调递减的数列 $\{t_n\}$ ，满足 $t_n \rightarrow a$ 。容易想到等式左边为

$$\lim_{x \rightarrow a+} \frac{f(x)}{g(x)} = \lim_{n \rightarrow \infty} \frac{f(t_n)}{g(t_n)}$$

但是等式右边的导数极限如何替换？笔者后来想到了用差商代替导数，本质上还是柯西中值定理

$$\frac{f(t_n) - f(t_{n+1})}{g(t_n) - g(t_{n+1})} = \frac{f'(\xi_n)}{g'(\xi_n)}$$

再由 $\lim t_n = \lim \xi_n = a$ ，可知

$$\lim_{n \rightarrow \infty} \frac{f(t_n) - f(t_{n+1})}{g(t_n) - g(t_{n+1})} = A$$

如此经过转化，得到了一个离散版本的罗比达法则

$$\lim_{n \rightarrow \infty} \frac{f(t_n)}{g(t_n)} = \lim_{n \rightarrow \infty} \frac{f(t_n) - f(t_{n+1})}{g(t_n) - g(t_{n+1})} = A$$

再设 $b_n = f(t_n)$ ， $a_n = g(t_n)$ ，作为推广，猜想下面形式是一种自然思路

$$\lim_{n \rightarrow \infty} \frac{b_n}{a_n} = \lim_{n \rightarrow \infty} \frac{b_n - b_{n+1}}{a_n - a_{n+1}} = A$$

作为推广，猜想上面的等式只针对数列，不要求一定存在相应的可导函数 f, g 满足 $b_n = f(t_n)$ ， $a_n = g(t_n)$ 。

那么除了 $\lim b_n = \lim a_n = 0$ 这个条件外，还需要额外条件吗？注意到罗比达法则要求分母 $g(x)$ 满足： $g'(x) \neq 0, \forall x \in (a, b)$ 。由导数介值定理可知，必有 $g'(x) > 0$ 恒成立，或者 $g'(x) < 0$ 恒成立。总之 $g(x)$ 是严格单调函数。于是猜想分母 $a_n = g(t_n)$ 必须是严格单调数列。

综上，猜出了 Stolz 定理：设数列 $\{a_n\}, \{b_n\}$ 满足 $\lim b_n = \lim a_n = 0$ 且 $\{a_n\}$ 是严格单调递减的，若存在极限

$$\lim_{n \rightarrow \infty} \frac{b_n - b_{n+1}}{a_n - a_{n+1}} = A$$

则有

$$\lim_{n \rightarrow \infty} \frac{b_n}{a_n} = \lim_{n \rightarrow \infty} \frac{b_n - b_{n+1}}{a_n - a_{n+1}} = A$$

这里 $\{a_n\}$ 是严格单调递增的也可以，此时考虑 $-a_n$ 即可。

Stolz 定理的证明可在任何一本数学分析教材中找到，证明就不细说了。猜想的出发点在于从可导函数上取离散点列，接下来需要把导数翻译成离散语言，想到的突破口是“差商”，最后发现根据导数介值定理，需要分母是单调数列。如此类比猜出了 **Stolz** 定理。不过，如同数学中常见的，**Stolz** 定理的证明与罗比达法则毫无关系，这里主要揭示了两者的相似性不是偶然的，是有内在联系的。两个不同事物的内在联系往往就是问题的核心所在，这种事在数学里非常多。这给我们的启示是：看问题不要孤立，要多注意内在联系。

当然，单就 **Stolz** 定理的证明来说，笔者觉得核心是分母 a_n 的单调性，这个性质在证明中起到关键作用，大家要多体会。

2. 约当测度与 Lebesgue 测度

Lebesgue 测度的想法其实比较自然，主要是想推广长度、面积和体积这些基本概念到一般情况。为简单起见，这里只讲一维情形。诚然，对于比较规整图形，求长度、面积等容易的多。但对有一些比较糟糕的点集，如何定义“体积”概念就是一个首要问题。

假设对点集 E ，有一个惟一的非负实数 $m(E)$ 与之对应，它是体积的推广，姑且取个名词叫“测度”。测度既然是体积的推广，它应该兼容原有的体积特征，而不产生矛盾，这是前提。下述性质是体积应该满足的，测度也应该满足：

1. 非负性，即 $m(E) \geq 0$ ；
2. 有限可加性，即 $E \cap F = \phi$ （空集）时，有 $m(E \cup F) = m(E) + m(F)$ ，这种性质可以推广到有限个情形；
3. 正则性，即对区间 $[0, 1]$ ，其测度 $m([0, 1]) = 1$ 且 $m([a, b]) = b - a$ 。

这里允许闭区间退化为一点，即 $b = a$ 情况，于是对单个点 P 有 $m(P) = 0$ 。性质 3 相当于引入一个基准（单位），便于其它测度值与之比较。这很像人民币 100 元、50 元、20 元、10 元和 1 元，是基本计价标准。

如果想要推广体积概念，首先需要考虑能否推广（在兼容意义下）上述 3 个基本性质。不难看出性质 1 和性质 3 无法再推广了，剩下的就是性质 2 的推广了。Lebesgue 当初将有限可加性推广至可数可加性，这是一个质的飞跃，也是 Lebesgue 测度理论的关键地方。关于 Lebesgue 测度的完整内容可参考《实变函数》教材，这里只说说核心想法。

学完 Lebesgue 测度回过头想想，如果推广有限可加性到无穷，那么只能到可数无穷为止。否则的话对单个点有 $m(P) = 0$ （这是基本性质），而区间 $[0, 1]$ 有不可数个点构成，且不说“不可数加法”如何定义，而相加结果到底是 0 还是 1，也会引起混乱，与传统认识长度相违背。所以干脆到可数可加性为止。

于是抽象出勒贝格测度公理：

设 $\mathcal{A} = \{E\}$ 是直线 \mathbf{R} 上的一个集合族，使得对每个 $E \in \mathcal{A}$ ，都有惟一的实数 $m(E)$ 与之对应，满足：

1. 非负性，即 $m(E) \geq 0$ ；
2. 可数可加性，即如果 $E_1, E_2, \dots, E_n, \dots$ 两两不相交，则

$$m(\bigcup_{n=1}^{\infty} E_n) = m(E_1) + m(E_2) + \dots + m(E_n) + \dots$$

3. 正则性，即 $m([a, b]) = b - a$ 。

但光有干巴巴的 3 条公理，似乎还是寸步难行，到底测度 $m(E)$ 是什么样子呢？这需要进一步引入 L 测度定义。幸好当年的若尔当测度（有限可加性）在这方面已经有了探索性的想法。比如求一个图形面积，可以画出一些矩形格子，满足覆盖住图形。然后把矩形加细，仍然覆盖住图形。这样能逐步逼近图形面积。或者干脆直接类比 Riemann 积分中达布上和概念。

借助上述类比，可推广外测度概念。对于点集 $E \in \mathcal{A}$ ，选择可数个开区间 $I_1, I_2, \dots, I_n, \dots$ 的并集，满足

$$E \subset \bigcup_{n=1}^{\infty} I_n$$

然后令

$$m^*(E) = \inf \left\{ \sum_{n=1}^{\infty} |I_n| \right\}$$

这是一种非常自然的想法。这种定义 $m^*(E)$ 称为外测度。

当年讨论若尔当测度是已经有了外测度和内测度的想法了，直观说是外逼

近和内逼近, 简称“外压内挤”。读者也可以类比 Riemann 积分理论中的达布上和与达布下和。

不过 Lebesgue 遇到的困难可能更大, 它所考虑到的图形可能没有“内部”(如有理数集)。外逼近可以满足(因为是可覆盖的), 而如何处理内逼近就是一个难题了。Lebesgue 突发奇想: “内逼近”想法推广是对其补集外逼近。具体说, 设 $E \subset [a, b]$, E 在 $[a, b]$ 中的相对补集定义为:

$$E^C = [a, b] - E$$

于是, 自然将外测度概念应用到相对补集上, 诱导出内测度概念

$$m_*(E) := [a, b] - m^*(E^C)$$

最后, 效仿上积分和下积分思想, 定义如果对 $E \in \mathcal{A}$, 有内外测度相等

$$m_*(E) = m^*(E)$$

则称 E 是 Lebesgue 可测的, 其对应的 Lebesgue 测度记为

$$m(E) := m_*(E) = m^*(E)$$

注: 之所以引入了内测度概念, 是因为外测度只满足次可加性

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} m^*(E_n)$$

有反例表明可以取得小于号, 反例的构造要用到选择公理。这说明可以被外测度量度的集合范围有些大, 需要再附加一些条件缩小范围才行。数学就是这样, 常常是如果有反例出现, 则对原问题附加条件, 缩小范围, 排除反例。

3. Fourier 级数与 Fourier 变换

数学分析中还有一个有趣的现象, 即某个定理可能成对儿出现, 有一个离散版本, 很可能存在一个相似的连续版本, 反之亦然。前面提及的关于数列极限的 Stoltz 定理与可微函数的罗比达法则就是代表。

在后续的 Fourier 级数理论中还会遇到另外一对儿: Fourier 级数与 Fourier 变换, 即离散 VS 连续。从理论和实用角度看, Fourier 变换用途更广。比如在 PDE 理论以及信号与系统中。

采用复数形式表述, 一个周期为 T 的可积函数的 Fourier 级数可以表示成:

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{i \frac{2\pi n t}{T}}$$

其中系数

$$c_n = \frac{1}{T} \int_0^T f(t) e^{-i \frac{2\pi n t}{T}} dt$$

这里 $n = 0, \pm 1, \pm 2, \dots$. 为了表述方便, 以下记 $l = T/2, \alpha_n = n\pi/l$.

为了把 Fourier 级数想法推广到非周期函数 (定义在整个实轴上) 上, 一个自然想法是先考虑 $[-l, l]$ 情形, 再令 $l \rightarrow \infty$. 于是, 形式上有

$$f(t) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} (2c_n l) e^{in \frac{\pi}{l} t} \frac{\pi}{l}$$

以及

$$c_n = \frac{1}{2l} \int_{-l}^l f(t) e^{-i\alpha_n t} dt$$

为考察 $l \rightarrow \infty$ 情形, 先定义函数

$$c(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

当 l 很大时, $c(\alpha_n)$ 与 $2c_n l$ 相差很小, 这样得到近似关系

$$f(t) \approx \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} c(\alpha_n) e^{i\alpha_n t} \frac{\pi}{l}$$

注意到

$$\alpha_{n+1} - \alpha_n = \frac{\pi}{l}$$

是一个等距分割, 近似关系式右侧是一个 Riemann 和, 令 $l \rightarrow \infty$, 得到积分

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c(\omega) e^{i\omega t} d\omega$$

其中 “Fourier 系数的推广” (连续型) $c(\omega) = \hat{f}(\omega)$ 为

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

称作函数 $f(t)$ 的 Fourier 变换。由 $\hat{f}(\omega)$ 得到 $f(t)$ 的积分称作 Fourier 逆变换。

以上的推导是形式推导, 主要表明类比思考过程。实际上可以直接定义 Fourier 变换, $\hat{f}(\omega)$ 然后证明在某些条件下, Fourier 逆变换公式成立:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega t} d\omega$$

系数 2π 的处理有些麻烦, 所以导致有的 Fourier 变换定义是带系数的, 有的定义不带系数, 阅读文献时应该注意作者如何定义 F 变换, 以避免不必要的麻烦。

Fourier 变换有着物理意义, 主要体现在信号与系统中 (主要是频谱分析),

将 $f(t)$ 看做时域信号，经过 **Fourier** 变换得到 $\hat{f}(\omega)$ 变成了频域信号。数学上，积分是一种平滑算子，一个函数与核函数作用，经过积分后得到的新函数的性质往往比原来的函数要好（比如可微性方面）。实际应用中，根据经验，频域问题通常比时域问题要简单，一个复杂的时域信号 $f(t)$ 看不出规律，但经过 **Fourier** 变换后得到频域信号 $\hat{f}(\omega)$ 很可能就有规律了。

关于 **Fourier** 变换在数学研究中的作用可参考数理方程教材或者 **PDE**（偏微分方程）方面的教材。笔者的一点经验是：**Fourier** 变换和 **Laplace** 变换把微分运算变成代数运算，而代数运算相对简单。**Fourier** 变换的更深应用要涉及广义函数论和椭圆算子等深入领域。

4. 椭圆函数加法公式

下面的例子深一些，使用的类比思想非常经典，主要是高斯和阿贝尔两位大师的工作。下面的文章本质是两位大师的“合力之作”，笔者加入一点类比评注，使得思维上衔接更紧凑，权当抛砖引玉。

文章主要处理一类称作雅可比椭圆正弦函数 $\text{sn } u$ 。这是一类特殊的椭圆函数，本文采用阿贝尔的记号 $\varphi(u)$ 来表示标准记号 $\text{sn } u$ 。

第一类椭圆积分形式为

$$u = \int_0^z \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

其中，参数 k 满足 $0 < k^2 < 1$ 。由此可以定义出一个反函数 $z = \varphi(u)$ ，这是椭圆积分的反演（逆向思维），一般称 $\varphi(u)$ 为椭圆函数（双周期亚纯函数）。

利用积分反演定义新的函数并不是什么少见的事情，众所周知

$$u = \int_0^z \frac{dx}{\sqrt{1-x^2}} = \arcsin z$$

定义了正弦函数 $z = \sin u$ 。这其实是椭圆积分中 $k=0$ 对应的退化情形（极端情况），下面借助正弦函数加法公式作为类比来研究椭圆函数加法公式。

我们知道，正弦函数 $\sin u$ 有加法公式

$$\sin(u+v) = \sin u \cos v + \cos u \sin v$$

写成更对称的导数形式是

$$\sin(u+v) = \sin u \sin' v + \sin' u \sin v$$

注意到 $\varphi(u)$ 的退化情形 ($k=0$ 时) 即为 $\sin u$, 我们可以猜想 $\varphi(u+v)$ 中有因子

$$\varphi(u)\varphi'(v) + \varphi'(u)\varphi(v)$$

于是, 不妨猜想

$$\varphi(u+v) = \frac{\varphi(u)\varphi'(v) + \varphi'(u)\varphi(v)}{\psi(u, v, k^2)}$$

写成商的形式纯粹是为了运算上的方便, 后面我们会看出原因。接下来的重点是猜测待定函数 $\psi(u, v, k^2)$ 的形式。

由正弦函数加法公式可知

$$\psi(u, v, 0) = 1$$

椭圆积分两边对变量 z 求导, 得到

$$\frac{du}{dz} = \frac{1}{\sqrt{(1-z^2)(1-k^2z^2)}}$$

所以有导数关系 (考虑反函数微分方程)

$$\varphi'(u) = \frac{dz}{du} = \sqrt{(1-\varphi^2(u))(1-k^2\varphi^2(u))}$$

为方便起见, 引进辅助函数 (类似于 $\cos u$)

$$f(u) = \sqrt{1-\varphi^2(u)}, \quad g(u) = \sqrt{1-k^2\varphi^2(u)}$$

于是 $\varphi'(u) = f(u)g(u)$, 以及 $\varphi(0) = 0, f(0) = g(0) = 1, \varphi'(0) = 1$

注意到 $\varphi(u+v)$ 中 u 和 v 是轮换对称的且满足

$$\psi(0, v, k^2) = \psi(u, 0, k^2) = 1$$

为了得到 $\psi(u, v, k^2)$ 的进一步信息, 我们考虑另一种退化情形 $k^2 = 1$. 此时, 原积分化为初等积分

$$u = \int_0^z \frac{dx}{1-x^2} = \frac{1}{2} \ln \frac{1+z}{1-z}$$

即

$$z = \varphi(u) = \frac{e^{2u} - 1}{e^{2u} + 1}$$

又因为 $e^{2u} = \frac{1+\varphi(u)}{1-\varphi(u)}$, 由此不难得到下面的加法公式 (对应 $k^2 = 1$ 情况)

$$\varphi(u+v) = \frac{\varphi(u) + \varphi(v)}{1 + \varphi(u)\varphi(v)}$$

因为此时 $\varphi'(u) = 1 - \varphi^2(u)$ ，直接计算，得

$$\begin{aligned} \varphi(u)\varphi'(v) + \varphi'(u)\varphi(v) &= \varphi(u)(1 - \varphi^2(v)) + (1 - \varphi^2(u))\varphi(v) \\ &= (\varphi(u) + \varphi(v))(1 - \varphi(u)\varphi(v)) \end{aligned}$$

由此可知

$$\psi(u, v, 1) = (1 + \varphi(u)\varphi(v))(1 - \varphi(u)\varphi(v)) = 1 - \varphi^2(u)\varphi^2(v)$$

再注意到 $\psi(0, v, k^2) = \psi(u, 0, k^2) = 1$ 及 $\varphi(0) = 0$ ，故猜想

$$\psi(u, v, k^2) = 1 - k^2\varphi^2(u)\varphi^2(v)$$

是一种比较自然和整齐的思路。

于是我们归纳猜想出加法公式：

$$\varphi(u+v) \approx \frac{\varphi(u)\varphi'(v) + \varphi'(u)\varphi(v)}{1 - k^2\varphi^2(u)\varphi^2(v)} =: D(u, v)$$

一般来说，知道了猜想的结论（概率意义上），总是可以验证的，下面直接证明： $\varphi(u+v) = D(u, v)$ 。这个精彩的证明出自数学大师阿贝尔之手！

证：利用偏导数公式，经过相对繁琐的计算，可知

$$\begin{aligned} & [1 - k^2\varphi^2(u)\varphi^2(v)]^2 \frac{\partial D}{\partial u} \\ &= f(u)g(u)f(v)g(v)[1 + k^2\varphi^2(u)\varphi^2(v)] - \varphi(u)\varphi(v)[g^2(u)g^2(v) + k^2f^2(u)f^2(v)] \end{aligned}$$

由此不难发现轮换对称性，剩下只需求解偏微分方程

$$\frac{\partial D}{\partial u} = \frac{\partial D}{\partial v}$$

考虑变换 $u = (s+t)/2, v = (s-t)/2$ ，于是

$$\frac{\partial D}{\partial t} = \frac{1}{2} \left(\frac{\partial D}{\partial u} - \frac{\partial D}{\partial v} \right) = 0$$

故存在函数 ϕ ，使得

$$D = \phi(s) = \phi(u+v)$$

又 $v=0$ 时，有 $\phi(u) = D(u, 0) = \varphi(u)$ ，这样我们证明了

$$\varphi(u+v) = \frac{\varphi(u)\varphi'(v) + \varphi'(u)\varphi(v)}{1 - k^2\varphi^2(u)\varphi^2(v)}$$

这个例子巧妙地将“类比、猜想与证明”融合到一起，比较能代表笔者对数学研究的观点。

一般的椭圆积分形式如下：

$$\int R(z, \sqrt{p(z)}) dz$$

这里 R 是一个有理函数， $p(z)$ 是一个三次或四次的多项式，且有两个不同的零点。数学上已经证明了，椭圆积分是非初等积分。

例：下面的积分称为双纽线积分，这个积分在计算双纽线曲线长度时自然出现，高斯就是通过考察这个积分，发现了椭圆函数的

$$\int \frac{1}{\sqrt{1-z^4}} dz$$

椭圆积分的一个重要来源是在计算椭圆周长时发现的，故而得名。由此可知，椭圆弧长积分是非初等积分。

评注：画家当初入门时都是从临摹开始的，学习数学也是如此。我们学习前人的劳动成果，慢慢提高自己的水平，然后再创造自己的作品。类比，则是认识数学的重要方法（如婴儿学步），希望通过上面 4 个有技术含量的具体例子，能使得大家对数学中的类比思想有一个具体体会，避免泛泛论道。

第7章 猜想与形式推导

貌似是大一的时候，选修了一门数学系开设的“数学思维欣赏”类课程，这门课程由数学系的好几个老师轮番讲，其中一位老师的观点是：要想搞好数学，得会猜。提出一个猜想，如果最终被证明是正确的，则美其名曰是洞察力。当然，有的猜想好猜，有的不那么好猜。

从文章标题“猜想与形式推导”看，是一个矛盾体，笔者试图将直觉主义与形式主义结合起来，说的通俗一些，本文重点即谈谈数学中的“连猜带证”方法。关于形式推导，需要强调的是“形式推导本身在逻辑上是严格”，关键在于推导过程的前提是假设的或者猜想的。通俗地说，默认某些前提假设是正确的，然后开始形式推导。比如假设方程

$$\frac{dy}{dx} = y, y(0) = 1$$

的解可以写成幂级数 $y(x) = a_0 + a_1x + a_2x^2 + \dots$ ，这个级数还可以逐项求导。于是代入原方程，得到递推关系 $na_n = a_{n-1}$ ，再由 $a_0 = y(0) = 1$ 可以求出 $a_n = 1/n!$ 。于是得到原方程的解

$$y(x) = 1 + x + \frac{1}{2!}x^2 + \dots + \frac{1}{n!}x^n + \dots$$

所以形式推导的关键在于不是推导本身，而是推导的前提。而这看似不起眼的前提却是关键所在。很多人用很繁琐的数学工具做一个物理模型，效果却非常差，其根源往往不在于数学工具本身，而在于其前提条件是否适用。抛开前提假设或猜想，去讨论数学工具本身，这是本末倒置。

就笔者所见，现行的数学教材大都采用形式化写法，即先给出若干定义，然后给出 $A \Rightarrow B \Rightarrow C$ 之类的证明。虽然逻辑上无懈可击，但是读起来却比较乏味，收获不多。其实读读数学史可以知道，一个定理从发现到证明，不可能是完全理性的，“连猜带证”的现象比比皆是。

一位数学大师甲可能经过一种天马行空一般的途径 A，巧妙地猜出结论 B，然后用当时流行的“严密方法”（可能在现代看来还是不严密）给出定理 A 的证明 C。随着时间的流逝，定理 A 却被张冠李戴地冠以数学大师乙的名字，而教材中给出的证明却是完全不同于 C 的方法 D。通常后人们只能读到结论 B 和证明 D。至于 C 有可能在老掉牙的数学历史文献中的角落找到，而当初的途径 A 则可能永远消失在历史的尘埃中。

途径 A, 即高斯所说的“脚手架”, 定理 B 即瑰丽的建筑, 不论在教材或者是学术论文中, 总是把“脚手架”拆除掉, 留下僵硬的严密证明。每当笔者读到一个定理时, 总不禁要问, 为什么只给出“证”, 没有“猜”呢? 很多结论看起来不是那么显然啊。在有了一些科研经历后, 更是有这种感慨, 现实问题比课本复杂多了, 有着太多的变量, 有着太高的不确定性。“猜”和“证”的相合, 才是完整的科研。

看过笔者文章的人都知道, 笔者特别称赞张筑生老师的《数学分析新讲》教材, 不仅仅因为这是自己的第一本数学分析书, 还因为张老师的写法启发性特别强, 对“猜”的论述非常独到。本章试图效仿新讲, 给出一些“猜”和“证”的例子, 分享一下笔者的学习感悟, 读者可以从中学习一些探索方法, 有助于数学分析的学习。需要说明的是限于篇幅, 笔者重点谈谈“猜”, 至于标准的严格证明, 可参考《新讲》或者其它教材。

先从一个例子谈起: $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$. 这个等式最早由欧拉证明, 后来人们发

现了许多证明。标准的证明是考察偶函数 $f(x) = |x|$ 的在 $[-\pi, \pi]$ 上的 Fourier 级数展开式, 然后计算出 $b_n = 0$, 以及

$$\begin{aligned} a_0 &= \frac{2}{\pi} \int_0^{\pi} x \, dx = \pi \\ a_n &= \frac{2}{\pi} \int_0^{\pi} x \cos nx \, dx = \frac{2}{\pi n^2} [(-1)^n - 1] \end{aligned}$$

再由 $f(x) = |x|$ 是分段单调的连续函数知其 Fourier 级数收敛到自身

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

即 (下面利用了 $a_{2n} = 0$ 这一结论)

$$|x| = \frac{\pi}{2} - \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{\cos(2n+1)x}{(2n+1)^2}$$

注意到

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \sum_{n=1}^{\infty} \frac{1}{(2n)^2} + \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \Rightarrow \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{4}{3} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2}$$

令 $x=0$ ，可知

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{4}{3} \times \frac{\pi^2}{8} = \frac{\pi^2}{6}$$

比较巧妙的方法如“二重积分法”，即

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \int_0^1 \int_0^1 \frac{1}{1-xy} dx dy$$

将被积函数看成是等比级数，于是

$$\begin{aligned} \int_0^1 \int_0^1 \frac{1}{1-xy} dx dy &= \int_0^1 \int_0^1 \sum_{n=0}^{\infty} (xy)^n dx dy = \sum_{n=0}^{\infty} \int_0^1 \int_0^1 x^n y^n dx dy \\ &= \sum_{n=0}^{\infty} \int_0^1 x^n dx \int_0^1 y^n dy = \sum_{n=0}^{\infty} \frac{1}{n+1} \frac{1}{n+1} \\ &= \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} \end{aligned}$$

于是，只需要计算出上面的二重积分即可。这个积分可以通过巧妙的坐标变换求出来，细节见《数学天书中的证明》一书，建议大家读读。

该问题尽管有很多巧妙的证明，但欧拉本人的想法仍然非常有启发性，其想法类比了多项式因式分解。

注意到 $\sin n\pi = 0$ ，其中 $n = \pm 1, \pm 2, \dots$ 为方程 $\sin \pi x = 0$ 的所有非零根，再类比多项式因式分解，欧拉猜想出无穷乘积

$$\frac{\sin x}{x} = \left(1 - \frac{x^2}{\pi^2}\right) \left(1 - \frac{x^2}{2^2 \pi^2}\right) \left(1 - \frac{x^2}{3^2 \pi^2}\right) \cdots = \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2 \pi^2}\right)$$

最后，利用正弦函数 $\sin x$ 的级数展开式，并将等式右边无穷乘积展开成幂级数，然后比较等式两边 x^2 项的系数，即得到

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

当然，欧拉当时并没有考虑收敛性问题（那个时代皆如此），有点阴差阳错得到了正确的结论。但是在科研中需要明白“有思路总比没思路强，猜出了结果，证明起来就相对容易了”这个道理。如果，抛开数学证明的严格性要求，欧拉的方法无疑更巧妙（至少给出了猜测的方向）。

不论是古人，还是现代人，实际科研中很少有把所有条件都验证完毕后，

才开始行动的。经常是默认某些条件成立，然后往下推导，遇到问题再跳回去检验和修改，这是一种反复跳跃性模式。

有的情况是猜出了结果，然后用形式推导来反推，最后按正方向来证明。比如对 \mathbf{R} 上的函数 $f(x)$ ，通过一些例子，猜想 $f(x)$ 可以表示成一个偶函数与奇函数之和（相当于“有序”分解）。初看起来有些棘手，但如果我们假设结论成立，即 $f(x) = s(x) + c(x)$ （奇偶函数之和），则根据定义，有

$$f(x) = s(x) + c(x) \text{ 以及 } f(-x) = s(-x) + c(-x) = -s(x) + c(x)$$

然后就像解方程一样，求出 $s(x)$ 和 $c(x)$ 表达式。严格的证明就是直接验证

$$f(x) = \frac{f(x) - f(-x)}{2} + \frac{f(x) + f(-x)}{2}$$

很多定理就是类似证明的，不过一般只给出严格的证明（但缺少探索的信息）。这个例子还算比较显然的（至少好猜），有些复杂的例子就不那么好猜了。

比如考察两个黎曼可积函数 $f(x), g(x)$ ，这里 $g(x+T) = g(x)$ 是周期函数。则有如下的黎曼-勒贝格引理：

$$\lim_{n \rightarrow \infty} \int_0^T f(x)g(nx) dx = \frac{1}{T} \int_0^T g(x) dx \int_0^T f(x) dx$$

将整数 n 换成一般的实数 λ ，结论仍然成立。

上面这个等式不太容易直接想到。首先，类比 Fourier 级数中的结论

$$\lim_{\lambda \rightarrow \infty} \int_a^b f(x) \sin \lambda x dx = \lim_{\lambda \rightarrow \infty} \int_a^b f(x) \cos \lambda x dx = 0$$

则自然猜想出下面的等式（当然不一定无条件成立）：

$$\lim_{n \rightarrow \infty} \int_0^T f(x)g(nx) dx = 0$$

令 $f(x) \equiv 1$ ，上面等式恒成立，则有

$$\int_0^T f(x)g(nx) dx = \int_0^T g(nx) dx = \frac{1}{n} \int_0^{nT} g(t) dt = \int_0^T g(t) dt$$

上式的极限是 0，因此只能是

$$\int_0^T g(x) dx = 0$$

考虑 $g(x) = \sin x$ 这个特例，显然满足题意。但对一般的周期函数 $g(x)$ ，在一个周期内的积分不一定为 0（如 $g(x) = 1 + \sin x$ ），通常是

$$\int_0^T g(x) dx \neq 0$$

不过，借助几何直观，我们发现将周期函数 $g(x)$ 适当向上（或向下）平移 C 个

单位之后，得到的新周期函数 $\varphi(x) = g(x) + C$ 满足周期积分为 0，即

$$0 = \int_0^T \varphi(x) dx = \int_0^T g(x) dx + C \int_0^T dx$$

由此得到

$$C = -\frac{1}{T} \int_0^T g(x) dx$$

因为我们“猜想”

$$\lim_{n \rightarrow \infty} \int_0^T f(x) \varphi(nx) dx = 0$$

所以自然进一步猜想

$$\lim_{n \rightarrow \infty} \int_0^T f(x) g(nx) dx = \lim_{n \rightarrow \infty} \int_0^T f(x) (\varphi(nx) - C) dx = \frac{1}{T} \int_0^T g(x) dx \int_0^T f(x) dx$$

猜想出结果之后，总是可以验证的，而且证明起来也有了方向。具体证明方法可选择利用阶梯函数逼近 $f(x)$ 来证明，这里从略。

一般来说，猜出结果总比证明要容易一些。特别是我们可以通过一些特例归纳出一般性的结论，然后进一步考虑给出严格的证明。但不得不承认，产生出一个好的想法并不容易，依赖所谓的直觉或经验，这方面需要坚持不懈地练习与思考。比如，猜想一个“好的”函数 $f(x)$ 可以展开成幂级数（泰勒级数）就是一个大胆而深刻的猜想，即

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots$$

有了这个想法后，再利用形式推导确定系数相对容易多了。令 $x=0$ ，得到 $f(0) = a_0$ ；上面等式两边对 x 求导，再令 $x=0$ ，得到 $f'(0) = a_1$ ；如此类推，得到一半的关系式 $f^{(n)}(0) = n! a_n$ 。

在泰勒级数问世之前，牛顿等数学大师已经导出了 $e^x, \sin x, \cos x$ 等函数的幂级数表达式，通过类比得到一般的猜想 $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots$ ，由此诞生了解析函数。

站在后人的角度看前人，精彩之余也确实感到数学的严格性问题，实际上很多大师因为当时的整体环境限制也得到了一些荒谬的结论。所以才有了后来柯西与魏尔斯特拉斯等人关于微积分基础的数学工作。

大胆猜想，形式推导，案例验证和严格证明这样一条龙的方式是笔者推荐的数学思维方法。

最后我们从工科生的独特角度谈谈复变函数中的柯西定理来结束本文，限于篇幅，这里不对定理成立的一般性条件做精确表述，只介绍一些想法。严格而完整的证明请参考阿尔福斯的《复分析》一书。

下面先给出工科“三不管原则”

1. 不管三七二十一，先对 $f(x)$ 展开成泰勒级数（或多项式）再说；
2. 不管三七二十一，先对级数逐项积分或逐项求导再说；
3. 不管三七二十一，先对含参数积分在积分号下求导再说。

注：数学上严格证明了以上“三不管原则”对解析函数成立，这里只是为了知识串联方便而已。

这种“三不管原则”思路更多见于工科生的解题或科研中，物理和力学中很多公式的推导也是利用了“三不管原则”的，笔者当年也经常使用这三个原则的。就笔者了解，一般工科《高等数学》或《微积分》教材大多不讲授“一致收敛”概念，更不用说“控制收敛”概念，直接应用上述三不管原则，这在数学系看来，有些难以接受。所幸，实践表明“三不管”方法非常简单实用，好用就是合理（工程中的经典原则）。

在自学完《数学分析新讲》后，被欧拉公式

$$e^{ix} = \cos x + i \sin x$$

的奇妙所吸引。不禁想到，要是一般的复函数 e^z ，或者更一般的复函数 $f(z)$ 是不是也会有一些奇妙的结论呢？心动不如行动，考虑当变量为函数的情况，有没有类似推广。于是决定采用如下方法。首先，尽可能少定义新的概念，能从微积分中 copy 就 copy，最好是直接拿来主义，然后再套用“三不管原则”。

根据复数实数化原则，即 $z = x + iy$ ，将复数 z 看成两个实数外加一个虚数单位 i ，于是先研究 $f(z) = f(x + iy) = u(x, y) + iv(x, y)$ 。自己的工具其实就是微积分里那些工具，导数、积分和级数。

首先，连续性定义不难 copy，接下来就是推广导数，类比定义复导数

$$\lim_{h \rightarrow 0} \frac{f(z+h) - f(z)}{h} =: f'(z)$$

如果此极限存在，则称 $f(z)$ 在 z 点可导，并把此极限记为 $f'(z)$ 。如果在一个区

域 D 内, $f'(z)$ 处处存在, 则称 $f(z)$ 在 D 内解析或全纯。

注意到, 复数 h 可以以很多种途径趋近于 0 , 于是上面的定义很像二元函数的连续性与方向导数定义。如果一个定义是普适的, 那么对特例情况也成立。故沿着两个特殊方向 x 轴和 y 轴定义的复导数应相等, 由此导出 C-R 方程:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

而当 u, v 有连续的一阶偏导数时, 由 C-R 方程也可以推出导数 $f'(z)$ 存在。复变函数教材中严格证明了, 如果导数 $f'(z)$ 存在, 那么 C-R 方程成立, 且 u, v 有连续的一阶偏导数 (实际上是任意阶偏导数都连续)。

下面是定义积分。因为复变函数 $f(z)$ 实际是实 2 维的 (在复平面 C 上), 但又必须当做一个整体 (复 1 维)。类比曲线积分, 一个想法是定义在一条曲线 C

$$\int_C f(z) dz := \int_a^b f(z(t)) z'(t) dt$$

此想法还是“复数实数化思想”, 即意识到 $z = x + iy$ 和 $f(z) = u(x, y) + iv(x, y)$, 而曲线的参数方程是 $x = x(t), y = y(t)$ 。本质上, 该想法是将复积分看成两个实曲线积分的组合。但是对同一条曲线 $C: z(t) = x(t) + iy(t)$, 可能有不同的参数方程, 不过根据 1 阶微分形式不变性原理, 上面积分的不依赖于参数方程的具体选择。

有了积分定义, 可以进行计算了, 不难得到

$$\int_C f(z) dz := \int_C (u + iv)(dx + idy) = \int_C u dx - v dy + i \int_C v dx + u dy$$

看到这个表达式之后, 出于习惯, 自己联想到当曲线 C 是闭曲线时, 似乎可以应用格林公式, 计算表明当导函数 $f'(z)$ 连续, 对闭曲线 C 有

$$\begin{aligned} \int_C f(z) dz &= \int_C u dx - v dy + i \int_C v dx + u dy \\ &= -\iint_D \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) dx dy + i \iint_D \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right) dx dy = 0 \end{aligned}$$

这里应用了前面的 C-R 方程。

于是, 得到一个著名的结果 (柯西定理)

$$\int_C f(z) dz = 0$$

上面这个定理是复变函数的基础，地位相当于微积分中牛顿-莱布尼茨公式

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

对比一下曲线积分与路径无关的内容，可定义一个只依赖于起始位置的“点函数” $F(z)$ ，它不依赖于路径的选取

$$F(z) = \int_0^z f(\tau) d\tau$$

这个函数 $F(z)$ 是解析的。

出于习惯，得到一个抽象的结论之后，总想着拿例子验证一下。选择简单的幂函数 z^n 来验证，支持自己的判断。对于柯西定理，实际上可以证明只需要 $f'(z)$ 存在即可，无需 $f'(z)$ 连续的假设。特别地，若曲线 C 包围的区域内存在有限个可去奇点 ξ_i ，则柯西定理仍然成立。这里可去奇点 ξ_i 定义为

$$\lim_{z \rightarrow \xi_i} (z - \xi_i) f(z) = 0$$

不管三七二十，先把 $f(z)$ 展开成泰勒级数再说。

$$f(z) = f(a) + f'(a)(z-a) + \frac{1}{2!} f''(a)(z-a)^2 + \dots$$

两边直接求积分则很平凡，但发现对等式

$$\frac{f(z)}{z-a} = \frac{f(a)}{z-a} + f'(a) + \frac{1}{2!} f''(a)(z-a) + \dots$$

两边求积分（注意到等式右边除第一项外，每项都是多项式），得到

$$\int_C \frac{f(z)}{z-a} dz = f(a) \int_C \frac{dz}{z-a}$$

这里 C 总是默认为一条可求长简单闭曲线。

如果 C 是圆周 $C: z = a + re^{i\theta}$ 那就好办，直接代入得到

$$\int_C \frac{dz}{z-a} = \int_0^{2\pi} \frac{ire^{i\theta} d\theta}{re^{i\theta}} = 2\pi i$$

但是闭曲线 C 不是圆周怎么办？自己忽然想到了格林公式里面的“挖奇点”技巧，在闭曲线 C 内做一个包含点 a 的充分小的圆 C_ϵ ，则闭曲线 C 与圆 C_ϵ 所围成的区域没有奇点，可以应用格林公式（或柯西定理），于是得到

$$\int_{C-C_\epsilon} = 0$$

这里要考虑曲线的定向。由此得到

$$\int_C \frac{dz}{z-a} = \int_{C_\epsilon} \frac{dz}{z-a} = 2\pi i$$

代回原积分，得到另一个有趣的结果（柯西积分公式）

$$f(a) = \frac{1}{2\pi i} \int_C \frac{f(z)}{z-a} dz$$

有意思，一点的函数值竟然可以用包围该点的闭曲线的积分来确定。

将上面的公式看做一个含参变量 a 的积分，再来一下积分号下求导，得到

$$f'(a) = \frac{1}{2\pi i} \int_C \frac{f(z)}{(z-a)^2} dz$$

反复积分号下求导，得到

$$f^{(n)}(a) = \frac{n!}{2\pi i} \int_C \frac{f(z)}{(z-a)^{n+1}} dz$$

看到导数竟然与积分发生“等式联系”，自己想到了泰勒级数与傅里叶级数中的系数，前者的系数通过导数确定，或者有的系数通过积分确定。想不到，竟然在复数里取得了思想上的“统一”。

我们已经得到了解析函数 $f(z)$ 的导函数公式

$$f'(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{(\zeta-z)^2} d\zeta$$

那么如果 $f(z)$ 是有界函数呢？设 $|f(z)| \leq M$ ，然后取 $C = \{z: |z-z_0|=R\}$ 为一个圆周，代入上面式，那么有不等式

$$|f'(z)| \leq \frac{1}{2\pi} \int_{|z-z_0|=R} \frac{M}{R^2} d\zeta = \frac{M}{R}$$

特别地，若 $f(z)$ 在整个复平面上解析且有界，则令半径 $R \rightarrow +\infty$ ，得到 $|f'(z)|=0$ ，于是 $f(z)=const$ 常数。

（刘维尔定理）复平面上解析且有界的函数只有常函数。

复平面上解析的函数称作整函数。这个定理有些其貌不扬，但非常实用。比如用它证明代数基本定理。思路是反证法，假设非常数复多项式无零点，即恒有 $P(z) \neq 0$ ，考虑 $1/P(z)$ ，则它是复平面上解析且有界的函数，于是根据刘维尔定理 $1/P(z)$ 是常数，从而复多项式 $P(z)$ 也是常数，矛盾！

关于刘维尔定理的推广是“毕卡小定理”：

设 $f(z)$ 是复平面上的整函数，若 $f(z)$ 有两个值取不到，则 $f(z)=const$ 。

接着前进，我们看到前面导数与积分的等式说明了：如果 $f(z)$ 是解析函数，那么 $f'(z)$ 也是解析函数。再联想到，

$$F(z) = \int_0^z f(\tau) d\tau$$

由此得到一个定理 (Morera):

设复变函数 $f(z)$ 在区域 D 内连续，且对 D 内任意一条简单闭曲线 C 有

$$\int_C f(z) dz = 0$$

则 $f(z)$ 是解析函数。

证明很自然，任取一点 $z_0 \in D$ ，则解析函数

$$F(z) = \int_{z_0}^z f(\tau) d\tau$$

处处可导且 $F'(z) = f(z)$ 。根据上面的定理 $F(z)$ 是解析函数，则 $F'(z)$ 也是解析函数。

上面这些结论是复变函数中的标准内容，都有严格的数学证明。这里只是作为例子示范一下“如何利用形式推导”来进行知识的串联。

评注：以上故事当然是虚构的，不过那“三不管原则”（三板斧）倒是真的，笔者见过的许多人（包括我自己）都这么用过。物理学大师费曼算是工科生吧，据他自己在自传中描述到，他一直搞不懂围道积分，所以一直用积分号下求导这一方法，用的是炉火纯青！

猜想与形式推导结合起来，枯燥的数学证明也可以变成艺术。

第8章 硬算是门基本功

听很多人说过，数学分析这门课程技巧很多，特别活。确实，学好数学分析必须掌握分析中的一些基本技巧，如逐次逼近、分段估计、基底分解，反演变换等等。有些方法之精巧，构思之巧妙，甚至令人望而生畏。

客观说，数学分析里也有技巧性要求较低的方法，比如计算。很多人将计算归到微积分中，认为和数学分析是两回事。笔者没有这方面成见，数学分析和微积分本是同源，先贤大师们都不认为是两家，后人却强分彼此，不由想起了华山派气宗和剑宗的故事。有句话说，天下武功唯快不破。那么也可以说，数学分析为“算”不破。做题也好，搞科研也罢，很多好的想法或猜想就是在计算中发现的，数论中的类似例子就很多（如素数定理）。也有一些分析定理，其证明本质是一系列计算。

数学大师中的计算高手要数欧拉和高斯了，华人数学家也有计算高手，如华罗庚和陈省身。华老是计算高手地球人都知道，如当年搞统筹法，经常与计算打交道。另一个例子是华老当年曾破译了日军密码。话说国军谍报人员俞大维，当年是留美数学博士，在数理逻辑方面颇有造诣，但最近却遇到了破解日军军事密码的困难，一大堆工作人员忙了几个月都无法破解。一次偶然的机会，经蔡孟坚将军牵线，俞大维请华老帮忙破解密码。华老连夜观察、反复比对，仔细寻觅密码中数字的规律，彻夜未眠。进过一夜的工作，华老成功地破译了日军的军事密码。据当事人私下回忆：日军那时使用的密码技术，是把原来的文件，俗称“明文”，用数学方法变换一下，称为“加密”。加密后的文件，俗称“密文”。“密文”传输出去，即使被截获，别人也如同雾里看花，难解其意。华老发现了日军密码的秘密：从明文变换到密文的加密过程，日军用的是数论中的“默比乌斯函数”。那么，采用反函数，密文立即被还原成明文，日军密码就这样被破译了。华老的这种洞察力离不开深厚的计算功底和对数论的精通，二者缺一不可。

相比华老，陈老爷子也是计算高手就鲜为人知了。虽然微分几何中计算也不少，但与丘成桐的偏微分方程路线不同，陈老爷子主攻整体微分几何，更看重代数拓扑等整体性数学工具。但从一件小事可以看出陈老爷子的计算功力，伍洪熙曾经说过：陈老爷子私下对他说过，老爷子当年的成名之作 Gauss-Bonnet 公式的内蕴证明那篇文章，所有的推导都是心算的。这篇文章有中文翻译，大

家可以在《陈省身文集》中找到，只要看看这篇文章，就能体会到老爷子的计算功底。

举了上面华老和陈老的例子意在纠正纯数学中一种不太正常的风气，如数学系的只会证明就行了，计算能力差无所谓，何况现在是电脑时代了。很多纯数学大师是计算高手，只是他们自己不显山露水而已。当然，计算能力特别烂的也可以成为数学大师（如拓扑和微分动力系统方面的数学大师斯梅尔，菲尔兹得主），不过这不是理由。特别对于分析方向，若计算能力不行的话，基本就走不通了。

很多数学定理的发现或者证明确实是算出来的，即便将来打算搞基础数学，适当加强自己计算能力也是有好处的（特别是电脑时代，数值实验变得很容易）。有些东西你不亲自动手算算，你是看不出规律的，你不积累，如何爆发！

回顾历史，许多大数学家都是擅长计算的，比如笔者的数学偶像高斯，后半辈子在搞天文计算。那时没计算机，基本靠手算，天文数字很好算吗？不过后人整理高斯的手稿时，发现他很少有算错的。

以高斯发现素数定理为例。记 $\pi(x)$ 表示不超过 x 的素数的个数，所谓素数定理，指下面极限表达式

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x / \ln x} = 1$$

高斯记函数 $\text{Li } x = \int_2^x \frac{dt}{\ln t}$ （这是一个非初等积分），这是 $\pi(x)$ 的另一个渐进表达式。由洛比达法则可知

$$\lim_{x \rightarrow \infty} \frac{\text{Li } x}{x / \ln x} = 1$$

函数 $\text{li } x$ 按如下方式定义

$$\text{li } x = \lim_{\eta \rightarrow 0} \left(\int_0^{1-\eta} + \int_{1+\eta}^x \right) \frac{dt}{\ln t}$$

显然 $\text{Li } x = \text{li } x - \text{li } 2$ 。

与前人思路不同高斯不是简单研究 $\pi(x)$ 的数值分布特点，而是考察比值 $\pi(x)/x$ 的分布特点，猜测是否有简单函数来近似代替。

x	$\pi(x)$	$\frac{x}{\log x}$	$\text{li } x$	$\frac{\pi(x)}{\text{li } x}$	$\frac{\pi(x)}{x}$
1 000	168	145	178	0.94...	0.1680
10 000	1 229	1 086	1 246	0.98...	0.1229
50 000	5 133	4 621	5 167	0.993...	0.1026
100 000	9 592	8 686	9 630	0.996...	0.0959
500 000	41 538	38 103	41 606	0.9983...	0.0830
1 000 000	78 498	72 382	78 628	0.9983...	0.0785
2 000 000	148 933	137 848	149 055	0.9991...	0.0745
5 000 000	348 513	324 149	348 638	0.9996...	0.0697
10 000 000	664 579	620 417	664 918	0.9994...	0.0665
20 000 000	1 270 607	1 189 676	1 270 905	0.9997...	0.0635
90 000 000	5 216 954	4 913 897	5 217 810	0.99983...	0.0580
100 000 000	5 761 455	5 428 613	5 762 209	0.99986...	0.0576
1 000 000 000	50 847 478	48 254 630	50 849 235	0.99996...	0.0508

表格出自华罗庚的《数论导引》

按数学家迪厄多内的说法，高斯考虑 $(x, x+1000)$ 内的所有素数，通过大量的计算，高斯从数值表中观察到，当 x 很大时，有下面的近似性质：

$$\frac{\pi(x+1000) - \pi(x)}{1000} \text{ 正比于 } \frac{1}{\ln x}$$

数值计算表明，用 $\text{Li } x$ 来近似表示 $\pi(x)$ ，效果不错。

笔者相信如果不是高斯事先计算了大量的素数，他也不可能发现观察到素数分布密度特征的。当然，这里面也包含了高斯的数学直觉。笔者也猜想，古人搞天文计算，经常与大数打交道，不可能不熟悉对数表，有的大师更是将对数表背下来了，相信高斯就是其中之一（各种资料显示，股神巴菲特也把复利表或对数表背下来了）。这里需要指出，背对数表也有窍门，经常要与插值计算相结合，这样可以减少工作量。至于素数定理的具体证明，请参看本书附录。

素数定理的例子有些高深，可以再举一个亲切的例子，求函数乘积 $(uv)^{(n)}$ 的导数展开公式，也称莱布尼茨公式。这个公式证明不是很复杂，结论也不是很难记，不知大家有没有算过？

$$\begin{aligned}(uv)' &= u'v + uv' \\(uv)'' &= u''v + 2u'v' + uv'' \\(uv)''' &= u'''v + 3u''v' + 3u'v'' + uv''' \\&\dots\end{aligned}$$

如果你有耐心，计算次数比较多，如 3 次、4 次、5 次，合并结果中的同类项后，你会发现两个规律：

1. 各项系数是二项式系数（杨辉三角）；

2. 函数 u, v 的导数阶数之和都为 n 。

如果你记忆力足够好或高中学的扎实，你会立刻发现这很像二项式展开式，所以你可以大胆的猜想一般结果是：

$$(uv)^{(n)} = \sum_{k=0}^n C_n^k u^{(k)} v^{(n-k)}$$

这就是后来的莱布尼茨的公式。你可以尝试用数学归纳法严格证明它。这个例子说明并不是所有的数学定理都隐藏的很深，很难发现规律，数学有时候也很简单和自然。

当然，上面的求导公式例子比较简单，计算量比较小。其它问题可就不一定了，比如微分方程 $y'' + p(x)y' + q(x)y = 0$ ，笔者经常发现很多人询问这个方程的解是什么，求解答等等。估计很多人是在科研过程中遇到这个问题。

数学系的显然知道，上面的二阶变系数微分方程通常没有“初等解”，所以很多人一厢情愿地想找到某个解的想法泡汤了。然后，很多数学系的人对此问题就不管了，反正没有初等解。

其实没有初等解不代表没有解，方程 $y'' + p(x)y' + q(x)y = 0$ 的解也不是解不出来。不过你得换一种思维习惯，接受“级数解”。诚然，初等函数的范畴太狭窄了，而非初等函数又似乎神秘莫测。不过，古人想出了一招，可以研究相当多的非初等函数，那就是幂级数和三角级数。很多非初等函数是可以表示成幂级数的，比如大名鼎鼎的椭圆函数。级数，数学分析教学中被低估了的强大工具。实际上当 $p(x)$ 和 $q(x)$ 满足某些解析条件时，方程 $y'' + p(x)y' + q(x)y = 0$ 的解可以表示成下面幂级数形式

$$y = \sum_{n=0}^{\infty} c_n (x - x_0)^n$$

将这个幂级数解代回原方程，然后得到系数 c_n 之间的递推关系，再由初值条件可以一一算出，于是求出了解。具体过程详见《数学物理方法》教材，这里不多说了。这种方法非常实用，就是计算量比较大（因为 $p(x)$ 和 $q(x)$ 也要展开成幂级数）。

幂级数方法也可以处理一些非初等积分，例如根据勒让德等人在椭圆积分方面的研究结果，第一类椭圆积分的标准型为

$$E(k) = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{1-k^2 \sin^2 \varphi}}$$

其中参数 $0 < k^2 < 1$. 将被积函数对 k^2 做二项级数展开, 再对 φ 逐次积分, 计算得

$$E(k) = \frac{\pi}{2} \left(1 + \frac{1}{4}k^2 + \frac{9}{64}k^4 + \frac{25}{256}k^6 + \cdots \right)$$

注意, 上面得到的是精确解, 而不是数值积分中的近似解。实际应用中也不一定全用幂级数, 也可以使用三角级数, 或者其它特殊函数。记得有本名著叫《特殊函数概论》, 有兴趣的读者可以去看一看。

据笔者了解, 有人排斥“硬算”还有一个莫名的理由----“不体现数学美”。在这些人看来, 只有简洁优美的数学证明才值得看。而与硬算有关的东西, 哪怕是证明, 通常很暴力, 甚至 ugly! 这种想法千万要不得 (每个学科皆如此), 除非天赋异禀 (如哈代和狄拉克), 否则这样的人或者没有科研经验, 或者本身不能给出证明, 却始终幻想“存在”某条“数学捷径”。数学史一再证明了, 很多优美深刻的数学定理的“第一个证明” (优先权), 通常不完美, 有时也不是很严格, 甚至很 ugly!

为了说明这一点, 举一个笔者见过的典型例子-----Fourier 级数 (够深刻优美吧) 的“发现”。注意, 笔者强调的是发现, 而不是证明。下述的形式推导, 本质上是一系列很暴力的硬算。

当年数学家傅里叶通过物理中的研究热传导现象, 提炼出一个偏微分数学方程:

$$\begin{cases} \frac{\partial^2 T}{\partial x^2} = k^2 \frac{\partial T}{\partial t} \\ T(0, t) = T(\pi, t) = 0, t > 0 \\ T(x, 0) = f(x), 0 < x < \pi \end{cases}$$

除偏导数方程外, 另外两个是边值条件和初值条件。 k^2 是热传导系数, 通常不是整数。方程对应物理模型是一根长度为 π 的细长杆, 两端温度为 0, 侧面绝热。这里不严格讨论物理过程和方程是如何建立的 (可参考《数学物理方法》教材), 只探讨傅里叶当年是如何来求解这个方程的梗概, 方法非常经典 (也很暴力),

即“分离变量法”：傅里叶假设方程的解可以分离变量

$$T(x, t) = X(x)Y(t)$$

然后代入原方程得到 $X''(x)Y(t) = k^2 X(x)Y'(t)$ ，于是

$$\frac{X''(x)}{k^2 X(x)} = \frac{Y'(t)}{Y(t)} =: -\lambda$$

显然，上面方程若恒等，只能等于常数 $-\lambda$ （这里常数 $\lambda > 0$ ）。对 $\lambda = 0$ 和 $\lambda < 0$ 两种情况可以排除，读者可自行推导一下。

记 $\lambda := \mu^2$ ，其中 $\mu > 0$ 。考虑常微分方程

$$X''(x) + \mu^2 k^2 X(x) = 0, \text{ 边界条件 } X(0) = 0, X(\pi) = 0$$

上面方程的通解是 $X(x) = A \cos \mu k x + B \sin \mu k x$ ，代入边界条件得到 $A = 0$ ，以及 $B \sin \mu k \pi = 0$ ，由此得到 $\mu k = n$ （整数），于是常数

$$\lambda_n = \mu^2 = \left(\frac{n}{k}\right)^2$$

再求解微分方程 $Y'(t) + \lambda Y(t) = 0$ 。最终得到这样一个特解

$$T_n(x, t) = b_n e^{-\lambda_n t} \sin nx$$

傅里叶的想法是把所有满足条件的解都拿来主义，放到一起（所谓线性叠加原理），得到该问题的级数解：

$$T(x, t) = \sum_{n=1}^{\infty} b_n e^{-\lambda_n t} \sin nx$$

注意到，还有一个初始条件 $T(x, 0) = f(x)$ ，于是

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx$$

上面等式右边每一项是一个周期函数，也是奇函数，可以奇延拓到 $[-\pi, 0]$ 上去，进而按照周期延拓到整个实数轴上去。

傅里叶并没有停止脚步，他发现如果把边值条件修改一下，变成

$$\frac{\partial T}{\partial x}(0, t) = \frac{\partial T}{\partial x}(\pi, t) = 0$$

同样的方法，能得到一个含有余弦级数的偶函数解

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx$$

考虑到一个函数 $f(x)$ 能表示成一个偶函数与奇函数之和

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2}$$

傅里叶提出一个极为大胆的猜想：

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

这就是名震千古的 **Fourier 级数**！

接下来需要确定系数 a_n, b_n ，可以采用欧拉的积分方法，也就是现行教材中利用三角函数积分的正交性标准方法。其中系数 a_n, b_n 由以下积分确定

$$\begin{aligned} a_0 &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) dt \\ a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt dt \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt dt \end{aligned}$$

严格来说，上面的 **Fourier 级数** 等式未必成立。通常关于 **Fourier 级数** 的收敛性有很多判别定理，如狄利克雷条件等。站在现代角度评估，傅里叶当年的想法是形式推导，外加一堆蛮力硬算（当然，后人们补充了若干证明，使之成为严格的数学理论）。但正是傅里叶的以上很暴力、不严格、甚至异想天开的想法，开辟了一个新的数学天地。在笔者看来，这些想法不比当今教材中的严格证明逊色，甚至有过之而无不及。

傅里叶也考虑了函数 $f(x)$ 定义在实数轴上的情况，并且用他的不严密、很暴力的形式推导得到了 **Fourier 变换**（用现代数学语言描述）

$$\mathcal{F}[f(t)] = \hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt$$

与 **Fourier 逆变换**

$$f(t) = \mathcal{F}^{-1}[\hat{f}(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega t} d\omega$$

几百年来，**Fourier 变换理论**（习惯称为 **Fourier 分析**）仍然是数学领域的热点，同时也获得了巨大的应用成功。**Fourier 级数** 的相关理论是很深刻和优美的，但谁能想到其发现过程如此暴力。客观说，分离变量法还算相对不那么暴力的方法，数理方程中还有更暴力的方法。

笔者后来发现原来《新讲》第三册中也给出了上面的分离变量法，得到了

Fourier 级数表达式，不过采用的例子是弦振动方程，而不是热传导方程，但方法大同小异。可惜的是，也许出于教材写法的限制，张老师没有从创立者的角度来写文章，笔者最初读起来没有什么太多激动感。上面关于 Fourier 级数发现过程是在《古今数学思想》这套书中看到的。

多年后笔者回想，为什么同样的文章读起来没有热血的感觉呢？为什么小孩子对很多大人习以为常的事情那么好奇，觉得不可思议，大人们却无动于衷呢（可能是大人们知道太多了）。在数学中进行换位思考，假设自己处于创立者的角度，换个角度看问题，就别有洞天了！不仅仅是数学领域，同样的问题，不同的思考与解读，结果就不同了。

最后举一个例子，这又是一个经典的硬算例子，出自数学大师高斯之手，笔者通过扩充写成文章，收录进笔者编辑整理的《高斯经典文章及相关数学工作汇编》一书（有人已经上传到新浪共享）。

算术-几何平均数列与椭圆积分

SCIBird

谨以本文纪念一代数学大师高斯

摘要：本文主要以《高斯全集》第三卷所提供的材料为基础扩充整理而成，尝试用相对通俗和现代的语言来解释下高斯在“算术-几何平均数列与椭圆积分”方面的数学工作。这篇文章难度不大，读者具有大学二年级微积分水平即可。

数学大师高斯曾经考虑过下面的数列

$$\begin{cases} a_{n+1} = \frac{a_n + b_n}{2} \\ b_{n+1} = \sqrt{a_n b_n} \end{cases}$$

其中 $a_0 = a, b_0 = b$ ($a > b > 0$)，我们不妨设称此数列为“算术-几何平均数列”，简称为 A-G 平均数列。由均值不等式，得 $a_n \geq b_n$ ，进而得到

$$a \geq a_1 \geq a_2 \geq \cdots \geq a_n \geq b_n \geq \cdots \geq b_1 \geq b$$

再由单调有界数列必有极限，知两数列必有极限。为证明两极限相等，需要一点技巧。根据“闭区间套”定理可归纳出如下猜想，

$$0 \leq a_{n+1} - b_{n+1} \leq \frac{a_n - b_n}{2}$$

上面不等式左边显然成立，将递推关系代入知右边不等式等价于 $b_n \leq \sqrt{a_n b_n}$ ，显然也成立。由“闭区间套”定理知 A-G 数列存在共同极限，记为

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n =: M(a, b)$$

高斯的主要工作就是给出了极限 $M(a, b)$ 的具体表达式

$$M(a, b) = \frac{\pi}{2G(a, b)} \quad (1)$$

其中

$$G(a, b) = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}}$$

这个 $G(a, b)$ 称为高斯型椭圆积分，也称为**第一类椭圆积分**。高斯的工作正是揭示了 A-G 平均数列与椭圆积分之间存在奇妙的联系。

在知道积分表达式 $G(a, b)$ 的前提下，我们可以证明等式 (1) 成立，过程如下：

考虑积分变量代换

$$\sin \varphi = \frac{2a \sin \theta}{(a+b) + (a-b) \sin^2 \theta}$$

容易验证当 θ 变量从 0 变化到 $\pi/2$ 时，变量 φ 也在改范围内单调递增变化。微分上式，得

$$\cos \varphi d\varphi = 2a \frac{(a+b) - (a-b) \sin^2 \theta}{[(a+b) + (a-b) \sin^2 \theta]^2} \cos \theta d\theta$$

另一方面，由 $\sin^2 \varphi + \cos^2 \varphi = 1$ 得

$$\cos \varphi = \frac{\sqrt{(a+b)^2 - (a-b)^2 \sin^2 \theta}}{(a+b) + (a-b) \sin^2 \theta} \cos \theta$$

由此可建立等式

$$d\varphi = 2a \frac{(a+b) - (a-b) \sin^2 \theta}{(a+b) + (a-b) \sin^2 \theta} \frac{d\theta}{\sqrt{(a+b)^2 - (a-b)^2 \sin^2 \theta}}$$

注意到

$$\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi} = a \frac{(a+b) - (a-b) \sin^2 \theta}{(a+b) + (a-b) \sin^2 \theta}$$

所以得到“加法公式”

$$\frac{d\varphi}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}} = \frac{d\theta}{\sqrt{\left(\frac{a+b}{2}\right)^2 \cos^2 \theta + ab \sin^2 \theta}}$$

取 $a_1 = \frac{a+b}{2}$, $b_1 = \sqrt{ab}$, 则

$$G(a, b) = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}} = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{a_1^2 \cos^2 \theta + b_1^2 \sin^2 \theta}}$$

反复迭代上面的“加法公式”，得

$$G(a, b) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{a_n^2 \cos^2 \theta + b_n^2 \sin^2 \theta}}$$

由不等式关系 $a_n \geq b_n$, 得到

$$\frac{\pi}{2a_n} < G(a, b) < \frac{\pi}{2b_n}$$

上面不等式两边对 n 取极限，利用夹逼定理，得

$$G(a, b) = \frac{\pi}{2M(a, b)}$$

这样我们就证明了前面的结论，以上证明是后人补上的。

值得注意的是我们是在已知极限的表达式前提下来证明的，知道了结果总是一个有利条件。但是，如果仅从 A-G 数列出发，则很难猜出极限 $M(a, b)$ 的积分表达式。这个积分表达式又相当整齐，很长一段时间，笔者一直猜想高斯可能是先研究积分变换的“加法公式”

$$G(a, b) = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}} = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{\left(\frac{a+b}{2}\right)^2 \cos^2 \theta + ab \sin^2 \theta}}$$

然后自然导出 A-G 数列的。这样比直接从 A-G 数列出发要容易些。

但是，高斯实际采用的是直接求解方法，利用幂级数方法从 A-G 数列出发直接导出了极限 $M(a, b)$ 的椭圆积分表达式。下面我们就介绍下高斯的杰出思路：

考察 A-G 数列和极限 $M(a, b)$ 的结构, 我们可以归纳出下述几条性质:

1. $M(a, a) = a$;
2. $M(ta, tb) = tM(a, b)$, $t \geq 0$ (齐次性);
3. $M(a, b) = M(\frac{a+b}{2}, \sqrt{ab})$, 此为函数的加法公式, 相当于说

$$M(a, b) = M(a_1, b_1) = M(a_2, b_2) = \dots$$

性质 3 涉及的这个“加法公式”极为重要, 要多留心。

高斯先考察函数 $M(1+x, 1)$ 的幂级数展开

$$M(1+x, 1) = 1 + Ax + Bx^2 + Cx^3 + \dots \quad (2)$$

由性质 3 中的加法公式, 知

$$M(1+x, 1) = M(1+\frac{1}{2}x, \sqrt{1+x}) \quad (3)$$

为去掉根号表达式, 高斯考虑变换 $x = 2t + t^2$, 代入 (3) 得

$$M(1+x, 1) = M(1+t+\frac{t^2}{2}, 1+t) = (1+t)M(1+\frac{t^2}{2(1+t)}, 1)$$

由以上各式得

$$\begin{aligned} & 1 + A(2t+t^2) + B(2t+t^2)^2 + C(2t+t^2)^3 + \dots \\ &= (1+t) + A\frac{t^2}{2} + B\frac{t^4}{4(1+t)} + C\frac{t^6}{8(1+t)^2} + \dots \end{aligned}$$

高斯将上式右边各项做二项级数展开, 并与左边比较各项系数, 通过硬算依次解得

$$A = \frac{1}{2}, B = -\frac{1}{16}, C = \frac{1}{32}, D = -\frac{21}{1024}, E = \frac{31}{2048}, \dots$$

这样就在形式上可以确定了幂级数 $M(1+x, 1)$ 的各项系数, 及其级数表达式

$$M(1+x, 1) = 1 + \frac{1}{2}x - \frac{1}{16}x^2 + \frac{1}{32}x^3 - \frac{21}{1024}x^4 + \dots \quad (4)$$

某种意义上, 我们也求出了 A-G 数列的形式极限 $M(a, b)$, 但尚不能与椭圆积分建立联系。

这时体现了高斯的耐心, 他继续考察函数 $M(1+x, 1-x)$, 由性质 3 得

$$M(1+x, 1-x) = (1-x)M(1+\frac{2x}{1-x}, 1)$$

再将 $\frac{2x}{1-x}$ 做级数展开并利用 (4), 得

$$M(1+x, 1-x) = 1 - \frac{1}{4}x^2 - \frac{5}{64}x^4 - \frac{11}{256}x^6 - \dots$$

似乎有些规律了, 高斯同时考察了该函数级数倒数的展开 (耐心的体现),

奇迹发生了！有非常整齐的结果：

$$\frac{1}{M(1+x, 1-x)} = 1 + \frac{1}{4}x^2 + \frac{9}{64}x^4 + \frac{25}{256}x^6 + \cdots \quad (5)$$

此前，根据勒让德等人在椭圆积分方面的研究成果，第一类椭圆积分的标准型为

$$E(k) = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{1-k^2 \sin^2 \varphi}}$$

其中参数 $0 < k^2 < 1$ 。将被积函数对 k^2 做二项级数展开，再对 φ 逐次积分，计算得

$$E(k) = \frac{\pi}{2} \left(1 + \frac{1}{4}k^2 + \frac{9}{64}k^4 + \frac{25}{256}k^6 + \cdots \right)$$

通过将上式与（5）式对比，高斯猜想

$$E(k) = \frac{\pi}{2M(1+k, 1-k)}$$

取 $k = \frac{\sqrt{a^2 - b^2}}{a}$ ，再利用加法公式和齐次性，得

$$M(1+k, 1-k) = M(1, \sqrt{1-k^2}) = M(1, b/a) = M(a, b)/a$$

代回 $E(k)$ 的积分表达式，可知

$$M(a, b) = \frac{a\pi}{2E(k)}$$

经过简单的变换，表示成积分形式

$$G(a, b) = \frac{1}{a} E(k) = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}}$$

其中 $G(a, b) = \frac{\pi}{2M(a, b)}$ ，这样我们就建立了 A-G 平均数列极限 $M(a, b)$ 与椭圆积分之间的联系，一个优美而深刻的表达式！

需要指出的是上面使用的技巧是“形式幂级数法”，尽管高斯在论证过程中将变量 x, t 限制在 $(-1, 1)$ 内（出于级数收敛性的考虑），但在数学上，我们目前

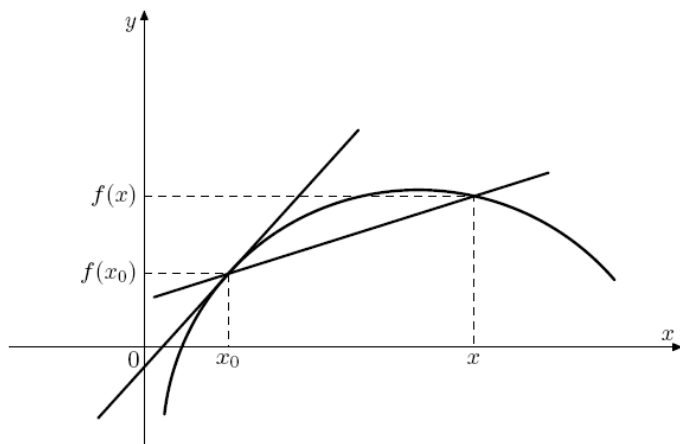
还没有证明函数 $M(1+x, 1)$ 可以展开成幂级数，故称之为“形式幂级数法”。但我们“猜出”极限的表达式后，有了明确方向，总是可以验证的（文章开头那个积分变换的证明是严格的）。

评注：笔者的一点感想

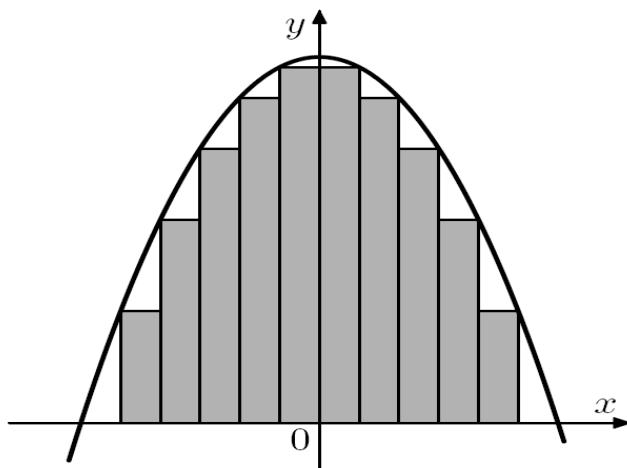
1. 知道极限表达式后，结论总是可以验证的；
2. 高斯的思维痕迹大致是：推导出关于 M 的一个等式，将 M 展开成幂级数并代入等式确定系数。通过大量的硬算和孜孜不倦刨根问底的精神找到了 A-G 数列与椭圆积分的联系；
3. 科研实践表面，现实状况多是“连猜带证（形式和探索性计算或推导）”，通过大量的试算（数值实验）来试图发现和归纳一些规律。考虑到高斯得到正确结论后尚且有这么大的计算量，之前探索性计算量可想而知，记得华老称高斯为老谋深“算”；
4. 单纯的计算意义不大，研究者要凭借自己的知识和经验来寻找“规律”，这需要一定知识储备，一定工作量，可能还要有神来之笔；
5. 单纯的计算意义不大不代表不需要计算，实际上只要多翻阅下《高斯全集》（网上有电子版，拉丁文）就会发现高斯很爱“硬算”，这里的硬算包含着大量纯计算和形式硬推导（如幂级数乘法）。总之，硬算是一门重要的基本功。

第9章 发掘几何直观

数学分析之所以上手很容易的原因之一是，其低维（1、2、3 维）情况有很好的几何直观，可以用作类比，至少理解定理的表述相对容易（证明难度是另外的问题）。所以对一些分析定理或公式感到困惑，而无法前进时，那么最好退一步，发掘一下其低维（特别是 2 维）情况的几何直观，这样有助于理解定理的意义。当然，也可以用低维（特别是 2 维）情况的几何直观作为例子来验证分析定理（结论）本身的合理性。这些都是很好的学习习惯。比如导数可视为切线的斜率，积分可视为求曲边梯形面积。



书上举的积分例子通常是求抛物线与 x 轴所围图形面积，这也是积分的最原始思想体现。



抛物线 $y = 1 - x^2$

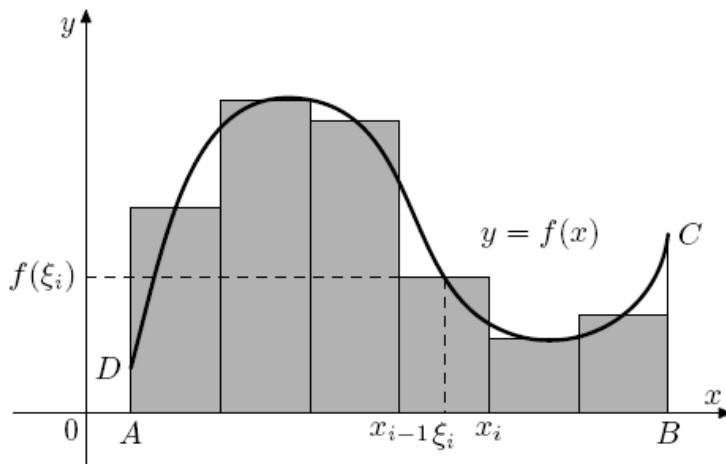
书中使用方法就是经典的“矩形分割法”（Riemann 和）。根据经验和直觉，只要矩形越来越密，则与实际面积越来越接近，这就是积分法的原始思想。

当然，能看出微分与积分两者之间是互逆运算关系并不简单（针对连续函

数)。积分一般定义为 **Riemann** 和的极限（这个极限有点特殊）

$$\int_a^b f(x)dx = \lim_{\max\{\Delta x_i\} \rightarrow 0} \sum f(\xi_i)\Delta x_i$$

这个极限与 **Riemann** 和的分割方法选取无关。

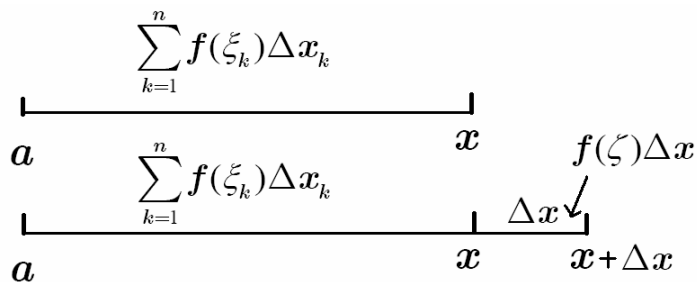


Riemann 积分示意图

换个视角，考虑变上限积分

$$\int_a^x f(t) dt$$

如下图，对区间 $[a, x + \Delta x]$ 上的分割，求 **Riemann** 和实际上只是在 $[a, x]$ 的 **Riemann** 和基础上增加了 $f(\zeta)\Delta x$ ，所以两个 **Riemann** 和的差商的极限是 $f(\zeta)$ ，即 $\Delta x \rightarrow 0$ ，自然有 $\zeta \rightarrow x$ ，于是 $f(\zeta) \rightarrow f(x)$ （这里利用了函数的连续性）。



上述角度说明：在 **Riemann** 和的定义过程中，我们已经把 $f(x)\Delta x$ （开始就可以取 $\zeta = x$ ）悄悄塞进去了，差商过程只不过把 $f(x)\Delta x$ 分离出来再把 Δx 消掉，自然剩下 $f(x)$ 了，因此从定积分的定义角度看：变上限定积分是 $f(x)$ 的一个原函数其实是非常自然的事情。由此得到微积分基本定理：

设 $f(x) \in R[a, b]$ ，且存在连续函数 $F(x) \in C[a, b]$ 满足 $F'(x) = f(x)$ ，即 $f(x)$ 存在原函数 $F(x)$ ，则

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

因为以上公式指出了微分运算与积分运算之间的互逆关系，所以被称作微积分基本定理。证明不难，直接利用 Riemann 积分的定义即可，任取区间 $[a, b]$ 的一个分割 $\{x_i\}$ ，并记 $x_0 = a$, $x_n = b$ ，则由拉格朗日中值定理，得

$$F(b) - F(a) = \sum_{k=1}^n F(x_k) - F(x_{k-1}) = \sum_{k=1}^n f(\xi_k) \Delta x_k$$

注意到等式右边即是 Riemann 和，取极限并利用极限的唯一性，得

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

这个证明并不复杂，一些工科教材也讲述定积分 Riemann 和的定义，但不知道为什么多数工科微积分教材没有收录进去。

不过笔者感觉上面的几何解释在理解“微分与积分两者之间是互逆运算关系”时，仍然不够自然。笔者认为，最自然的还是物理的例子---距离&速度。从物理角度理解这个公式可能比几何更自然些。

设直线上单向连续运动质点 t 时刻对应的距离函数为 $S(t)$ ，问在 $[t_1, t_2]$ 这段时间内质点运行的距离？将 $[t_1, t_2]$ 分成很多小时间段，每小段时间如此之短，对应速度可视为不变，则

$$S(t_2) - S(t_1) \approx \sum_i v(t_i) \Delta t_i \rightarrow \int_{t_1}^{t_2} v(t) dt$$

不难理解 $S'(t) = v(t)$ 关系，即

$$v(t) := \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} = S'(t)$$

以上一种比较自然地理解牛顿-莱布尼茨公式的思路。说到这里，再说明一下，物理的例子也可以视作几何直观的一种。几何大概是与物理联系最密切的数学分支了，高层次的物理也如此（如广义相对论）。

微分中值定理也有很明显的几何直观。首先是费马极值原理，对于光滑曲线 $y = f(x)$ ，其在一点 x_0 处取得极值，则有

$$f'(x_0) = 0$$

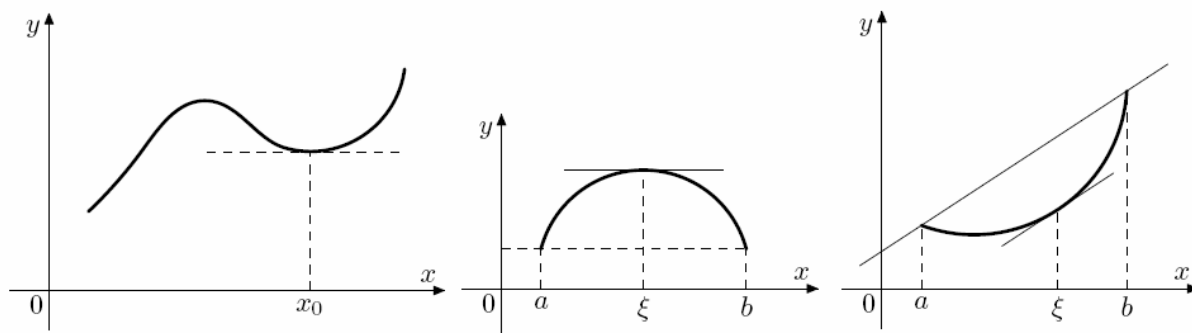
多画画图，可以发现极值点处的切线应该是水平的，切线斜率在该点两侧异号。罗尔中值定理中条件 $f(a) = f(b)$ 是为了确保在 (a, b) 内至少有一个最值，因而也是极值，所以根据费马引理 $f'(\xi) = 0$, $\exists \xi \in (a, b)$ 。这里指出，有时候 (a, b) 内非最值（极值）点 μ ，也能使得 $f'(\mu) = 0$ 。

拉格朗日中值定理可以看成经过坐标轴旋转后得到的罗尔中值定理,

$$f(b) - f(a) = f'(\xi)(b - a), \exists \xi \in (a, b)$$

将拉格朗日中值定理中的曲线写成参数形式 $x = g(t), y = f(t)$, 则得到柯西中值定理

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(\xi)}{g'(\xi)}, \exists \xi \in (a, b)$$



费马极值原理与微分中值定理示意图

很多时候, 不一定有现成的几何模型, 这就需要发挥想象力去构造一个几何模型。除去个别几何方面天赋异禀的人之外(笔者见过这种人), 这种能力需要长时间的培养与训练。比如理解函数乘积的微分公式

$$d(uv) = u dv + v du$$

上面这种性质称为导数性质或莱布尼茨性质。可以抽象出下图来理解

dv	$u dv$	$du dv$
v	uv	$v du$
	u	du

根据微分的基本思想, 略去高阶小量 $du dv$, 依照定义显然有

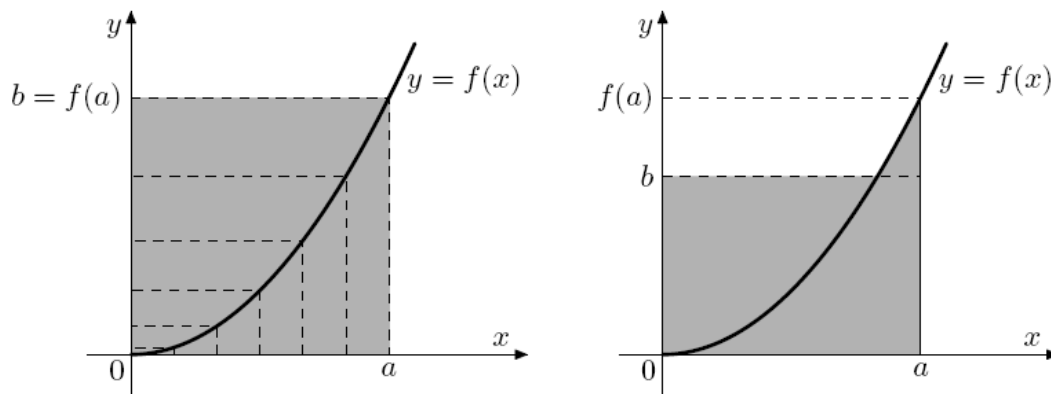
$$d(uv) = u dv + v du$$

一些抽象的定理, 画画图之后就非常直观了, 比如下面的 Young 不等式。设连续函数 $f(x)$ 在 $[0, +\infty)$ 上严格单调递增, $f(0) = 0$, $f^{-1}(y)$ 表示 f 的反函数。则当 $a, b > 0$ 时, 有

$$ab \leq \int_0^a f(x) dx + \int_0^b f^{-1}(y) dy$$

等号成立当且仅当 $b = f(a)$.

粗看 Young 不等式一头雾水, 怎么还有反函数? 但只要画画图, 发现这个定理的内容还是很直观的, 如下图 ($0 < b < f(a)$ 情形没有画出):



Young 不等式的几何直观

上面举了很多例子, 说明了几何直观在验证概念或定理时的好处, 这是一个好习惯。同时, 例子不断深入也说明几何直观是有层次的, 需要思维上一层层抽象。有的几何直观非常抽象, 常常是辛苦思索后的结果, 同时也需要一些灵感。一个经典例子是拉格朗日乘数法, 该方法是处理条件极值问题的首选方法, 实用性比较强。条件极值问题是一个很常见问题, 现实世界中的大多数问题 (无论数学, 物理, 还是其它问题) 都是有条件的。

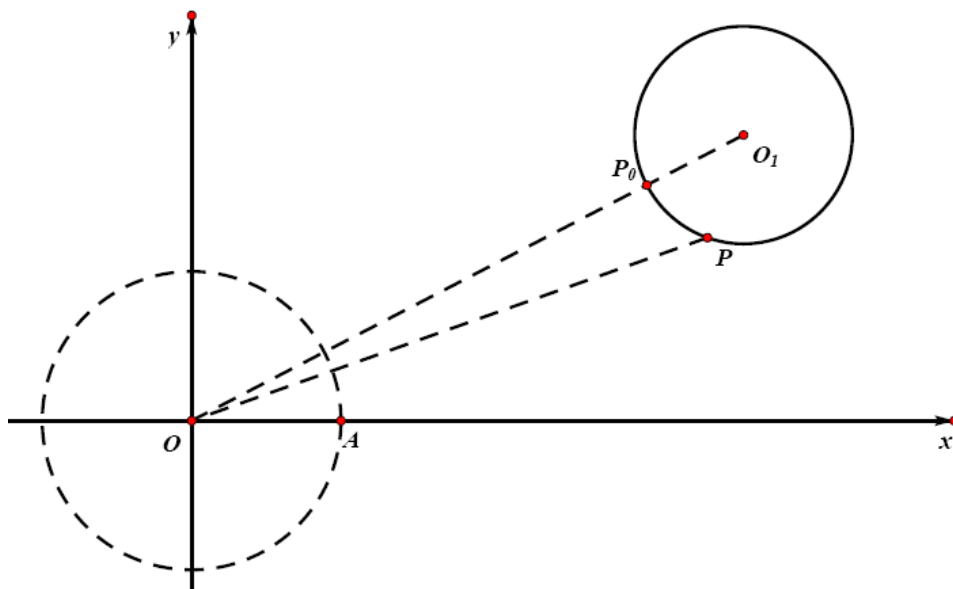
这个方法的原理并不难, 就像数学归纳法一样, 框架现成的, 直接拿来主义就行了。至于解偏导数方程求驻点, 那是另外一个问题了。最令初学者 (甚至学过数分很多年的人) 感到困惑的地方之一是----如何想出拉格朗日乘数法这么巧妙的方法呢?

理解的途径之一是几何直观。为了降低难度, 便于理解拉格朗日乘数法的几何思想, 先看一个经典的高中问题:

如下图, 给定圆 $\Gamma: (x-a)^2 + (y-b)^2 = R^2$, P 为圆 Γ 上的动点, 圆 Γ 的圆心为 O_1 , O 为坐标原点。求线段 $|OP|$ 的最大值和最小值。

这道题的难度不大, 关键在于其背后几何思想。以微积分的角度看, 记 P 点坐标为 (x, y) , 设函数 $f(x, y) = x^2 + y^2 = |OP|^2$, 则原问题可以转化为一个标准的条件极值问题:

求函数 $f(x, y) = x^2 + y^2$ 在约束条件 $(x-a)^2 + (y-b)^2 = R^2$ 下的极值。



一道高中常见的条件极值问题

本题当然可以用拉格朗日乘数法来求解。但站在几何直观角度看，却别有洞天。做一个辅助动圆 $\Gamma': x^2 + y^2 = r^2$ ，同时满足圆 Γ 与圆 Γ' 相交非空。不难看出，两圆相切时，动径 r 取得极值。

这里的动圆 $\Gamma': x^2 + y^2 = r^2$ 是为了辅助思考而引进的，所以图中用虚线表示。我们抓住最核心的两点：

1. 辅助动圆；2. 极值点对应相切点。

笔者正是抓住这两个核心想法，才想通拉格朗日乘数法的。最初的想法是考虑函数 $w = f(x, y, z)$ 在约束条件 $M: g(x, y, z) = 0$ 下的极值问题。想法也是几何方式的思考。如同上面的例子，自己考虑的是“动曲面”（也称等高面）

$$\sum_C = \{(x, y, z): f(x, y, z) = C, C \in \mathbf{R}\}$$

其中 C 是常数。显然， \sum_C 是一族曲面。

接下来考虑曲面族中，使得 $\sum_C \cap g(x, y, z) = 0$ 非空的那些曲面。假设点 $P(x_0, y_0, z_0) \in g(x, y, z) = 0$ 为极值点，使得 $f|_M(P)$ 取极值。类比上面两个圆的例子，我们自然猜想：

曲面族 \sum_C 中的曲面 $f(x, y, z) = f(x_0, y_0, z_0)$ 与 $M: g(x, y, z) = 0$ 相切换句话说，可猜想两者在点 $P(x_0, y_0, z_0)$ 处有共同的切平面。

$M: g(x, y, z) = 0$ 在 P 点的切平面方程 (点法式) 为:

$$g_x(P)(x - x_0) + g_y(P)(y - y_0) + g_z(P)(z - z_0) = 0$$

接下来如何证明曲面 $f(x, y, z) = f(x_0, y_0, z_0)$ 在 P 点也是这个切平面?

后来想到了切向量方法 (也是切平面的定义)。考虑曲面 $M: g(x, y, z) = 0$ 上, 点 $P(x_0, y_0, z_0)$ 附近的一条曲线

$$\gamma(t) = (x(t), y(t), z(t))$$

其中, 曲线 $\gamma(t)$ 是在曲面 $M: g(x, y, z) = 0$ 上, 满足 $\gamma(0) = (x_0, y_0, z_0)$, 对应 P 点切向量为:

$$\gamma'(0) = (\dot{x}(0), \dot{y}(0), \dot{z}(0))$$

设函数 $\varphi(t) = f(x(t), y(t), z(t))$, 则 $t = 0$ 是 $\varphi(t)$ 的极值点, 故 $\varphi'(0) = 0$, 即

$$f_x(P)\dot{x}(0) + f_y(P)\dot{y}(0) + f_z(P)\dot{z}(0) = 0$$

当时发现 $f(x, y, z) = f(x_0, y_0, z_0)$ 在 P 点的法向量就是

$$\{f_x(P), f_y(P), f_z(P)\}$$

而曲线 $\gamma(t) = (x(t), y(t), z(t))$ 可以取任意方向, 即切向量

$$\gamma'(0) = (\dot{x}(0), \dot{y}(0), \dot{z}(0))$$

方向可以任意。

上面证明了 $f(x, y, z) = f(x_0, y_0, z_0)$ 在 P 点的法向量与任意切向量 $\gamma'(0)$ 垂直, 这说明 $\gamma'(0)$ 也在 $f(x, y, z) = f(x_0, y_0, z_0)$ 在 P 点的切平面上。

认识到两个曲面在极值点 P 处有共同的切平面, 于是其法向量是平行的 (前面两圆的例子也支持这一点)。因此极值点 P 应满足法向量方程:

$$\nabla f(P) = k \nabla g(P)$$

为了整齐起见, 习惯上记 $\lambda = -k$, 并设拉格朗日函数

$$F(x, y, z) = f(x, y, z) + \lambda g(x, y, z)$$

显然

$$\nabla F(P) = 0 \Rightarrow \nabla f(P) = -\lambda \nabla g(P)$$

以上是从几何直观角度来思考拉格朗日乘数法, 指出构造拉格朗日函数想法的几何源头。上面的内容只是探索性讨论, 很多地方还需要严格证明。笔者最初查看了很多数学分析教材, 都没有上述想法, 等后来自己想明白后, 偶然间在卓里奇的《数学分析》上册中找到了类似想法, 而且写的更详细, 推荐大家去看一下, 这本书还是很重视几何直观的。

最后这个例子改编自一道物理竞赛决赛试题，题目有天文学背景，笔者把它整理成一篇小文章。文章的内容不深，所用的数学方法也就是余弦定理和导数求极值，但想法却极富创造性。将物理、几何与分析三者融合为一体，这里推荐一下。

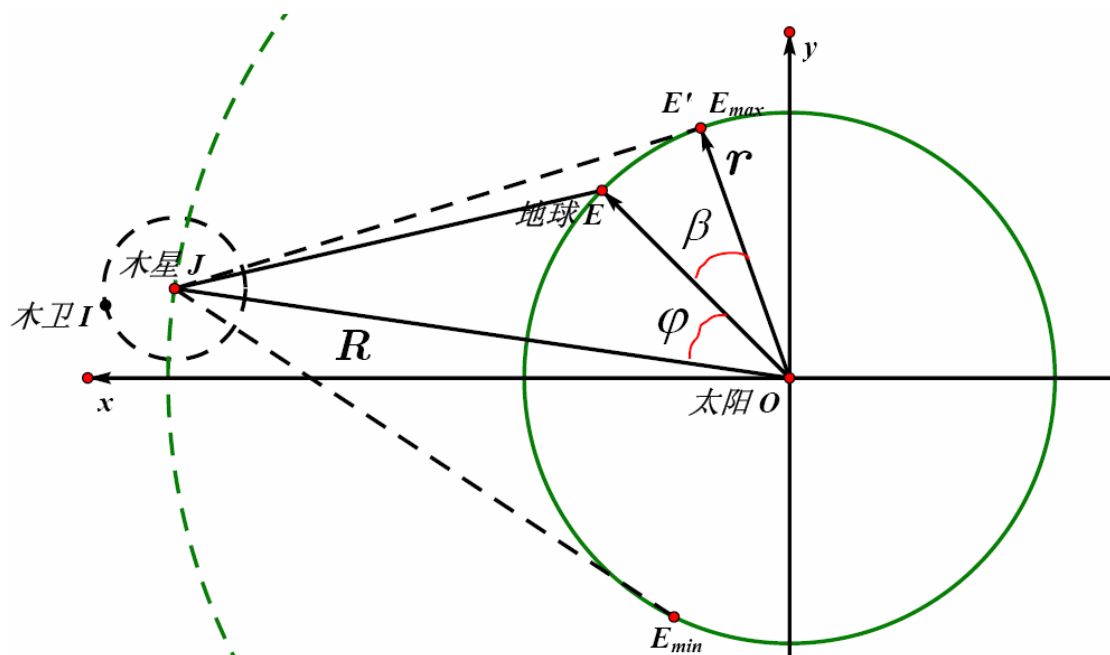
利用“木卫食”现象测光速

历史上丹麦科学家罗默（出生于 1644 年）利用在地球位于其公转轨道的不同位置处从地球上测得的木星的卫星周期，首次求出了光的传播速度。我们来回顾下罗默的杰出想法（经过改进），考虑如下简化模型：

通过观测离木星最近的一个卫星木卫 I 的运动，在地球公转轨道上各处从地球上测得木卫 I 的所有运动周期（即“木卫食”周期）值中（为方便记 $T_0 = 42.5$ 小时），最大的比 T_0 多 15 秒，最小的比 T_0 小 15 秒。

假定地球、木星的公转轨道是同一平面内的圆轨道，木卫 I 绕木星的运动轨道也是该平面内的圆轨道；地球的轨道半径为 $r = R_E = 1.5 \times 10^8 \text{ km}$ ，木星的轨道半径为 $R = R_J = 7.8 \times 10^8 \text{ km}$ ，木卫 I 的轨道半径为 $R_1 = 4.2 \times 10^5 \text{ km}$ 。要求：利用这些数据并作合理近似求出光的传播速度 c 。

注：“木卫食”，即木星卫星进入木星背影区时，使得木星卫星反射的太阳光被木星遮挡，而不能被地球上的观测者观测到的现象。



木星、地球运动示意图

假设木星、地球均做顺时针匀速圆周运动。木星 J 位置和地球 E 位置如图所示, 此时 OE 与 OJ 夹角为 φ , 利用余弦定理可知 JE 距离满足:

$$JE = L(\varphi) = \sqrt{R^2 + r^2 - 2Rr \cos \varphi}$$

设木卫 I 的真实运动周期为 τ , 这个 τ 与木星、地球公转周期相比很小。现在假定地球在 τ 时间内偏转了 β 角, 角 β 很小且为常数 (图中为显示清楚, 比例上进行了夸张处理, 且忽略了木星的移动)。

利用万有引力公式和匀速圆周运动公式, 可以求出地球和木星转动的角速度之比为

$$\omega_E / \omega_J = (R/r)^{3/2} \approx 12 = 1/k$$

于是, 地球相对于太阳 O 转动了 β 角, 则木星相对于太阳 O 转动了 $k\beta$ 角。

当木卫 I 运动到木星背后 (相对于地球) 则发生木卫 I “木卫食” 现象, 理论上通过观测木卫 I 的两次“木卫食”现象时间间隔, 可以测出其运动周期 τ 。但实际测量木卫 I 运动周期却得到一系列不同的测量值。罗默经过长时间思考这个矛盾后, 提出了一个大胆的划时代想法: “光速是有限的”。

实际上, 因为木卫 I 轨道半径相对于木星与地球距离非常小, 所以可以忽略木卫 I 运动的影响。经过真实运动周期 τ 后, 地球从位置 E 运动到位置 E' 。在光速有限的假定下, 地球公转和木星公转产生了光程差 $\Delta l = JE' - JE$ (这里的计算还需考虑木星公转影响), 因此与真实运动周期 τ 相比, 实际测量值会多出一个时间差 $\Delta t = \Delta l / c$,

光程差近似为 (下面这个等式考虑了木星公转的影响)

$$\Delta l = JE' - JE = L(\varphi + \beta - k\beta) - L(\varphi) \approx L'(\varphi)(1 - k)\beta$$

利用求导法则, 通过一些繁琐但初等的运算, 可以求出 $\cos \varphi = r/R$ 时, $L'(\varphi)$ 有最大值 (本质上是解方程 $L''(\varphi) = 0$, 求 $L'(\varphi)$ 的极值点), 此时取最大光程差。对应图中的几何位置, 即最大测量周期位置 E_{\max} 满足: $JE_{\max} \perp OE_{\max}$, 即与 JE_{\max} 圆相切。利用 $L'(\varphi)$ 为奇函数特点, 可知另一个切点 E_{\min} 为最小测量周期点。

由对称性, 可知木卫 I 实际运动周期为: $\tau = (T_{\max} + T_{\min}) / 2 = T_0$, 以及

$$\Delta T = T_{\max} - \tau = T_{\max} - T_0 = 15 \text{ s}$$

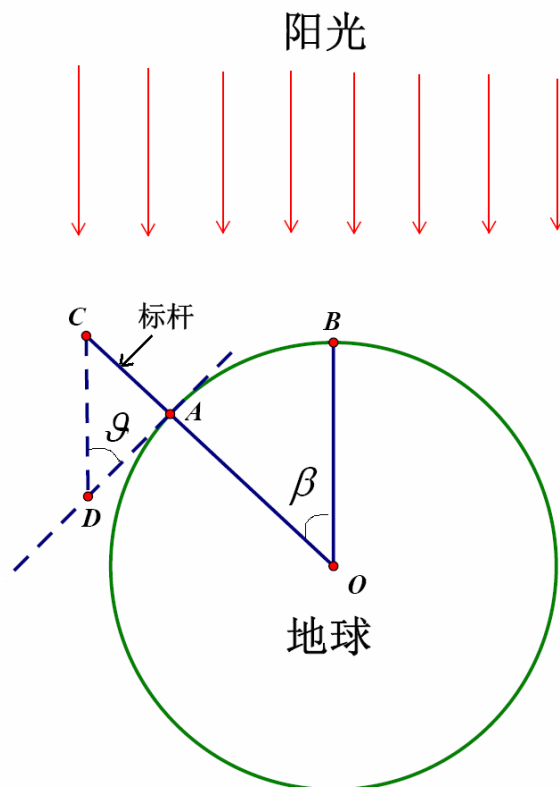
因为地球自转周期为 1 年, 所以可算出经过 T_0 时间后, 地球大约转了 $\beta = 1.75^\circ$ 。

这个角 β 非常小，所以在小角近似情况下（如上图），光程差 $L'(\varphi)(1-k)\beta$ 近似为弧长 βr （即弧 EE' ）。于是 $c\Delta T = \Delta l = \beta r$ ，可知光速约等于：

$$c = \frac{\beta r}{\Delta T} \approx 3.05 \times 10^8 \text{ m}$$

现代公认的光速值为 $c = 3.0 \times 10^8 \text{ m}$ ，考虑到罗默生活的那个年代，可见其洞察力之深刻！当然，因为地球实际运动轨道为椭圆，再加上技术原因，对木卫 I 观测误差比较大，罗默采用日地椭圆轨道的半长轴作为日地距离，测得的光速为 $c \approx 2.25 \times 10^8 \text{ m}$ ，这个结果在数量级上是正确的。重要的是他的杰出想法，笔者认为罗默这一伟大想法完全可以媲美古希腊的科学家埃拉托色尼利用太阳直射现象测量地球周长的伟大想法。

古希腊不少人相信地球是圆的，因为圆是最完美的，况且月亮也是圆的。公元前 240 年前后，在希腊的亚历山大城图书当馆长的埃拉托色尼注意到在夏至的中午，阳光可以直射到位于亚历山大城附近的小镇塞恩（地点 B）的一口枯井的井底，直立的物体没有影子，也就是说太阳正好就悬挂在塞恩城的正上方，他发现在同一天同一时间，在亚历山大城（地点 A）地面上的物体有一段很短的影子，阳光是斜射进亚历山大城的。埃拉托色尼由此萌发一个大胆的想法——估算地球周长！他考虑了如下模型（为形象，下图的比例有些夸张）：



利用太阳直射现象测量地球周长示意图

假设地球是圆的，阳光是平行光。图中 A、B 两地距离 s 已知。在 A 地立一个标杆 AC，测出中午时的影子长为 AD，则根据三角关系 $\tan \vartheta = AC / AD$ ，由此可确定 ϑ 角。由内错角和直角三角形关系可知 $\vartheta + \beta = 90^\circ$ ，于是地球周长：

$$C = \frac{s}{\beta} \times 360 = \frac{s}{90 - \vartheta} \times 360$$

评注：在写书的过程中，自己产生了一个想法，是不是单写一章应用数学，说明数学在现实世界的应用。不过因为其它章节框架已经写好了，不容易替换，于是放弃了单写一章的想法。转而选择一些应用例子，放到某些章节中（如本章）。上面两个例子也说明了一个道理，做出伟大的发现不一定要用高深的数学工具，关键是你要有想法，特别是创造性的想法。数学系的教学体系很容易固化思维，这方面数学系和物理系经常吵架。也许数学的证明是僵硬的，但其中的探索与发现却是鲜活的，这种观点是本书的基调，本书写作上也尽量做到这一点。

第10章 要有逆向思维

逆向思维或逆向思考似乎与常规逻辑不太合群，有点不走寻常路的感觉。但历史一再证明了逆向思维的重要性，其在人类认识世界的过程中往往起到奇兵的作用。数学中的逆向思维首推反证法。相信大家对反证法已经非常熟悉了，就像老朋友一样。不少定理直接证明可能有些困难，但借助反证法证明就容易多了，比如利用反证法证明素数有无穷多个。这个例子非常经典：

反证假设素数只有有限个，记为 p_1, p_2, \dots, p_n ，现在考虑数 $p_1 p_2 \cdots p_n + 1$ ，根据假设可知，它应该是合数，从而包含至少两个素因子。但是 p_1, p_2, \dots, p_n 都不是 $p_1 p_2 \cdots p_n + 1$ 的素因子，矛盾。

另一个有名的例子是： $1+2+3+\cdots+100=?$

高斯给出的解答是考虑倒排相加关系

$$\begin{array}{r} 1+2+3+\cdots+100 \\ 100+99+\cdots+2+1 \end{array}$$

于是 $1+2+3+\cdots+100=101 \times 100 \div 2=5050$ 。这里的倒排相加就是逆向思维的体现，该方法是等差数列求和方法中经典技巧。

逆向思维的例子很多，有些太寻常了，以至于没有留意。比如反函数的概念，这是拓展函数研究范围的一种方法。按笔者的想法，多项式和有理函数是熟悉的，因为可以与整数加减乘除运算作为类比。但是，对数函数、指数函数与三角函数这三类函数就不那么熟悉了，比如 $\sqrt{2}^{\sqrt{3}}$ 应该如何理解？

按照数学史，对数函数先于指数函数出现，通常定义为

$$\ln x = \int_1^x \frac{dt}{t}$$

被积函数是有理函数，我们很熟悉。通过变量代换，可以得到积分等式

$$\int_1^{xy} \frac{dt}{t} = \int_1^x \frac{dt}{t} + \int_1^y \frac{dt}{t}$$

这就是对数的乘法公式： $\ln xy = \ln x + \ln y$ 。

对数函数 $y = \ln x$ 的反函数为指数函数 $x = e^y$ ，习惯上记为 $y = e^x$ 。指数函数有加法公式 $e^{x+y} = e^x \cdot e^y$ 。这里的底数 e 这样定义，它使得

$$1 = \int_1^e \frac{dt}{t}$$

对数函数 $y = \ln x$ 导数可以直接求出来，即 $y' = 1/x$ （利用变上限积分求导

法则)。指数函数 $y = e^x$ 的导数可以间接求出来，即利用它的反函数导数，然后倒过来即可，由此可知 $(e^x)' = e^x$ 。这比直接利用导数定义要便捷一些。

借助于欧拉公式 $e^{ix} = \cos x + i \sin x$ ，我们实际上可以由指数函数来研究三角函数。当然，也可以由其它方法来研究三角函数，如利用积分反演定义

$$x := \arcsin y = \int_0^y \frac{du}{\sqrt{1-u^2}}$$

然后考虑 $x := \arcsin y$ 的反函数 $y = \sin x$ 。

对于数学家来讲，研究一个函数的反函数是一个很自然的事情，研究一种运算的逆运算也是很自然的。数学分析中的典型代表要数不定积分 \int 了。最初不定积分运算 \int 看做是导数的逆运算，即按

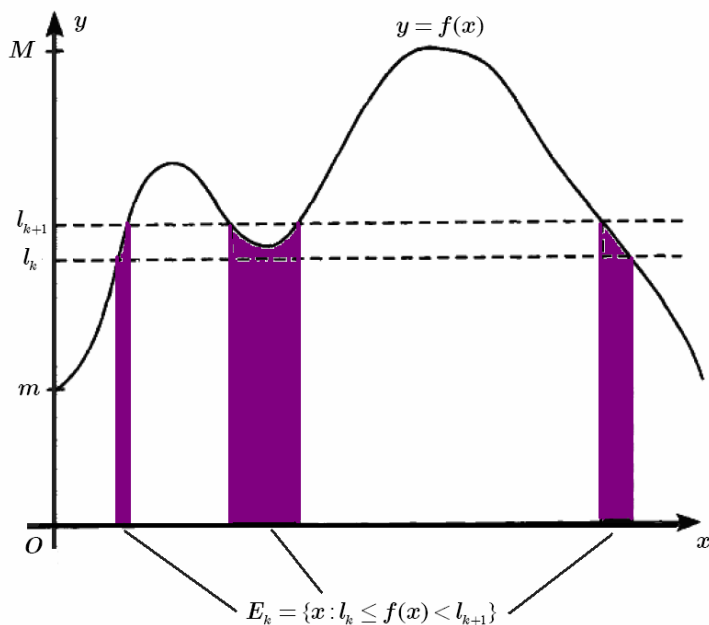
$$\int := \left(\frac{d}{dx} \right)^{-1}$$

来理解。所以有的书上称不定积分为“反导数”。这里完全可以用其它符号来表示不定积分 \int 。只是后来人们发现了牛顿-莱布尼茨公式，发现连续函数的变上限定积分就是一个原函数，所以采用了相似的符号。不过必须强调的是，不定积分 \int 表示一族函数而不是一个，即

$$\int f(x) dx = F(x) + C$$

不要漏掉任意常数 C 。

以上例子都是很自然的。类似的例子有很多，一个非平凡的例子是 Lebesgue 积分，其原始想法是分 y 轴！可用下图来说明。



Lebesgue 积分示意图

回忆一下 **Riemann** 积分的定义，其想法是分 x 轴，然后考虑下面和式（称为 **Riemann** 和）：

$$\sum_{i=1}^n f(\xi_i)(x_i - x_{i-1})$$

最后令 $\max\{x_i - x_{i-1}\} \rightarrow 0$ 时取极限，这个极限就是对应的定积分。

但是当函数 $f(x)$ 剧烈震荡时，上述 **Riemann** 和的极限可能不存在。即可以控制 $\{x_i\}$ ，但无法控制 $f(\xi_i)$ 。此时，传统的分 x 轴方法遇到了困难（存在非 **Riemann** 可积函数）。

作为逆向思维，勒贝格的想法是分 y 轴（即控制 $f(\xi_i)$ ），过程如下：

设 $m \leq f(x) \leq M$ ，然后分割函数的值域 $[m, M]$ ，得到分割点

$$m = l_0 < l_1 < \cdots < l_n = M$$

再记集合

$$E_k = \{x \in [a, b] : l_k \leq f(x) < l_{k+1}\}$$

上述点集的 **Lebesgue** 测度记为 $m(E_k)$ 。考察和式

$$\sum l_k m(E_k)$$

如果 $\max\{l_{k+1} - l_k\} \rightarrow 0$ 时，上述和式的极限存在，则记为

$$\int_E f(x) dx$$

称作 **Lebesgue** 积分。这种积分对于 **Riemann** 积分是兼容的，即 **R** 可积函数也是 **L** 可积的，且两者相等。上面的 **Lebesgue** 积分定义与 **Riemann** 积分的定义基本是一致的，但关键的地方在于如下点集（引入了 **L** 测度）：

$$E_k = \{x \in [a, b] : l_k \leq f(x) < l_{k+1}\}$$

之所以借助 **Lebesgue** 测度 $m(E_k)$ ，是因为对于复杂函数来讲，点集 E_k 未必能表示成有限个区间的并集。关于 **Lebesgue** 测度的想法，读者可参考本书第 6 章的内容。

笔者当初自学 **Lebesgue** 积分时，内心有一个困惑。因为一维积分可以看做是求图形面积，即求

$$I = \iint_D dx dy$$

站在求面积（测度）角度看，分 x 轴与分 y 轴地位是对等的，为什么后者对应的 **Lebesgue** 积分更强大呢。这个问题困扰了自己很久，后来发现原因：

通常的积分是针对函数 $f(x)$ 的，而函数的定义要求对每一个 x ，有惟一的数值 $f(x)$ 与之对应。所以分 x 轴的时候， $f(x)$ 的表现不那么糟糕。但是，如果分 y 轴，那么根据函数定义，同一个数值 y ，可能有一堆 x 与之对应，此时已经不存在函数关系 $x = g(y)$ 了。这一解释说明，站在函数积分的角度看，分 x 轴与分 y 轴并不对等。细想想确实如此，对于形态好的函数，分 x 轴与分 y 轴效果一样，对于一些形态比较糟糕的函数，需要分 y 轴了（同时借助 L 测度理论）。

需要强调的是：勒贝格本人的看法是，Lebesgue 积分的真正价值不是使得一些 R 不可积函数变得可积了， L 积分的真正价值是“大大放宽了积分与极限之间的运算条件，如控制收敛定理”。

（控制收敛定理）设有限可测集 E 上的可测函数序列 $f_n(x)$ 满足

$$|f_n(x)| \leq M, \forall x \in E, \forall n \in \mathbf{N}$$

其中 $M > 0$ 为常数。如果 $\lim f_n(x) = f(x)$ 几乎处处成立，那么 $f(x)$ 也是 L 可积的，且可以在积分号下求极限：

$$\lim_{n \rightarrow \infty} \int_E f_n(x) dx = \int_E f(x) dx$$

在整理本书的过程中，一些数学思考往事浮现出来。自己大一时学完微积分后，产生了一个疑问：如果只知道某个连续函数的 Fourier 系数是 a_n, b_n ，能否求出这个连续函数 $f(x)$ 。对于大一的工科生来说，Fourier 级数算是工科微积分中最难的部分了，也不大可能从这本书中找到疑问的答案。

不过，自己还是有些想法的，假设连续函数的 Fourier 级数是收敛的，猜想

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

自己相信这个猜想是对的，但当时无法给出证明。

与数学系教学不同，工科生通常和连续函数与可导函数打交道，不太关注函数可积性问题。所以自己当时没有考虑过连续函数的 Fourier 级数是否处处收敛的。现在知道了卡尔森定理：

连续函数 $f(x)$ 对应的 Fourier 级数几乎处处收敛到 $f(x)$ 自身。

这个看似平常的结论难度却是菲尔兹级别的。人们构造了具体的反例，即构造一个连续函数，其 Fourier 级数在某个点是发散的，反例的具体构造还是很麻烦

的。不过利用泛函分析中的共鸣定理，可以给出一个存在性反例的证明，感兴趣的读者可以查看下江泽坚的《泛函分析》。

真正解开自己心中疑惑的还是老朋友《数学分析新讲》（大二相见），在第三册中讲到了费耶和 $\sigma_n(x)$ 。记 $f(x)$ 对应的 Fourier 级数的部分和

$$S_m(x) = \frac{a_0}{2} + \sum_{n=1}^m a_n \cos nx + b_n \sin nx$$

然后再考虑部分和的算术平均和（C-1 求和法）

$$\sigma_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} S_k(x) = \frac{S_0(x) + S_1(x) + \cdots + S_{n-1}(x)}{n}$$

新讲中证明了费耶定理：

若 $f(x)$ 是连续函数，则费耶和 $\sigma_n(x)$ 一致收敛于 $f(x)$ 。

而新讲第一册中有一个不起眼的结论：若 $\{\alpha_n\}$ 是无穷小序列，则

$$\beta_n = \frac{\alpha_1 + \cdots + \alpha_n}{n}$$

也是无穷小序列。由此可知，设 $\{S_n\}$ 是一个收敛序列，即 $S_n \rightarrow A$ ，则 $\{S_n - A\}$ 是无穷小序列。由此很容易推出：

$$\frac{S_1 + \cdots + S_n}{n} \rightarrow A$$

回到自己的猜想。因为假设连续函数 $f(x)$ 的 Fourier 级数是收敛的，再根据上述结论可知若 $S_n(x) \rightarrow g(x)$ ，则有

$$\frac{S_0(x) + S_1(x) + \cdots + S_{n-1}(x)}{n} \rightarrow g(x)$$

再由费耶定理知 $g(x) = f(x)$ （极限的唯一性）。这就证明了自己当初的猜想。

大一做的 Fourier 级数方面的试题，基本都是计算各种初等函数的 Fourier 级数展开表达式，也就是算 Fourier 系数。所以自然产生一个想法：能不能已知 Fourier 系数反推出 $f(x)$ ，这个思考过程也算逆向思维吧。

逆向思维的另一个代表例子是猜想广义 Riemann-Lebesgue 引理：

设黎曼可积函数 $f(x), g(x)$ ，这里 $g(x+T) = g(x)$ 是周期函数。则

$$\lim_{n \rightarrow \infty} \int_0^T f(x) g(nx) dx = \frac{1}{T} \int_0^T g(x) dx \int_0^T f(x) dx$$

将整数 n 换成一般的实数 λ ，结论仍然成立。

上面这个等式的右边不太容易直接猜到，但如果借助逆向思维，则还是可以推出来等式右边是何种形式的。

首先，类比 Fourier 级数中的结论

$$\lim_{\lambda \rightarrow \infty} \int_a^b f(x) \sin \lambda x \, dx = \lim_{\lambda \rightarrow \infty} \int_a^b f(x) \cos \lambda x \, dx = 0$$

则自然猜想出下面的等式（当然不一定无条件成立）：

$$\lim_{n \rightarrow \infty} \int_0^T f(x) g(nx) \, dx = 0$$

令 $f(x) \equiv 1$ ，上面等式恒成立，则有

$$\int_0^T f(x) g(nx) \, dx = \int_0^T g(nx) \, dx = \frac{1}{n} \int_0^{nT} g(t) \, dt = \int_0^T g(t) \, dt$$

上式的极限是 0，因此只能是

$$\int_0^T g(x) \, dx = 0$$

考虑 $g(x) = \sin x$ 这个特例，显然满足题意。但对一般的周期函数 $g(x)$ ，在一个周期内的积分不一定为 0（如 $g(x) = 1 + \sin x$ ），通常是

$$\int_0^T g(x) \, dx \neq 0$$

不过，借助几何直观，我们发现将周期函数 $g(x)$ 适当向上（或向下）平移 C 个单位之后，得到的新周期函数 $\varphi(x) = g(x) + C$ 满足周期积分为 0，即

$$0 = \int_0^T \varphi(x) \, dx = \int_0^T g(x) \, dx + C \int_0^T dx$$

由此得到

$$C = -\frac{1}{T} \int_0^T g(x) \, dx$$

因为我们“猜想”

$$\lim_{n \rightarrow \infty} \int_0^T f(x) \varphi(nx) \, dx = 0$$

所以自然进一步猜想

$$\lim_{n \rightarrow \infty} \int_0^T f(x) g(nx) \, dx = \lim_{n \rightarrow \infty} \int_0^T f(x) (\varphi(nx) - C) \, dx = \frac{1}{T} \int_0^T g(x) \, dx \int_0^T f(x) \, dx$$

搞数学研究也类似，有些问题的关键不是数学工具的高深或证明技巧难度大，而是有没有想法。法国数学家勒让德研究了椭圆积分几十年，也写了三大本专著，如研究下面的第一类椭圆积分

$$u = \int_0^z \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

该积分在求椭圆弧长时出现，故而得名。

不过，勒让德在研究椭圆积分时缺少了逆向思维，没有想到研究其反函数 $z = \operatorname{sn} u$ （椭圆函数），从而与椭圆函数奠基人的历史机会擦肩而过。而高斯和阿贝尔想到了研究椭圆积分的反函数，成为了椭圆函数领域的奠基人。

其实利用积分反演（反函数）研究函数并不是什么少见的事情，众所周知

$$u = \int_0^z \frac{dx}{\sqrt{1-x^2}} = \arcsin z$$

定义了正弦函数 $z = \sin u$ 。这其实是椭圆积分中 $k=0$ 对应的退化情形。技术上，研究椭圆函数比研究椭圆积分简单，而且有很多新的发现，比如椭圆函数的双周期性等。庞加莱发现自守函数的过程也利用了逆向思维，即研究富克斯方程的反函数，这方面具体内容将在下一章中讨论。

听一位数学老师说过，每次有人来做报告他都尽可能去听，但 10 场报告也就能听明白 3 场，为什么还去听呢？这位老师的解释是，听的是 idea！也许报告人不经意间某句话，某个 idea，就能给予你意想不到的启示。

不仅仅是数学分析这门基础课，一些高深的数学领域也需要用到逆向思维，比如费马大定理的证明。

Fermat's Last Theorem，简称 FLT，国内一般翻译成“费马大定理”（其实叫费马猜想更符合实际），其内容如下：

当自然数 $n > 2$ 时，不定方程 $x^n + y^n = z^n$ 没有满足 $xyz \neq 0$ 的整数解。

这一定理的表述很初等，内容也不难理解，而且乍一看很像一道奥数题。但其证明却十分艰难，历经 350 多年才由英国数学家 Wiles 所证明，证明涉及椭圆曲线、模形式和伽罗瓦表示等许多现代数学工具，过程十分曲折，是一个 20 世纪的数学证明。

自然数 n 可以表示成 $n = mp$ （这里 p 是一个素数），于是费马方程写成

$$(x^m)^p + (y^m)^p = (z^m)^p$$

可以看出，只需要证明 $n=4$ 和 $n=p$ （这里 p 是一个奇素数）这两种情况费马

大定理为真即可。前者已经由费马本人利用无穷递降法所证明，于是只需证明 $n = p$ 这一情况为真。

设 p 是一个奇素数，于是可以将费马方程写成更对称的形式：

$$u^p + v^p + w^p = 0$$

其中 $u = x, v = y, w = -z$ 。这与原命题是等价的。

在证明费马大定理过程中，一个关键性的突破就是应用了逆向思维。

假设费马大定理不成立，取这样一组非零互素整数解 (a, b, c) ，满足：素数 $p \geq 5$ ， $a^p + b^p + c^p = 0$ ， b 是偶数， $a = 4k + 3$ 。由此定义出一条特殊的椭圆曲线：

$$E_{a,b,c} : y^2 = x(x - a^p)(x + b^p)$$

这条曲线现在一般称为 Frey 曲线。

数学家 Frey 猜想这条曲线是如此的“怪异”，以至于它不应该存在，所以用它可以证明费马大定理，这真是一个大胆而神奇的想法！这里的“怪异”用数学术语说是不具有某种数学性质 *modular*。

Frey 本人尝试证明 $y^2 = x(x - a^p)(x + b^p)$ 不具有数学性质 *modular*，但是他的证明有严重的漏洞。这个工作实际上是由数学家 Ribet 在 Serre 工作的基础上给出了严格数学证明。

数学家 Wiles 正是得知 Ribet 的工作后，苦心钻研，面壁 9 年，证明了一大类（半稳定情况）椭圆曲线都具有数学性质 *modular*，这一大类椭圆曲线足够多，恰好包含 Frey 曲线，由此证明了费马大定理。Wiles 证明工作中的一个推论是如下定理：

（Wiles）

设 A, B 是非零互素整数，满足 $AB(A + B) = 16m, m \in \mathbb{Z} (m \neq 0)$ ，则曲线

$$E_{A,B} : y^2 = x(x - A)(x + B)$$

具有数学性质 *modular*。

费马大定理的证明：如上，取 $A = a^p, B = b^p$ ，并注意到 b 是偶数且素数 $p \geq 5$ ，因此 $AB(A + B)$ 是 32 的倍数。于是 Frey 曲线 $E_{a,b,c} : y^2 = x(x - a^p)(x + b^p)$ 包含在曲线族 $E_{A,B}$ 中，因此具有数学性质 *modular*，这与 Ribet 的结论矛盾！这

一矛盾说明，Frey 曲线不存在，于是费马大定理成立。

实际上，费马大定理的证明是 Ribet 和 Wiles 等多人工作的合力结果。但 Wiles 的工作难度更大，意义也更深远，因此才将菲尔兹特别奖和沃尔夫奖颁给了 Wiles. 上面的论述也表明了费马大定理实际上是用反证法间接证明的。

评注：在这个越来越浮躁的社会中，逆向思维是一种难得的品种。但应该避免走极端，钻牛角尖。逆向思维不是对着干，而是一种综合性的思维，一种创造性的思维。笔者比较推荐逆向思维中的“倒推法”，就是猜想 Riemann-Lebesgue 引理时用的那个方法，非常实用。

为了完整起见，简单介绍下数学性质 *modular* .

modular 一般翻译成“模性”，有好几种等价的数学表述，这里选择一种分析味道明显的数学表述，涉及到复变函数。

先引入一些概念和符号。记 Poincare 上半平面 $H = \{z \mid z = x + iy, y > 0\}$ ，而 $\Gamma_0(N)$ 表示如下集合

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{Z}, ad - bc = 1, c \equiv 0 \pmod{N} \right\}$$

定义分式变换（默比乌斯变换）

$$Mz := \frac{az + b}{cz + d}, \quad M \in \Gamma_0(N)$$

（模函数）

上半平面 H 上的亚纯函数 $f(z)$ 称为具有水平 N 的“模函数”，如果其满足：

$$f\left(\frac{az + b}{cz + d}\right) = f(z)$$

其中， $\forall M \in \Gamma_0(N), \forall z \in H$. 同时要求 $f(z)$ 在“尖点”处也是亚纯的。

根据定义，容易知道 $f(z+1) = f(z)$ ，这说明模函数是周期函数的一种推广，所以 $f(z)$ 存在 Fourier 级数展开

$$f(z) = \sum_{n=-\infty}^{+\infty} a_n e^{2\pi i n z}$$

称 $f(z)$ 在“尖点”处亚纯，如果对应的 Fourier 级数展开式为

$$f(z) = \sum_{n=0}^{+\infty} a_n e^{2\pi i n z}$$

也就是 $a_n = 0, n = -1, -2, -3, \dots$ 。

(椭圆曲线)

椭圆曲线指下面一类非奇异有理系数三次代数曲线

$$y^2 = Ax^3 + Bx^2 + Cx + D$$

其中，系数 A, B, C, D 为有理数且 $A \neq 0$ 。同时这些系数还必须使得等式右边没有重根（复数域上）。反应在几何直观上，就是三次曲线不自交或者没有尖点（非奇异性）。

通过线性有理变换，还可以将椭圆曲线变成下面的规范型

$$y^2 = 4x^3 + g_2x + g_3$$

其中系数 g_2, g_3 为整数，且使得等式右边没有重根。

(数学性质 modular)

称椭圆曲线 $E: y^2 = Ax^3 + Bx^2 + Cx + D$ 具有数学性质 modular，如果存在具有相同水平 N 的非常数模函数 $f(z)$ 和 $g(z)$ ，使得

$$f(z)^2 = Ag(z)^3 + Bg(z)^2 + Cg(z) + D.$$

简单说，性质 modular 说的是椭圆曲线可以被模函数参数化（更严格的说法是单值化）。就像单位圆 $x^2 + y^2 = 1$ 可以被三角函数 $x = \sin z, y = \cos z$ 参数化一样

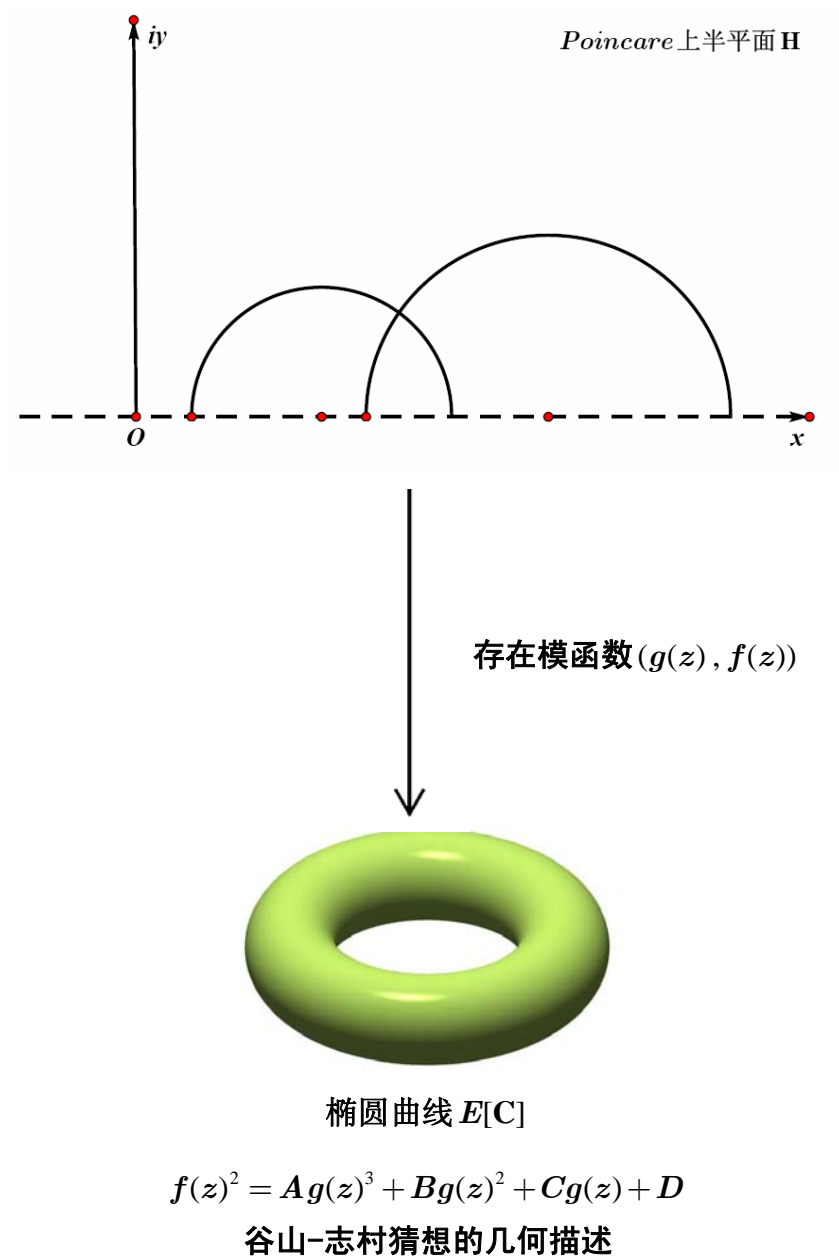
$$\sin^2 z + \cos^2 z = 1$$

我们在有关费马大定理的科普书中，经常看到“谷山-志村猜想”这个数学名词，这个猜想严格的表述如下：

谷山-志村猜想：有理域 \mathbb{Q} 上的椭圆曲线均可以用模函数来参数化。

椭圆曲线和模函数被认为是两个不同的数学世界，而“谷山-志村猜想”在这两个世界中建立起桥梁。能被模函数参数化的椭圆曲线称作“模椭圆曲线”。

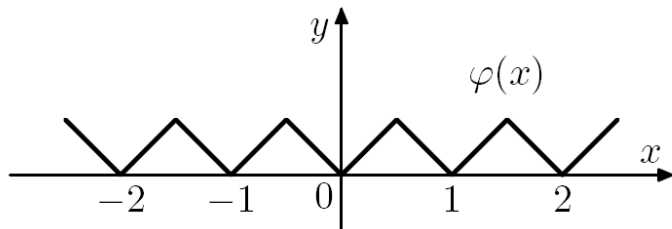
“谷山-志村猜想”可以用下面的几何图形加以描述：



第 11 章 数学中的构造方法

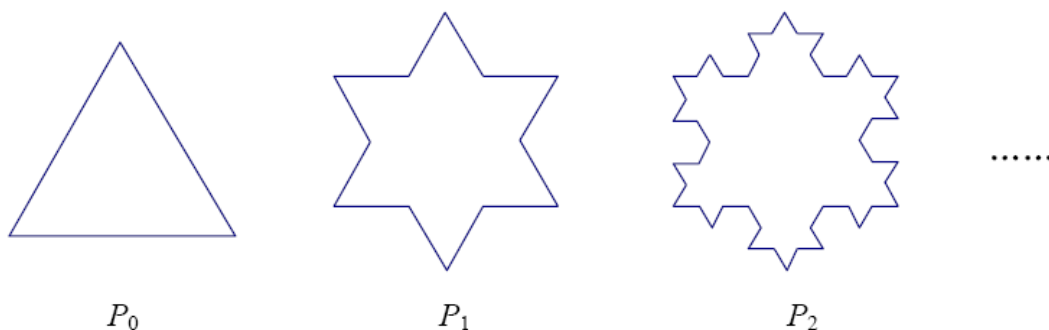
本章谈一谈数学中的构造方法。构造法属于数学学习中的较高层次，某种意义上属于数学创造力范畴。因为涉及创造力，所以很难用语言来具体形容，更不可能给出固定的模式来复制，有点只可意会，不可言传的感觉。笔者只能尝试多给出几个具体例子，通过案例来介绍数学中的构造方法。

笔者在写本文时，很自然想到了“反例”。确实，大部分反例都是采用构造法举出的。在数学分析中，最具代表性的要数“处处连续处处不可导函数”这个经典反例。记 $\varphi(x)$ 表示 x 与离它最近整数的距离，则 $\varphi(x+1) = \varphi(x)$ 。建议动手画画图，发现这是一个“锯齿函数”



令 $f(x) = \sum_{n=0}^{\infty} \frac{\varphi(4^n x)}{4^n}$, $x \in \mathbb{R}$. 此函数级数是一致收敛的，因此定义了一个连续函数，这个函数处处连续处处不可导。这个例子的具体证明见本书附录。

当然，站在几何直观角度，可以举出雪花曲线这个著名例子，这种曲线处处不存在切线。考虑下面的迭代过程：



具体操作是将每条边三等分，然后以中间的线段为边长，向外做出一个小等边三角形。如此反复操作……

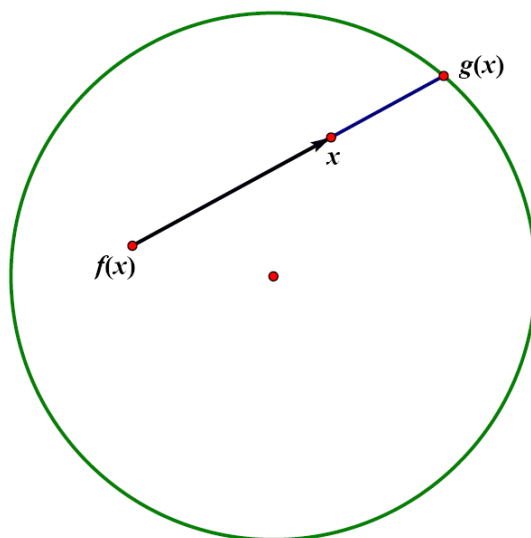
这样得到的极限曲线 $\lim P_n = P$ 具有这样的性质，它的周长是无穷大，但所包围的图形面积是有限的，为三角形 P_0 面积的 $8/5$ 倍。特别地，雪花曲线的边界太曲折了，以至于没有切线。雪花曲线也称科赫曲线，是分形几何的重要例

子。

上面反例的证明也许不难，但对于这种天书一般的想象力，很多人自然想一探究竟，这么巧妙的 idea 是如何想出来的？笔者曾试图归纳总结一下想象力或者创造力有没有什么规律，结果却是徒劳。大概试图用一种确定性方法来研究高度不确定的对象，可能出发点就不太对。这也许就是创造力本身被称赞为艺术的原因之一吧，数学中的构造性方法是创造力的典型代表。

数学中的构造性方法有几种体现：

1. 有的是否定某个想法（如上面的处处连续处处不可导函数）；
2. 有的是验证某个结论是否成立，比如我们研究某一类函数或泛函的下确界问题，猜出下界 C ，然后在构造一个函数序列，使得下界 C 能够被逼近，这说明 C 就是所求的下确界。
3. 有的是验证某些定理成立的前提条件，如果条件改一改定理结论是否成立？如著名的布劳威尔不动点定理断言从单位闭球到单位闭球的连续映射 $f: D^n \rightarrow D^n$ 必有不动点 $f(x_0) = x_0$ 。这是一个拓扑学定理，如果把单位闭球 D^n 换成一般的单位球面，是否存在不动点？容易证明，不动点不一定存在，如设映射 f 为对径映射 $f(x) = -x$ ，这也是一种构造法。实际上不动点证明的关键是利用了单位闭球 D^n “凸性”，下面这张图在不动点的各种证明（如本书附录）中经常遇见，这个构造的新映射 $g(x)$ 太重要了！



上面这几个例子还是局限在数学分析本身，其实构造法还可以走的更远一些，如分析和几何跨界。这里强调：本文所称“构造法”，不仅包括严格的数学

证明，也包括一些带有想象力的独特想法。下面用两个问题来说明一下。

第一个问题是北大 2009 年研究生考试的一道数学分析试题，问题如下：

已知 $x, y, z \geq 0$ ， $x + y + z = \pi$ ，试求 $2\cos x + 3\cos y + 4\cos z$ 的最大值和最小值。

本题的解法并不难，难的是如何求解超越方程。先消去 z 转化为无条件极值问题，则原问题等价于求下列二元函数的最值：

$$2\cos x + 3\cos y + 4\cos z = 2\cos x + 3\cos y - 4\cos(x+y) := f(x, y)$$

其中 $f(x, y)$ 定义在有界闭区域 $D = \{(x, y) | 0 \leq x, y \leq \pi, 0 \leq x + y \leq \pi\}$ 上。由区域 D 的紧性以及 $f(x, y)$ 的连续性可知，最大值和最小值存在。

求解过程很流程化，先求边界上最值，再求解内部的驻点值，然后放在一起比较大小即可。容易求出边界 ∂D 上最大值为 5，最小值为 1；求解内部的驻点值似乎要求解三角方程：

$$\begin{cases} f_x(x, y) = 4\sin(x+y) - 2\sin x = 0 \\ f_y(x, y) = 4\sin(x+y) - 3\sin y = 0 \end{cases}$$

客观说，笔者初解时被这个问题卡住许久，不大相信北大出这道题是要考生求解方程。经过思考之后，笔者发现，问题是要求函数的驻点处的函数值，而不是驻点值，也就是说只要证明驻点存在即可（不用太机械）。

一方面，偏导数方程给出了三角函数的“比例关系”。另一方面可观察到若驻点值存在，则 x, y 都是正数且满足 $x + y + z = \pi$ 。多少有些灵感划过心海吧，我意识到如果把 x, y, z 视作三边之长分别为 a, b, c 的三角形的内角，则由“正弦定理”（来自几何），得

$$\frac{c}{\sin(x+y)} = \frac{b}{\sin y} = \frac{a}{\sin x}$$

结合偏导数方程，得到关系式 $a = 2c, 3b = 4c$ 。令 $a = 6t, b = 4t, c = 3t$ ，则不难验证确实可以构成一个三角形，且满足此关系的三个内角是惟一确定的，这就证明了驻点的存在性。

余下就容易了，利用余弦定理可以算出

$$\cos x = -\frac{11}{24}, \cos y = \frac{29}{36}, \cos z = \frac{43}{48}$$

于是驻点对应函数值为

$$-2 \times \frac{11}{24} + 3 \times \frac{29}{36} + 4 \times \frac{43}{48} = \frac{61}{12} > 5$$

于是原问题所求最大值为 $\frac{61}{12}$ ，最小值为 1。出乎意料，这个最大值还真不好猜。

本题的巧妙之处在于以“正弦定理”为桥梁，将一个三角方程问题转化为一个三角形的边角关系问题，算得上是“数形结合”的例子吧。

另一个问题是一道智力题，一般化的表述为：

设在区间 $(0, 1)$ 内随机选取 n 个数，求这 n 个数中最小的数的期望。

这个问题不算难，有标准的解法。每个数在 $(0, 1)$ 都服从均匀分布，即这 n 个数是独立同分布的。记第 i 个数对应的随机变量是 X_i ，则原问题转化为求最小数的概率分布或概率密度

$$P\{\min(X_1, \dots, X_n) \leq x\} = ?$$

然后在按照数学期望公式直接计算即可。

首先，由独立同分布可知

$$\begin{aligned} P\{\min(X_1, \dots, X_n) \leq x\} &= 1 - P\{\min(X_1, \dots, X_n) > x\} \\ &= 1 - P\{X_1 > x, \dots, X_n > x\} = 1 - P\{X_1 > x\} \cdots P\{X_n > x\} \end{aligned}$$

其次由独立随机变量 X_i 服从 $(0, 1)$ 上的均匀分布

$$P\{X_i > x\} = \int_x^1 du = 1 - x$$

由上面可求出这 n 个的联合分布函数

$$F_{\min}(x) = P\{\min(X_1, \dots, X_n) \leq x\} = 1 - (1 - x)^n, \quad x \in (0, 1)$$

于是其对应的概率密度函数为

$$f_{\min}(x) = F'_{\min}(x) = n(1 - x)^{n-1}$$

所求最小数 $X = \min(X_1, \dots, X_n)$ 的数学期望为

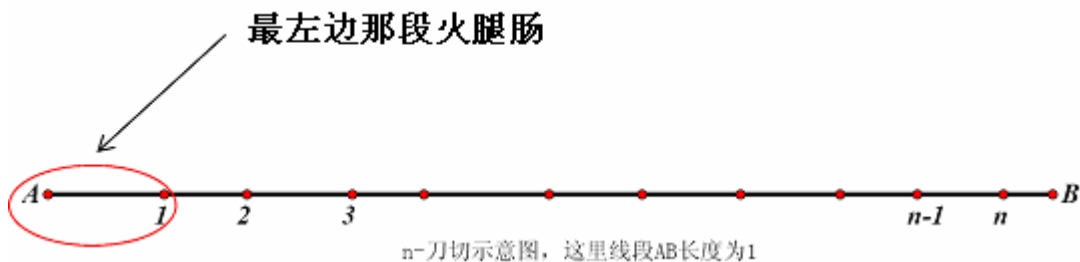
$$EX = \int_0^1 nx(1 - x)^{n-1} dx = nB(2, n) = \frac{1}{n+1}$$

计算中用到了 Γ 函数与 B 函数的一些基本性质。

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \quad \text{以及} \quad \Gamma(n+1) = n!$$

这是一个标准且严格的数学证明，也是笔者最初想到的方法。但总觉得不太自然，因为当时很快猜出了结果 $1/(n+1)$ ，总觉得有更直观的思路，感觉就在眼前，但却捕捉不到。晚上在超市里买火腿肠时来了灵感，笔者发现这个问题可以转化为更直观亲切的表述：

考虑一根长度为 1 的火腿肠，随机切 n 刀，分成 $n+1$ 段，问最左边的那一段长度期望是多少？



将这 $n+1$ 段火腿肠调整排列次序，仍然组成一根长度为 1 的火腿肠，构成另一个 n 刀切（某种平等性），可推测出这 $n+1$ 段火腿肠的长度期望均相等，都是 $1/(n+1)$ ，这个数就是我们所求的最小数的期望。直观上确实不难想到，最左边那段火腿肠没有什么特殊性。

上面两个问题第一个借助于几何来构造求解方法，第二个借助于物理来构造求解方法。将一个具体问题抽象化能有助于帮助我们接触到数学本质。反过来，将一个抽象问题具体化、直观化，有助于找到方向，毕竟构造法是一个具体过程。

笔者复习《微积分》时，出于学习需要，发现了微分中值定理的辅助函数构造法中最常见的“ $e^{\int P}$ 因子法”，该方法体现了中值定理与微分方程之间的联系，在解题中非常实用，这种方法很多人都这么用。

中值定理与微分方程关系紧密，比如 1 阶形式 $f'(\xi) + P(\xi)f(\xi) = 0$ 。将 ξ 视作变量 x ，得到一个一阶微分方程 $f'(x) + P(x)f(x) = 0$ 。分离变量，再积分得到

$$f(x)e^{\int P(x)} = 1$$

上式两边对变量 x 求导，再注意到关键因子 $e^{\int P}$ 是正的，可以消去，得到

$$f'(x) + P(x)f(x) = 0$$

就笔者所见，大多数 $f'(\xi) + P(\xi)f(\xi) = 0$ 类中值定理问题的辅助函数如下

$$F(x) = e^{\int P(x)} f(x)$$

然后验证 $F(x)$ 有“两个零点”，应用罗尔中值定理，得 $f'(\xi) + P(\xi)f(\xi) = 0$ 。

二阶中值定理不常见，如果是常系数的情况 $f''(\xi) + pf'(\xi) + qf(\xi) = 0$ ，且方程系数满足 $p^2 - 4q \geq 0$ 。此时可效仿一阶情况。这类二阶问题一般需要函数

$f(x)$ 有三个零点, 然后求导两次, 得到表达式。

辅助函数的构造依然按照解微分方程思路, 根据《新讲》第一册介绍的算子法, 引入微分算子 $D = d/dx$, 设方程 $t^2 + pt + q = 0$ 的两根为 λ_1, λ_2 , 则

$$(D - \lambda_1)(D - \lambda_2)f(x) = 0$$

解上面的微分方程, 仍然可以按照一阶方程的因子法。

先可构造辅助函数

$$F(x) = e^{-\lambda_2 x} f(x)$$

如果 $f(x)$ 有 3 个零点, 对上面公式求导可知 $g(x) = (D - \lambda_2)f(x)$ 有 2 个零点。

类似地, 再构造辅助函数 $\varphi(x) = e^{-\lambda_1 x} g(x)$, 求导可知

$$f''(\xi) + pf'(\xi) + qf(\xi) = 0$$

不难看出, 上面的辅助函数构造法本质是反解微分方程, 有点逆向思维的味道了。最后强调, 中值定理的试题比较活, 需要活学活用才行, 辅助函数构造方法很多 (不一定是解微分方程), 常需要分析法和归纳法相结合。

上面三个例子是笔者亲身经历的, 印象深刻, 同时难度也较低, 所以详细给出了具体推导过程。应该说, 数学中的构造方法变化多端, 少有定法, 本文涉及的几个微积分初等例子也不可能穷其所有。如果对提高数学创造力给出一个可行的具体办法 (带有个人看法), 除去天赋之外, 笔者给出的建议是: 熟能生巧, 多想多干。

下面的两个例子都是经典的构造技巧, 读者要体会一下。首先谈谈代数基本定理: **非常数多项式**

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0, \quad a_n \neq 0$$

有且仅有 n 个复零点。

代数基本定理最初由达朗贝尔和高斯所证明, 到底是谁首先证明出来的还有些争议, 因为以现代数学标准看, 两个人的方法都不是特别严密。不过多数人倾向于高斯先证明的, 而且高斯本人也把代数基本定理的证明作为他的博士论文。该定理的证明目前有几十种, 方法各异, 本书收录的这个证明构思巧妙, 一般课本里是找不到的。证明过程中只用到了复变函数中的柯西定理: $f(z)$ 是复平面 C 上的解析函数, D 为有界区域, 其边界 ∂D 是简单闭曲线, 则有

$$\int_{\partial D} f(z) dz = 0$$

利用反证法, 假设 $p(z)$ 无零点. 不妨设 $x \in \mathbf{R}$ 时, $p(x) \in \mathbf{R}$. 否则用 $p(z)\bar{p}(z)$ 代替 $p(z)$ 即可, 其中

$$\bar{p}(z) = \bar{a}_n z^n + \bar{a}_{n-1} z^{n-1} + \cdots + \bar{a}_1 z + \bar{a}_0, a_i \in \mathbf{C}$$

于是 $x \in \mathbf{R}$ 时, 实多项式 $p(x)$ 也无零点, 从而 $p(x)$ 在实数轴上不变号. 不妨设 $p(x) > 0$, 其中 $x \in \mathbf{R}$. 故积分

$$I = \int_0^{2\pi} \frac{d\theta}{p(2\cos\theta)} > 0$$

做变换 $z = \cos\theta + i\sin\theta$, 则 $2\cos\theta = z + z^{-1}$ 及 $dz = izd\theta$, 于是上述积分等价于下述单位圆周上的积分:

$$I = \frac{1}{i} \int_{|z|=1} \frac{dz}{zp(z+z^{-1})} = \frac{1}{i} \int_{|z|=1} \frac{z^{n-1} dz}{z^n p(z+z^{-1})}$$

记 $Q(z) = z^n p(z+z^{-1})$, 我们证明 $Q(z)$ 在复平面 \mathbf{C} 上恒不为零.

当 $z \neq 0$ 时, 根据条件 $p(x) > 0, x \in \mathbf{R}$ 可知 $Q(z) \neq 0$; 当 $z = 0$ 时, 利用二项式定理可知

$$Q(z) = z^n p(z+z^{-1}) = a_n + z(\cdots)$$

于是 $Q(0) = a_n \neq 0$. 综上所述, $Q(z)$ 在复平面 \mathbf{C} 上恒不为零.

令 $f(z) = z^{n-1} / Q(z)$, 则 $f(z)$ 是复平面 \mathbf{C} 上的解析函数, 根据柯西定理, 得

$$I = \int_{|z|=1} f(z) dz = 0$$

与前面的结论矛盾! 因此, 多项式 $p(z)$ 在复平面 \mathbf{C} 上至少有一个复零点.

记这个零点为 z_0 , 利用多项式除法得到 $p(z) = (z - z_0)h(z)$, 其中 $h(z)$ 是 $n-1$ 次多项式. 反复应用前面的论证结论, 可知多项式 $p(z)$ 有且仅有 n 个复零点.

复变函数教材中收录的代数基本定理证明方法通常是刘维尔定理和幅角原理, 不过这两个定理本身用到了较深的定理. 复变函数中的柯西定理相当于微积分中的牛顿-莱布尼茨公式, 算是最基本的分析定理了.

最后介绍下椭圆函数和自守函数的构造. 因为这属于相对高深的内容, 所以只能粗略地介绍一下关键想法了, 就当是一次高等数学科普吧.

现行教材上直接定义具有不共线的复周期 ω_1, ω_2 的亚纯函数

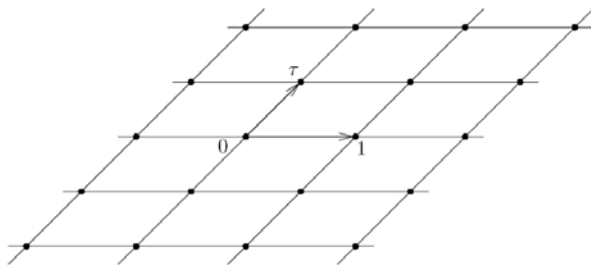
$$f(z + \omega_1) = f(z + \omega_2) = f(z)$$

为椭圆函数。不过历史上，高斯和阿贝尔提出了研究椭圆积分的反函数（反演），后称椭圆函数，我们在前面类比那一章已经介绍了这方面的内容。那么能不能不利用积分反演，直接构造双周期的椭圆函数。

直观上不难想象，椭圆函数的所有周期是复平面上一些格子（周期格）

$$\Lambda = \{m\omega_1 + n\omega_2 \mid m, n \in \mathbb{Z}\}$$

其中复数 ω_1 和 ω_2 是实线性无关的。



那么形式上，如下函数满足双周期性：

$$\sum_{\omega \in \Lambda} \varphi(z - \omega) \text{ 或 } \prod_{\omega \in \Lambda} \varphi(z - \omega)$$

这里函数 φ 待定。

之所以强调是形式上，是因为要使得上面的函数有意义，还要考虑收敛性问题。椭圆函数中最有名的要数魏尔斯特拉斯椭圆函数：

$$\wp(z) = \frac{1}{z^2} + \sum_{0 \neq \omega \in \Lambda} \left[\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right]$$

据说，这种形式的构造类比为余切级数展开（欧拉等式）

$$\pi \cot \pi z = \frac{1}{z} + \sum_{n=1}^{\infty} \left(\frac{1}{z+n} + \frac{1}{z-n} \right)$$

利用复积分可以证明非常数椭圆函数在一个周期四边形内必有奇点，且阶数至少是二（证明见《特殊函数概论》）。椭圆函数 $\wp(z)$ 的构造选择了一个二阶极点的情形（不妨设 $z=0$ 为二阶极点）。于是根据椭圆函数周期性要求，可知 $z=\omega$ 为二重极点。再由上面形式级数的思想，考察下列函数就自然了：

$$\sum_{0 \in \Lambda} \frac{1}{(z - \omega)^2}$$

但是这个级数不收敛，所以为满足收敛性还得再减去一项 $1/\omega^2$ 。当 $\omega=0$ 这种情况下， $1/\omega^2$ 发散，所以应该剔除，同时把 $1/z^2$ 单独写出来。于是级数

$$\wp(z) = \frac{1}{z^2} + \sum_{0 \neq \omega \in \Lambda} \left[\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right]$$

是收敛的, 具有二阶极点, 且满足 $\wp(z + \omega_1) = \wp(z + \omega_2) = \wp(z)$.

$\wp(z)$ 的导数 $\wp'(z)$ 也是椭圆函数, 它们之间满足上面的微分方程

$$\wp'(z)^2 = 4\wp(z)^3 + g_2\wp(z) + g_3$$

$\wp(z)$ 和 $\wp'(z)$ 也所以非常重要, 是因为它们不仅形式简单, 还可以作为“基函数”, 这种基函数分解思想 (当然不是线性的) 可以由下述定理描述:

设 $\wp(z)$ 的周期为 ω_1, ω_2 , 则所有以 ω_1, ω_2 为双周期的椭圆函数

$$f(z + \omega_1) = f(z + \omega_2) = f(z)$$

都可以表示成下列形式

$$f(z) = H(\wp(z)) + \wp'(z)Q(\wp(z))$$

其中, $H(z), Q(z)$ 为有理函数, 与 $f(z)$ 有关。

自守函数可以看作椭圆函数的推广, 创立者是庞加莱和克莱因, 下面主要介绍庞加莱的想法。

庞加莱最初研究对象是所谓富克斯方程

$$\frac{d^2 w}{dz^2} + p(z) \frac{dw}{dz} + q(z)w = 0$$

其中 $p(z)$ 和 $q(z)$ 为有理函数。假设方程有两个线性无关的解 $w_1(z)$ 和 $w_2(z)$, 富克斯的建议是研究它们的商

$$\varsigma(z) = \frac{w_1(z)}{w_2(z)}$$

当 z 沿着奇点邻域内一条曲线 Γ 环绕一圈时, $w_1(z)$ 和 $w_2(z)$ 变成了

$$w_1^*(z) = \alpha w_1(z) + \beta w_2(z)$$

$$w_2^*(z) = \gamma w_1(z) + \delta w_2(z)$$

这里的系数与曲线 Γ 有关。

在补充了其它条件后, 庞加莱得到了一个关系式

$$\varsigma(z) = \frac{\alpha \varsigma(z) + \beta}{\gamma \varsigma(z) + \delta}$$

其中 $\alpha, \beta, \gamma, \delta$ 组成了一个离散变换群 ($\alpha\delta - \beta\gamma \neq 0$)。

进一步, 庞加莱研究 $\varsigma(z)$ 的反函数 $z = z(\varsigma)$, 发现它应该满足

$$z(\varsigma) = z \left(\frac{\alpha\varsigma + \beta}{\gamma\varsigma + \delta} \right)$$

也就是说庞加莱发现应该有一种亚纯函数（即自守函数），满足：

$$F(z) = F \left(\frac{\alpha_i z + \beta_i}{\gamma_i z + \delta_i} \right)$$

当系数 $\alpha_i, \beta_i, \gamma_i, \delta_i$ 为整数，且 $\alpha_i \delta_i - \beta_i \gamma_i = 1$ 时， $F(z)$ 称为模函数。这种函数在数论中非常有用，费马大定理的证明就与模函数有关，这方面的简单介绍，读者可参考逆向思维那一章。因为历史原因，庞加莱称自守函数为富克斯函数，本书称为自守函数。

起初，庞加莱对自守函数冥思苦想了整整两个星期，企图证明它不存在，但这个想法以后被证明是错误的。后来，“一天晚上”，庞加莱说：“与往常不同，我喝了浓咖啡，因而辗转反侧，难以入眠，众多的思绪蜂拥而来，我感到它们在不断地冲突和碰撞……直到最后，它们一一相联，也就是说，形成了一个稳定的组合体。”对此，庞加莱还作了心理学的剖析，他说：“在这种情况下，我们似乎处于自身的无意识工作状态，虽然也部分地感到有某些超兴奋的有意识的思维成分，但总的来说，并不能改变无意识的特征。于是，我们就只能含糊糊糊地领略到两种思维机制的区别。”

下一步的工作就是企图找到函数的表达式。庞加莱回忆道：“我想要把这类函数表示成两个级数之商，这个思想是非常自然和有明确目标的，这时我想起了类似于此的椭圆函数的情形。如椭圆函数

$$\wp(z) = \frac{1}{z^2} + \sum_{0 \neq \omega \in \Lambda} \left[\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right]$$

我就设想，如果这两个级数存在，它们会有什么样的性质呢？循此向前，并没有遇到什么困难，我构造出了这两个级数，并称之为 \wp -级数”。“就在此时，我离开了我所居住的地方卡昂，在矿业学院的资助下，开始了地质考察的旅行生活。旅途中的许多事使我忘掉了数学工作。到了康斯坦茨湖，我们乘一辆马车到其他地方去，就在我把脚放到马车踏板上的那一刹那，一个思想突然闪现在我脑海中，而在此之前，我还从来没有想到过。这个思想就是我用以定义自守函数的变换与非欧几何的变换是等价的。

庞加莱所说的非欧几何变换指的是双曲几何中的默比乌斯变换：

$$M(z) = \frac{\alpha z + \beta}{\gamma z + \delta}$$

庞加莱上半平面 $\mathbf{H} = \{z \mid z = x + iy, y > 0\}$ 是双曲几何的经典模型。

当 $\alpha\delta - \beta\gamma > 0$ 时, 这些默比乌斯变换组成一个变换群, 称为双曲平面 \mathbf{H} 的“运动群”, 使得 \mathbf{H} 上的直线变成另一条直线, 默比乌斯变换同时也是保角和等距变换。按照克莱因的埃尔兰根纲领思想, 双曲几何的性质就是那些在双曲运动群下保持不变的性质。

为了表述简化起见, 我们用下面这个特殊离散的变换群

$$G = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbf{Z}, ad - bc = 1 \right\}$$

来描述庞加莱构造自守函数的想法 (专业一点的说法是 $G = \mathrm{SL}_2(\mathbf{Z})$)。庞加莱首先构造了自守形式, 效仿构造魏尔斯特拉斯椭圆函数 $\wp(z)$, 构造如下函数

$$\sum_{M \in G} H(M(z))$$

其中 $H(u)$ 为有理函数, 且在上半平面 \mathbf{H} 的边界上没有极点。

但上面的级数不一定收敛, 庞加莱的想法是加上一个收敛因子 $(cz + d)^{-2m}$, 最终构造出了现在称为 ϑ -级数的自守形式

$$\vartheta(z) = \sum_{i=1}^{\infty} \frac{1}{(c_i z + d_i)^{2m}} H\left(\frac{a_i z + b_i}{c_i z + d_i}\right), \quad m > 1$$

注意到若 $M_i, M_j \in G$, 则 $M_i M_j \in G$. 庞加莱进一步证明了

$$\vartheta\left(\frac{\alpha z + \beta}{\gamma z + \delta}\right) = (\gamma z + \delta)^{2m} \vartheta(z)$$

称为权 $2m$ 自守形式。

有了自守形式, 构造自守函数就容易了, 庞加莱考虑两个权 $2m$ 自守形式之比, 定义

$$F(z) = \frac{\vartheta_1(z)}{\vartheta_2(z)}$$

这个自守函数满足

$$F\left(\frac{az + b}{cz + d}\right) = F(z), \quad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$$

上述过程可以类比用 $\sin z$ 和 $\cos z$ 之比定义 $\tan z$ 的过程。

庞加莱利用 ϑ -级数构造出了自守函数实际上也证明了自守函数存在定理。如果取 $a = b = d = 1, c = 0$, 则 $F(z+1) = F(z)$, 这说明自守函数是周期函数的推广, 但有着更高层次的不变对称性 (在群 G 意义下)。

一个有趣的现象是: 分析学家偏爱椭圆函数, 而数论学家偏爱自守函数 (自守形式)。自守形式当下是数学界的大热门, 大名鼎鼎的 “Langlands 纲领” 就是与自守形式有关的。当然, 所用的数学工具也远远超出了 ϑ -级数的范围, 如大量运用表示论工具, 感兴趣的读者可以用 Google 搜索 “Langlands Program” 关键词。

评注: 上面举出了若干构造例子, 由浅入深, 甚至到了自守函数。意在表明数学思想是相通的。方法上难度有高低, 但思想是一致的。像级数构造、收敛因子等都是经典的数学方法。因为庞加莱是自己的偶像, 所以用的笔墨多一些, 也趁机借此机会展现一位数学大师的创造性思考过程。很多数学大师 (以高斯为代表) 对自己在数学上的创造过程总是惜墨如金, 不肯多说, 这不能不说是一种遗憾。

第12章 多读一些著名定理

笔者习惯将本科以及本科以下阶段称为少年阶段，这个阶段的学习和梦想其实是最大胆的，谁不曾梦想将来去证明一些大定理，做出一些伟大的发现。不过随着学习的深入，大多数人风格上逐渐趋于谨慎和保守，最终守在一小块土地上精耕细作。这是一种务实的态度，某种程度上也有些无奈。所以，趁着同学少年，意气风发时，多读一些著名定理就十分必要了。搞数学有时候也需要一种疯狂的执着精神，就像黑客一样。

从技术角度看，相比自己给出证明，读懂别人的现成证明要相对容易一些。当然，多数著名定理的证明需要引入一些高深的数学工具，特别是现代数学中那些大定理，即便是准确表述也是很困难的（涉及概念也较多）。不过，有些著名定理（如素数定理和不动点定理等）经过后人的不断努力挖掘，避免了引进比较高深的数学工具，技术门槛大大降低，已经可以写进本科教材了。注意，这里并非强调一定要用高中的数学方法来证明的著名定理，那不现实。

笔者为什么推崇《数学分析新讲》一书，除了教材本身写的平易近人之外，还有一个吸引笔者的地方，那就是《新讲》收录进很多著名定理，如：

1. 万有引力定理与开普勒三定律；
2. 代数基本定理；
3. 布劳威尔不动点定理；
4. 斯通-魏尔斯特拉斯定理；
5. 处处连续处处不可导函数；
6. 填满正方形的连续曲线；
7. 等周问题。

等等。按《新讲》作者张老师的说法：“花费这么多时间和精力学习微积分，学完之后究竟能做什么事情？好像除了求极值、求面积体积之外，就什么也不会了。笔者写这套书，就是想或多或少地改变一下这种状况。”

自己当年对张老师的这些话有点感觉，但不深刻。随着水平的不断提高，对这老师的这句话感悟越来越深。笔者高一时就已经知道了牛顿-莱布尼茨公式

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

当时也能熟练计算一些涉及多项式和三角函数的简单积分（太复杂的形式就算不出来了）。特别以自己能计算一些复杂曲边梯形面积和旋转体体积（别的同学

不会)而自傲,顺便也能解决一下简单的变力做功问题(笔者非常喜欢物理)。一直到大一,笔者一直以为微积分只能干这些活儿(求面积和体积之类的)。

转机发生在大二,当看到了两本书《数学分析新讲》和《数理方程与特殊函数》,这两本书涉及了许多较大定理(证明相对复杂),证明通常很长,涉及很多数学工具的综合运用。正是这两本书使笔者意识到原来微积分的威力不仅仅在于求极值、面积和体积啊,之前的见识真是井底之蛙!这是自高一自学微积分以来,认识上的第二次飞跃。

上面提到的牛莱公式是可以推广到 Lebesgue 积分上的,若 $f(x)$ 是 L 可积的,且除去有限个例外点之外有 $F'(x) = f(x)$, 则牛莱公式仍然成立。特别地,学过实变函数的读者还知道,存在一个“康托(三分集)函数”,除去一个不可数的零测集之外有 $F'(x) = f(x)$, 但牛莱公式不成立,这是一个经典反例,说明牛莱公式不能随意推广到零测集。

但是,恐怕很多人不了解下面这个最强牛莱定理:

设 $F(x) \in C[a, b]$ 是连续函数, $f(x) \in L[a, b]$ 是 Lebesgue 可积的,若至多除去一个可数集之外,有 $F'(x) = f(x)$, 则

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

如果说 Riemann 积分中的牛莱公式相对平淡无味的话,那么上面的这个定理就是笔者所言的著名定理了。实际上,“至多除去一个可数集之外,有 $F'(x) = f(x)$ ”这个条件满足时,此时 $F(x)$ 是绝对连续的。细心读者也能发现,上面的定理可以反推出康托函数反例中的零测集是不可数集。这个定理也暗示着我们,同样是零测集,但可数集和不可数零测集之间是有深刻差别的。

多看著名定理,你一般有这样的感觉:这些大定理的证明往往涉及到很多数学工具的应用,其中更包含许多创造性的想法和独具匠心的组合,这一点在一般教材上是学不到的。因为教材通常出于逻辑考虑,基本上是一章一个内容,例子和习题大多是针对该章节内容的,跨章节的综合题很少。稍不注意就变成了只见树木,不见森林了。比如很多人感到中值定理题中的辅助函数很难想,其实如果联想的常微分方程中的一些方法,很多辅助函数的构造还是很自然的。

有些人将多读著名定理片面的理解为多看教材。首先,多看教材,特别是名著,这是好事。其次,必须深刻意识到这样一个现实,教材中多数内容和例

子非常陈旧，特别是当下国内基础教材泛滥的情况下。真的能做到像《新讲》作者张老师那样精挑细选著名定理，然后加入教材的作者非常少。很多时候，这不是作者能力的问题。所以在掌握教材基本内容之后，多看一些高水平的杂志是一个好的选择，笔者比较推荐的是《美国数学月刊》和《数学译林》，这两本杂志通常会精选一些著名定理，难度也适中。当然，你能看明白更高水平的数学杂志那就更好了。

数学中泛泛而谈总是收获不多的，特别是本科阶段，最好不要养成这种习惯。读一本好书或者一个著名定理的证明，笔者习惯写一些评注，内容有的是一些自己的想法或解读，有的是补充一些技术细节，使得原文章从数学角度看，更自然一些。这方面的代表例子，请看本书附录中《素数定理的简单解析证明》长文，笔者喜爱这样的风格。

本章的内容是涉及著名数学定理的，因此得选择一个有难度，同时又很著名的定理，但所用方法还是数学分析的基本方法。最初打算用素数定理的解析证明作为代表，后来觉得不妥，证明毕竟要用到 $\zeta(s)$ 函数以及复变函数的一些知识。想来想去，最后选用的例子是数学大师柯尔莫哥洛夫的少年成名之作：

$L[0, 2\pi]$ 中几乎处处无界发散的 Fourier 级数反例。

证明选自汪林等编著的《实分析中的反例》一书，陈建功的《三角级数论》下册也收录进这个反例。反例的具体构造比较复杂（涉及许多不等式硬估计），但算不上高深，放到本章非常合适。

数学分析中有这样一个定理：

连续周期函数 $f(x)$ 的 Fourier 级数几乎处处收敛到 $f(x)$ 自身。

这个定理看似其貌不扬，难度却是菲尔兹奖级别的。上述结论对 $L^p[0, 2\pi]$ ($p > 1$) 空间中的函数也是成立的。但在 $L[0, 2\pi]$ 中未必成立，经典反例就是柯尔莫哥洛夫的这个例子。

网上流传着一个故事，据说是关于米尔诺或丹齐格的，大意是：

一位本科生上课迟到了，看到黑板上有一道题，以为是课后作业，就抄下来课下做。结果花费了好些天才做出来，并向老师抱怨作业题太难了。没想到老师激动地告诉他，这不是课后作业，而是一个未解的著名猜想。该学生因上课迟到没听见前半部分，误以为是作业，结果阴差阳错地做出来了。

故事的真实性恐怕无从考证了，这类故事是说，本科学生也可以做出世界

级的工作。当然，例子非常稀少。古人也有这方面的代表人物，如伽罗瓦。笔者阅读柯尔莫哥洛夫英文文集第一卷后，获得的信息是他确实是在大三的时候构造出本文的著名反例，并因此扬名数学界。比起米尔诺那个飘渺的故事，柯尔莫哥洛夫的例子确实有据可查，文章如下：（选自《柯氏文集》第 1 卷）：

1. A FOURIER-LEBESGUE SERIES DIVERGENT ALMOST EVERYWHERE*

The purpose of this paper is to give an *example of a summable¹ function with an almost everywhere (i.e. everywhere except on a set of measure zero) divergent Fourier series*).

必须指出，柯尔莫哥洛夫构造反例的想法不是偶然的，他的老师鲁金是函数论方面的大牛，曾提出一个猜想：存在连续函数在某个正测度集上发散。当然，卡尔森的工作说明这个猜想是不正确的。柯尔莫哥洛夫似乎在大二时参加了莫大数学系的一个三角级数讨论班，他本科的很多文章都是关于三角级数的。所以本文的反例不能完全说是天外飞仙，灵光一现的天才灵感，是有很多前期的工作作为铺垫的。任何一篇经典论文，或者一门绝技，都是经过辛苦努力换来的。

下面我们开始具体的数学推导，比较有技术含量，阅读要有耐心。按笔者的看法，读者应该紧紧抓住“无界矩形脉冲函数序列”（阶梯函数）这一主线，看看是如何从中巧妙地“构造”出调和级数的，柯尔莫哥洛夫这篇文章把阶梯函数的妙用发挥到极致了！

先从一个广义积分说起。通过适当的放缩，可以证明下面的广义积分是收敛的

$$\int_0^{\infty} f(x) dx = \int_0^{\infty} \frac{x dx}{1+x^6 \sin^2 x}$$

注意到 $x > 0$ 时，被积函数 $f(x) > 0$ ，但是 $f(n\pi) = n\pi$ ，这说明 $\overline{\lim_{x \rightarrow \infty}} f(x) = +\infty$ 成立。进一步研究函数图形发现 $f(x)$ 在 $x = n\pi$ 临近，函数图形急剧上升，其它地方急剧下降，这很像一个脉冲函数。更直观的例子是三角脉冲连续函数，图形是一些三角形，这些三角脉冲越来越窄，但峰值却越来越高，而每个三角形的面积却是有限的。

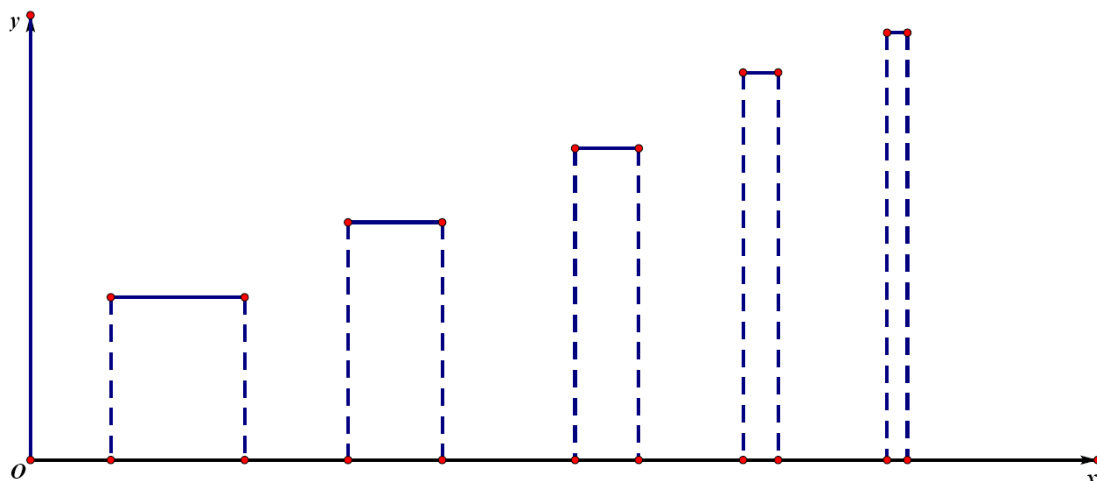
从直观化的角度看，面积可以看成高度与宽度两个变量的乘积，如果一个变量急剧增加，另一个变量急剧减少，那么乘积仍然可能是有限的。如果抛开连续性条件，可以构造一个无界矩形脉冲函数序列（阶梯函数）

$$f(x) = \frac{m_n}{n^2}, x \in \left[n - \frac{1}{m_n}, n + \frac{1}{m_n} \right], m_n = 2^n; f(x) = 0, \text{ 其它}$$

这个矩形脉冲函数非负，满足极限上 $\overline{\lim}_{x \rightarrow \infty} f(x) = +\infty$ 且广义积分

$$\int_0^{+\infty} f(x) dx = \sum \frac{m_n}{n^2} \cdot \frac{2}{m_n} = 2 \sum \frac{1}{n^2} < +\infty$$

笔者现在有个习惯，就是经常画一些示意图来帮助自己理解思考的问题，或者作为具体例子。无界矩形脉冲函数序列的图形大致如下：



无界矩形脉冲函数序列

构造反例函数的直观想法是：不能选择性态好的可积函数，要找一些很 **bt** 的可积函数，而“无界可积”函数就是这种 **bt** 函数。当然，不过为了降低构造难度，可以先构造一个可积却无界的函数序列。

设 $f \in L[0, 2\pi]$ 是绝对可积的，其对应的 **Fourier** 级数为

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

其中系数 a_n, b_n 由以下积分确定

$$\begin{aligned} a_0 &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) dt \\ a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt dt \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt dt \end{aligned}$$

上面级数之所以没有用“等号=”，是因为右边的级数未必收敛，即使收敛也未必收敛于 $f(x)$ 。关于 **Fourier** 级数的收敛性判别法有很多种，多数还互不相容，

详见大学数学分析教材，如《数学分析新讲》。

习惯上记部分和

$$S_n(f, x) = \frac{a_0}{2} + \sum_{k=1}^n a_k \cos kx + b_k \sin kx$$

将系数的积分表达式代入 $S_n(f, x)$ ，得到

$$\begin{aligned} S_n(f, x) &= \frac{a_0}{2} + \sum_{k=1}^n a_k \cos kx + b_k \sin kx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) dt + \frac{1}{\pi} \sum_{k=1}^n \left(\int_{-\pi}^{\pi} f(t) \cos kt \cos kx dt + \int_{-\pi}^{\pi} f(t) \sin kt \sin kx dt \right) \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{k=1}^n (\cos kt \cos kx + \sin kt \sin kx) \right] dt \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{k=1}^n \cos k(t-x) \right] dt \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \frac{\sin(n + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt \end{aligned}$$

下面的积分一般称作狄利克雷积分

$$S_m(f, x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \frac{\sin(m + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt$$

于是研究 Fourier 级数的收敛性可以转化为研究 $m \rightarrow \infty$ 时，上面积分的收敛性，这种研究收敛性的思路被当下数学分析教材所采用。上面这个积分在 $t = x$ 点可能有奇异性，造成一些麻烦。柯尔莫哥洛夫先考察一个简单情况，比如函数 $f(x)$ 在一个小闭区间 J 上为常数，其它地方为 0，并取点 x 为 J 之外的点，这样避免了对奇异性进行讨论。于是得到

$$S_m(f, x) = \frac{1}{\pi} \int_J f(t) \frac{\sin(m + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt$$

因为 $f(x)$ 为常数，所以由 Riemann-Lebesgue 引理

$$\lim_{\lambda \rightarrow \infty} \int_a^b f(x) \sin \lambda x dx = \lim_{\lambda \rightarrow \infty} \int_a^b f(x) \cos \lambda x dx = 0$$

可知 $m \rightarrow \infty$ 时，上述积分趋于 0，不过这是很平庸的结果。

柯尔莫哥洛夫考察一个动态过程，用一个函数序列 $\varphi_n(t)$ 代替 $f(t)$ ，而用闭区间序列 J_i 代替闭区间 J ，特别地，这里 $\varphi_n(t)$ 可能与 m 有关，那么就不能简单地套用 Riemann-Lebesgue 引理了，需要更精细地分析。他最后选择的函数序列

$\varphi_n(t)$ 就是无界矩形脉冲函数序列, 暂且要求

$$\int_{J_i} \varphi_n(t) dt = \frac{2}{n}$$

其它条件后面会一一说明。

在无数次试探性计算 (期间有很多无用功) 之后, 利用含绝对值不等式

$$|a| = |b + a - b| \geq |b| - |a - b|$$

柯尔莫哥洛夫发现了一个十分重要的不等式关系:

$$\begin{aligned} & \left| \int_{J_i} \varphi_n(t) \frac{\sin(m + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt \right| \\ &= \left| \int_{J_i} \varphi_n(t) \frac{\sin(m + \frac{1}{2})(A-x)}{2 \sin \frac{1}{2}(A-x)} dt + \int_{J_i} \varphi_n(t) \frac{\sin(m + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt - \int_{J_i} \varphi_n(t) \frac{\sin(m + \frac{1}{2})(A-x)}{2 \sin \frac{1}{2}(A-x)} dt \right| \\ &\geq \left| \int_{J_i} \varphi_n(t) \frac{\sin(m + \frac{1}{2})(A-x)}{2 \sin \frac{1}{2}(A-x)} dt \right| - \left| \int_{J_i} \varphi_n(t) \left(\frac{\sin(m + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} - \frac{\sin(m + \frac{1}{2})(A-x)}{2 \sin \frac{1}{2}(A-x)} \right) dt \right| \end{aligned}$$

选择这样的正实数 A , 满足 $(m + \frac{1}{2})A = 2\ell\pi$ 且 $0 < |A-x| \leq r_i$, 于是上面第一项化为

$$\left| \int_{J_i} \varphi_n(t) \frac{\sin(m + \frac{1}{2})(A-x)}{2 \sin \frac{1}{2}(A-x)} dt \right| = \left| \frac{\sin(m + \frac{1}{2})(A-x)}{2 \sin \frac{1}{2}(A-x)} \right| \int_{J_i} \varphi_n(t) dt \geq \frac{2 |\sin(m + \frac{1}{2})x|}{nr_i}$$

这里利用了不等式 $|\sin \frac{1}{2}(A-x)| \leq \frac{1}{2} |A-x| \leq \frac{1}{2} r_i$ 进行放缩, 从后文中可以看到巧妙地设置 A 和 x 的取值范围, 可以从 nr_i 中可以“构造”出调和级数。

第二项积分的估计需要一些技巧, 注意到下面恒等式

$$\frac{1}{2} + \cos t + \cos 2t + \cdots + \cos mt = \frac{\sin(m + \frac{1}{2})t}{2 \sin \frac{1}{2}t} =: D_m(t)$$

由此不难得到导数不等式: $|D'_m(t)| \leq m^2$.

令 $g(t) = D_m(t-x)$, 由拉格朗日中值定理得到

$$g(t) - g(A) = g'(\xi)(t-A)$$

代入到上面第二项积分中, 得到不等式估计

$$\begin{aligned} & \left| \int_{J_i} \varphi_n(t) \left(\frac{\sin(m + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} - \frac{\sin(m + \frac{1}{2})(A-x)}{2 \sin \frac{1}{2}(A-x)} \right) dt \right| \\ &= \left| \int_{J_i} \varphi_n(t) g'(\xi)(t-A) dt \right| \leq m^2 \int_{J_i} \varphi_n(t) dt = \frac{2m^2 d}{n} \end{aligned}$$

这里 $0 < |t - A| \leq d$.

综合上面的结果, 得到一个核心不等式:

$$\left| \int_{J_i} \varphi_n(t) \frac{\sin(m + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt \right| \geq \frac{2 |\sin(m + \frac{1}{2})x|}{nr_i} - \frac{2m^2 d}{n}$$

柯尔莫哥洛夫想法的精妙之处在于, 通过适当设置 x, t, A 这三者之间的关系, 使得上面的不等式中“出现”了调和级数。具体构造工作十分具有技巧性, 笔者相信这是大量试探性 (甚至多数没有意义) 的艰苦计算后才获得的。

引理: 对每个正整数 n , 可以构造出函数序列 $\varphi_n(x)$ 和集合 $E_n \subset [0, 2\pi]$, 以及正实数 M_n 和整数 q_n , 满足列条件:

1. $\varphi_n(x) \geq 0, \int_0^{2\pi} \varphi_n(x) dx = 2$;
2. φ_n 在 $[0, 2\pi]$ 上是有界变差函数;
3. 测度 $\lim_{n \rightarrow \infty} m(E_n) = 2\pi, \lim_{n \rightarrow \infty} M_n = \infty$;
4. 对于任意一点 $x_0 \in E_n$, 存在标号 $p_{n, x_0} \leq q_n$, 使得 φ_n 在该点的 Fourier 级数部分和满足 $|S_{p_{n, x_0}}(x_0, \varphi_n)| \geq M_n$.

引理的证明是归纳构造式的。

在闭区间 $[0, 2\pi]$ 上选取 n 个点, 记 $A_k = 4k\pi / (2n+1)$, 这里下标范围是 $k=1, 2, \dots, n$. 取奇整数 $\lambda_1=1, \lambda_2, \dots, \lambda_n$, 后面还需要补充一些假设条件。按如下方式定义一系列数 $\{m_k\}$ 与有限闭区间序列 J_k :

$$m_1 = n, 2m_k + 1 = \lambda_k(2n+1), J_k = \left[A_k - \frac{1}{m_k^2}, A_k + \frac{1}{m_k^2} \right]$$

由此得到一个不等式 $m_k \geq n$.

再构造“无界矩形脉冲函数序列 $\varphi_n(t)$ ”如下:

$$\varphi_n(t) = \begin{cases} m_k^2 / n, & t \in J_k \\ 0, & t \in [0, 2\pi] - \bigcup_{k=1}^n J_k \end{cases}$$

显然积分满足引理中的 1 和 2

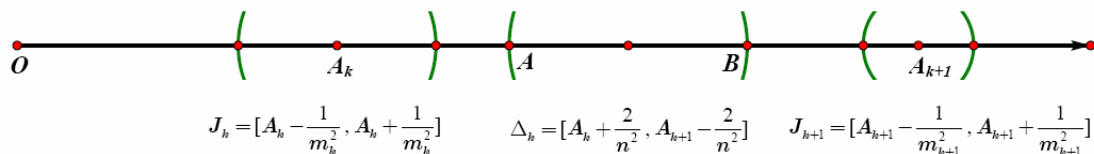
$$\int_{J_i} \varphi_n(t) dt = \frac{2}{n}$$

在此基础上，我们的目标是要构造一个正测度集 E （测度尽可能大），使得 $\forall x \in E$ ，对应 Fourier 级数部分和 $S_m(x, f)$ 是无界数列。同时还应该使得 E 与闭区间序列 J_i 的交集为空。

定义闭区间序列 $\{\Delta_i\}$

$$\Delta_i = \left[A_i + \frac{2}{n^2}, A_{i+1} - \frac{2}{n^2} \right], \quad i = 1, 2, \dots, n-1$$

建议读者多画画图，比如下图：



不难发现以下性质：当整数 n 固定时，有限闭区间序列 J_k 的半径是逐渐递减的，而有限闭区间序列 Δ_k 的半径保持不变。两个闭区间序列 $\{\Delta_i\}$ 与 $\{J_k\}$ 互不相交。这里选择增量 $1/m_k^2$ 和 $2/n^2$ 是出于收敛性考虑。

有了上面的铺垫可以证明引理了。假设 m_1, m_2, \dots, m_{k-1} 已经取定，由 Riemann-Lebesgue 引理知，可以选取充分大的整数 m_k ，使得 $\forall x \in \Delta_{k-1}$ ，

$$\left| \int_{\bigcup_{i=1}^{k-1} J_i} \varphi_n(t) \frac{\sin(m_k + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt \right| \leq 1$$

这样选择 x, t 不会使得分母产生奇异性，因为 $\forall x \in \Delta_{k-1}$ 以及 $\forall t \in \bigcup_{i=1}^{k-1} J_i$ 时，距离

$|t-x| \geq \rho_n > 0$ ，其中常数 ρ_n 仅与 n 有关。

所以可以用归纳法构造出数列 $\{m_i\}$ 。

上面讨论的是 $i < k$ 的情形。当 $i \geq k$ 且 $\forall x \in \Delta_{k-1}$ ，可以得到距离关系式

$$0 < A_i - x \leq A_i - A_{k-1} = (i-k+1) \frac{4\pi}{2n+1}$$

$$\frac{2m_k+1}{2}(A_i - A_k) = (i-k)2\pi\lambda_k$$

$$|t - A_i| \leq \frac{1}{m_i^2}$$

对照前文, 取 $A = A_i, r_i = A_i - A_{k-1} = (i-k+1)\frac{4\pi}{2n+1}$ 以及 $d = \frac{1}{m_i^2}$, 得到不等式

估计 (注意 $i \geq k, m_i \geq m_k$)

$$\left| \frac{1}{\pi} \int_{J_i} \varphi_n(t) \frac{\sin(m_k + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt \right| \geq \frac{2 |\sin(m_k + \frac{1}{2})x|}{n\pi r_i} - \frac{2m_k^2 d}{n\pi} \geq \frac{|\sin(m_k + \frac{1}{2})x|}{\pi^2(i-k+1)} - \frac{2}{n\pi}$$

即不等式

$$\left| \frac{1}{\pi} \int_{J_i} \varphi_n(t) \frac{\sin(m_k + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt \right| \geq \frac{|\sin(m_k + \frac{1}{2})x|}{\pi^2(i-k+1)} - \frac{2}{n\pi}$$

注意上面不等式中第一项里“出现”了与 i 有关的调和级数通项, 后面证明发散时用到了调和级数相关数学性质。

我们得到函数 $\varphi_n(t)$ 在点 $x \in \Delta_{k-1}, k=2, 3, \dots, n$ 处的 Fourier 级数部分和

$$\begin{aligned} |S_{m_k}(x, \varphi_n)| &= \left| \frac{1}{\pi} \int_0^{2\pi} \varphi_n(t) \frac{\sin(m_k + \frac{1}{2})(t-x)}{2 \sin \frac{1}{2}(t-x)} dt \right| = \left| \int_{\bigcup_{i=1}^{k-1} J_i} + \int_{\bigcup_{i=k}^n J_i} \right| \\ &\geq \frac{|\sin(m_k + \frac{1}{2})x|}{\pi^2} \sum_{i=k}^n \frac{1}{i-k+1} - \frac{2}{\pi} - 1 > \frac{|\sin(m_k + \frac{1}{2})x|}{\pi^2} \sum_{i=1}^{n-k+1} \frac{1}{i} - 2 \end{aligned}$$

由上面不等式可知, 当 $2 \leq k \leq n - \sqrt{n}, x \in \Delta_{k-1}$ 时

$$|S_{m_k}(x, \varphi_n)| \geq \frac{1}{2\pi^2} |\sin(m_k + \frac{1}{2})x| \ln n - 2$$

这里利用了调和级数里的近似关系 (再取 $k = n - \sqrt{n}$)

$$\sum_{i=1}^{n-k+1} \frac{1}{i} > \ln(n-k+1) > \ln \sqrt{n} = \frac{1}{2} \ln n$$

由上面推导可以看出, 为研究发散性, 可考虑 Δ_{k-1} 的满足下面条件的子集:

$$\sigma_{k-1} = \{x \in \Delta_{k-1} : \frac{1}{2\pi^2} |\sin(m_k + \frac{1}{2})x| \geq \frac{1}{\sqrt{\ln n}}\}$$

令 $q_n = m_n, M_n = \sqrt{\ln n} - 2$ 以及 $E_n = \bigcup_{i=1}^{n-[\sqrt{n}]-1} \sigma_i$, 则 $\lim_{n \rightarrow \infty} M_n = \infty$. 下面进一步证

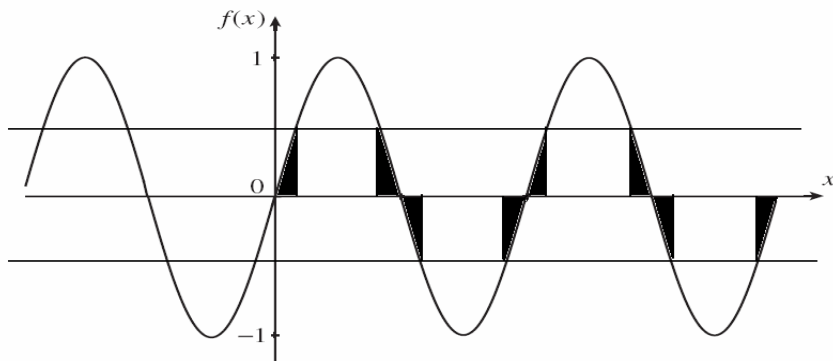
明测度 $\lim_{n \rightarrow \infty} m(E_n) = 2\pi$.

一种证明思路是考察满足下面不等式的点集

$$\frac{1}{2\pi^2} |\sin(m_k + \frac{1}{2})x| \leq \frac{1}{\sqrt{\ln n}} \quad (*)$$

动手画一画图像就知道，满足上面不等式的点集，是零点周围的一些对称闭区间的并集。其中零点指下面函数的零点

$$y = \sin(m_k + \frac{1}{2})x = \sin \lambda_k(2n+1) \frac{x}{2}$$



满足不等式(*)的点集(黑色部分)

每个小区间中恰有一个零点，小闭区间长度都相等，记为 δ ，则不难算出

$$\delta = 2 \frac{\arcsin \frac{2\pi^2}{\sqrt{\ln n}}}{\lambda_k(2n+1)/2} = O\left(\frac{1}{\lambda_k n \sqrt{\ln n}}\right)$$

注意到函数

$$y = \sin(m_k + \frac{1}{2})x = \sin \lambda_k(2n+1) \frac{x}{2}$$

在区间 $[0, 2\pi]$ 上共有 $\lambda_k(2n+1)+1$ 个零点，且相邻两个零点的距离是 $2\pi / \lambda_k(2n+1)$ ，而区间 Δ_{k-1} 的长度为 $4\pi / (2n+1) - 4/n^2$ 。所以函数 y 在区间 Δ_{k-1} 上至多有 $N = O(\lambda_k)$ 个零点。因此满足下面不等式

$$\frac{1}{2\pi^2} |\sin(m_k + \frac{1}{2})x| \leq \frac{1}{\sqrt{\ln n}}$$

的点集对应测度为

$$\delta N = O\left(\frac{1}{n\sqrt{\ln n}}\right)$$

于是

$$\lim_{n \rightarrow \infty} E_n = \lim_{n \rightarrow \infty} \sum_{i=1}^{n-[\sqrt{n}]-1} \left[\frac{4\pi}{2n+1} - \frac{4}{n^2} - O\left(\frac{1}{n\sqrt{\ln n}}\right) \right] = 2\pi$$

到此，我们证明了引理成立。

引理证明之后可以考虑构造反例了。柯尔莫哥洛夫的想法是把无界矩形脉冲函数序列 $\varphi_n(t)$ “叠加”到一起构造出一个 Lebesgue 可积函数 $\Phi(x)$ ，使得对应的 Fourier 级数发散。注意到是函数序列 $\varphi_n(t)$ 是无界的，因此必须设法加上一些系数来“约束”其行为（收敛性考虑），同时使得 Fourier 级数发散。

经过大量反复尝试之后，柯尔莫哥洛夫找到了一种函数子列构造法：

注意到 $M_n \rightarrow \infty$ ，因此利用归纳法可以构造一个递增的子列 $\{n_k\}$ ，满足：

1. 当 $1 \leq i \leq k-1$ 时， $q_{n_i} \leq \sqrt{M_{n_k}} / 2^k$ ；
2. $\sum_{i=1}^{k-1} \max_{m, x} |S_m(x, \varphi_{n_i})| \leq \frac{1}{2} \sqrt{M_{n_k}}$

条件 2 成立用到这样一个结论（证明见《傅里叶分析及其应用》）：

因为 φ_{n_i} 是有界变差函数，所以存在正的常数 B_{n_i} ，满足

$$\max_{m, x} |S_m(x, \varphi_{n_i})| \leq B_{n_i} < \infty$$

定义柯尔莫哥洛夫函数

$$\Phi(x) = \sum_{k=1}^{\infty} \frac{\varphi_{n_k}(x)}{\sqrt{M_{n_k}}}$$

根据实变函数中的 Levi 定理，若非负 L 可积递增函数列 f_n 的积分有上确界

$$\sup \int_E f_n < \infty$$

则存在可积函数 f ，使得 f_n 几乎处处收敛到函数 f ，且

$$\lim_{n \rightarrow \infty} \int_E f_n = \int_E f$$

注：上述 Levi 定理的证明详见徐森林的《实变函数论》。

将 Levi 定理应用到函数 $\Phi(x)$ 上，得到

$$S_i(x, \Phi) = \sum_{k=1}^{\infty} S_i(x, \varphi_{n_k}) / \sqrt{M_{n_k}}$$

设 $E = \overline{\lim_{k \rightarrow \infty} E_{n_k}}$ ，则 $m(E) = 2\pi$ 。根据上极限定义， $\forall x_0 \in E$ ，则 $x_0 \in E_{n_j}$ 对无穷

多个 j 成立。我们证明可积函数 $\Phi(x)$ 的 Fourier 级数在 x_0 点发散。

由等式

$$\begin{aligned} S_{p_j, x_0}(x_0, \Phi) &= \sum_{k=1}^{j-1} S_{p_j, x_0}(x_0, \varphi_{n_k}) / \sqrt{M_{n_k}} + S_{p_j, x_0}(x_0, \varphi_{n_j}) / \sqrt{M_{n_j}} \\ &\quad + \sum_{k=j+1}^{\infty} S_{p_j, x_0}(x_0, \varphi_{n_k}) / \sqrt{M_{n_k}} \end{aligned}$$

以及不等式

$$|S_k(x, f)| \leq (k + \frac{1}{2}) \int_0^{2\pi} |f(t)| dt$$

可知

$$\begin{aligned} |S_{p_j, x_0}(x_0, \Phi)| &\geq |S_{p_j, x_0}(x_0, \varphi_{n_j})| / \sqrt{M_{n_j}} \\ &\quad - \sum_{k=1}^{j-1} |S_{p_j, x_0}(x_0, \varphi_{n_k})| / \sqrt{M_{n_k}} - \sum_{k=j+1}^{\infty} |S_{p_j, x_0}(x_0, \varphi_{n_k})| / \sqrt{M_{n_k}} \\ &\geq \sqrt{M_{n_j}} - \frac{1}{2} \sqrt{M_{n_j}} - \sum_{k=j+1}^{\infty} 2(p_{n_j} + \frac{1}{2}) / \sqrt{M_{n_k}} \end{aligned}$$

但是, 由条件 1 可知

$$2p_{n_j} / \sqrt{M_{n_k}} \leq 1/2^{k-1}, \text{ 特别地, } 1/\sqrt{M_{n_k}} \leq 1/2^k$$

这说明上面不等式最后一项对应的级数是收敛的。

再由 $\sqrt{M_{n_j}} \rightarrow \infty$ 可知, 上极限 $\lim_{i \rightarrow \infty} |S_i(x_0, \Phi)| = +\infty, \forall x_0 \in E$.

于是柯尔莫哥洛夫函数 $\Phi(x) \in L[0, 2\pi]$, 但其对应的 Fourier 级数在 $[0, 2\pi]$ 上几乎处处无界发散。

评注: 柯尔莫哥洛夫后来还构造出绝对可积函数 $f(x)$, 其对应的 Fourier 级数在 $[0, 2\pi]$ 上处处无界发散, 其他人也构造了 Fourier 级数有界发散的反例, 这两个例子详见陈建功的《三角级数论》和汪林的《实分析中的反例》, 这里不再讨论了。笔者应对看完上面证明过程的读者表示感谢, 毕竟看完一大堆不等式硬估计, 感觉非常乏味枯燥, 费时费力。但这就是分析, 充满了大量的计算与估计。许多美好的数学定理背后, 就是这些枯燥的估计的结果。就像很多商业成功案例一样, 外表看起来很光鲜, 背后却凝结无数乏味和艰辛的工作。当然, 有的定理证明基本看不懂, 但定理本身好理解 (如费马大定理和庞加莱猜想), 这时记住结论也可以。多读著名定理不能机械地理解成读定理的证明。

天下事没有免费的午餐, 越是大成果, 越是绝技, 必然是付出艰辛努力的结果。笔者给出柯尔莫哥洛夫少年成名工作的详细数学推导, 正是以实际案例表达这种看法, 愿有心人能够体会数学证明之外的深意。

第13章 重视反例

反例是数学中的重要组成部分，但在教学中因为某些原因被低估了，原因或者因为有的反例证明太难（所用数学工具较深），或者是纯粹技巧性的构造（没有一般性），甚至或者认为“破坏”了数学的和谐性（如处处连续处处不可导函数）等等。

按笔者观点，反例属于数学中的“否定性定理”范畴，可能在某种程度上限制了我们认识数学的边界。但换一个角度讲，很多反例的存在使得数学多了层次性，反而激发了活力，打开了一个新的世界，比如：连续函数--光滑函数--解析函数。

最有名的反例之一要数“单位正方形的对角线长度不能表示成两个正整数之比”了，即 $\sqrt{2}$ 是无理数，这个反例引发了无理数井喷。如果耐心一些，细数数学分析教材（如《数学分析新讲》）中的反例，还是挺多的。如举出一个光滑函数，不存在泰勒级数，这个反例是函数：

$$f(x) = e^{-1/x}, x > 0; f(x) = 0, x \leq 0.$$

该函数在 $x=0$ 点的任意阶导数均为0，因此这个函数在 $x=0$ 点的泰勒级数恒为0，矛盾。

其它例子，如级数 $\sum a_n$ 收敛，则有 $\lim a_n = 0$ 。那么对收敛的广义积分

$$\int_0^{\infty} f(x) dx < \infty \text{ 似乎应该有 } \lim_{x \rightarrow \infty} f(x) = 0?$$

可惜这个结论不成立，如取 $f(x) = \sin x^2$ ，显然极限不存在。

及时要求 $f(x)$ 非负连续，那么仍然可能有

$$\lim_{x \rightarrow +\infty} f(x) = +\infty$$

这方面的经典反例是：广义积分

$$\int_0^{\infty} f(x) dx = \int_0^{\infty} \frac{x dx}{1 + x^6 \sin^2 x}$$

该广义积分收敛，但是 $f(n\pi) = n\pi$ ，显然 $\lim_{x \rightarrow \infty} f(x) = +\infty$ 。这个函数的图像是在点 $x = n\pi$ 处图像急剧上升，在两侧急剧递减。直观上，积分可以看做求面积，而面积可以近似看做是“宽 \times 高”，高度很大，但宽度很窄，相乘之后的数值仍然很小，这就是该反例的几何思路。

级数方面, 众所周知的结论是 $\sum_{n=1}^{\infty} \frac{1}{n} = +\infty$, 但是 $\sum_{n=1}^{\infty} \frac{1}{n^p} < +\infty, p > 1$. 于是

猜想: 对正项级数, 当 $a_n = o(\frac{1}{n})$, 是否有级数 $\sum_{n=1}^{\infty} a_n < +\infty$. 直观上, 似乎是对的,

但是级数 $\sum_{n=2}^{\infty} \frac{1}{n \ln n} = +\infty$ 就是这个猜想的反例。

上述几个例子粗看都是与直观相违背的, 这提醒我们在学习数学中不能丢弃直观思维的启发, 但直观想象不能代替严格的数学证明, 想当然的习惯要不得。当然, 分析教材中抛开几何直观, 而采用纯逻辑推导的写法值得商榷。

可能有朋友还是有这种感觉, 上面的例子还是有些做做, 似乎是 show 技巧。其实, 随着学习的深入, 特别是从一个高的视角“重温微积分”时, 会有新的感悟和发现。重温则不能再走老路, 必须注入新的想法。

比如, 泰勒级数那个反例引了解析函数的概念 (参考复变函数教材), 反例中 $x=0$ 是函数 $f(x)$ 的本性奇点, 很多重要的定理就是围绕奇点展开的, 如本性奇点的皮卡大定理。

还是说一些熟悉的例子吧, 先说说狄利克雷函数

$$D(x) = 1, x \in \mathbb{Q}; D(x) = 0, x \in \mathbb{R} - \mathbb{Q}$$

将 $D(x)$ 限制在区间 $[0, 1]$ 上就是一个著名的黎曼不可积函数的反例。

换一个角度看待这个反例。因为有理数可数, 将区间 $[0, 1]$ 上的有理数编号为 q_1, q_2, q_3, \dots , 构造函数列

$$D_n(x) = \begin{cases} 1, & x = q_1, q_2, q_3, \dots, q_n \\ 0, & \text{其它} \end{cases}$$

显然函数列 $D_n(x)$ 是 $[0, 1]$ 上的黎曼可积函数, 满足 $D_n(x) \rightarrow D(x)$ 且

$$\int_0^1 D_n(x) dx = 0$$

但是

$$0 = \lim_{n \rightarrow \infty} \int_0^1 D_n(x) dx \neq \int_0^1 \lim_{n \rightarrow \infty} D_n(x) dx = \int_0^1 D(x) dx$$

因为右边的狄利克雷函数不是 **Riemann** 可积的。

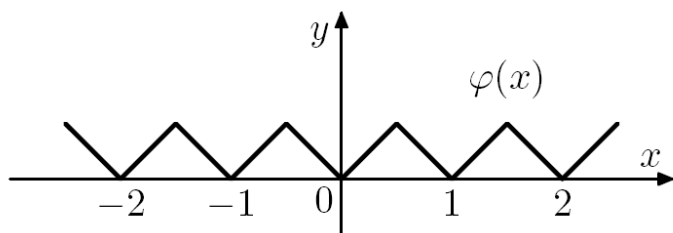
关于极限运算与积分运算什么时候可交换的问题研究，引出了控制收敛定理。而对极限函数的积分研究，是 Lebesgue 积分的来源之一

$$\lim_{n \rightarrow \infty} \int_0^1 D_n(x) dx = (L) \int_0^1 D(x) dx$$

此时极限函数是 L 可积的了。

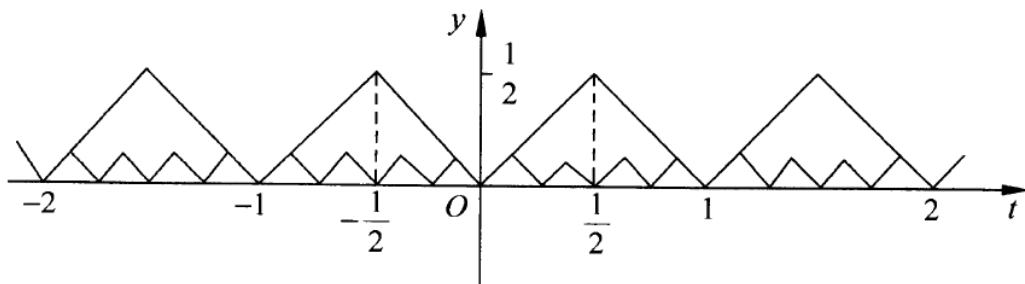
上面几个反例都是相对容易的，但也有更难的反例，如处处连续处处不可导函数。这个函数有很多种构造方式，比较直观的是《新讲》中收录的反例，构造如下：记 $\varphi(x)$ 表示 x 与离它最近整数的距离，则 $\varphi(x+1) = \varphi(x)$ 。

动手画画图，发现这是一个“锯齿函数”



令 $f(x) = \sum_{n=0}^{\infty} \frac{\varphi(4^n x)}{4^n}$ $x \in \mathbb{R}$ ，此函数一致收敛，故为连续函数，这个函数

处处连续处处不可导，具体证明见本书附录。本章只给出一个直观描述：随着 n 不断增大，锯齿函数 $\varphi(4^n x)$ 不断加密，而这些锯齿函数的叠加就“曲折了”。



如果读者了解一些分形几何方面的知识会发现上面的反例函数图像就是一种分形几何，极限图像非常曲折，难以想象。

换一个角度，人们之前为什么没有认识到“处处连续处处不可导函数”的存在呢？一种想法是，通常人们受直观影响，想象中的函数都是分段单调的。而“处处连续处处不可导函数”是无处单调的连续函数，很难想象。证明很简单，根据实变函数中的定理：单调函数几乎处处可导。利用泛函分析中的贝尔“纲定理”，可知道在连续函数中，“处处连续处处不可导函数”占绝大多数。这一惊人结论堪比“超越数比代数数多得多”的结论。

《新讲》中还给出了填满正方形的皮亚诺曲线例子，以及直圆柱的内接折面面积可以任意大（对比下曲线内接折线，可知这个反例很出乎意料）等反例，这里不一一介绍了。

上面的反例相对简单一些，分析中还有更复杂一些的反例。这里介绍两个反例，第一个是柯尔莫哥洛夫构造的 $L[0, 2\pi]$ 中几乎处处无界发散的 Fourier 级数反例，我们在第 12 章中详细介绍了这个著名反例。另一个反例是 H.Levy 构造的无解线性偏微分方程反例：

$$-\frac{\partial u}{\partial x} - i \frac{\partial u}{\partial y} + 2i(x + iy) \frac{\partial u}{\partial z} = f(x, y, z)$$

这里 $f(x, y, z) \in C^\infty(\mathbb{R}^3)$ 是一个光滑函数。此前人们相信线性偏微分方程一定有解的，但是 H.Levy 构造的这个反例引起了人们的思考，重新审视了线性偏微分方程理论。对这个反例具体证明感兴趣的读者，可以参考 F.约翰的《偏微分方程》一书第八章。

最后，再科普一个非常深刻的反例。这个反例太有名了，以至于数学系大三水平都应该知道-----她就是传说中的“米尔诺 7 维怪球”，怪球的发现被视作是微分拓扑学正式建立的标志。陈省身老爷子在《微分几何讲义》中给出了米尔诺最初的反例构造方法，不过太过于拓扑技巧化，这里采用后人发现的更简单而深刻的 7 维怪球反例，大致方法如下：

约定如下讨论的流形都是紧致可定向的。根据 Thom 的工作，设 M^7 是一个闭 7 维流形，则存在某个 8 维流形 B^8 ，以 M^7 为其边界，即 $M^7 = \partial B^8$ 。应该指出，满足 $M^7 = \partial B^8$ 的 B^8 不唯一。

米尔诺利用上述结论，并结合示性类理论和符号差定理，对 M^7 巧妙定义了一个 mod 7 整数微分同胚不变量 $\lambda(M^7)$ 。在 mod 7 意义下，整数 $\lambda(M^7)$ 不依赖于 B^8 的具体选取，故为 M^7 的微分同胚不变量。

对于标准的 7 维球面 S^7 ，有 $\lambda(S^7) = 0$ 。米尔诺构造了这样一个 7 维流形 M^7 ，它拓扑同胚于标准球面 S^7 ，但是 $\lambda(M^7) \neq 0$ 。于是利用反证法可知这个“7 维怪球 M^7 ”不微分同胚于标准球面 S^7 。在拓扑等价意义下，也称标准球面 S^7 上具有奇异微分结构。

7 维怪球的构造：定义解析函数：

$$f(z_1, z_2, z_3, z_4, z_5) = z_1^5 + z_2^3 + z_3^2 + z_4^2 + z_5^2, \quad z_i \in \mathbb{C}$$

方程 $f(z_1, z_2, z_3, z_4, z_5) = 0$ 的解集确定了一个通过原点的 8 维超曲面 $M^8 \subset \mathbb{C}^5$. 再设 $S^9: |z_1|^2 + |z_2|^2 + |z_3|^2 + |z_4|^2 + |z_5|^2 = 1$ 为标准的 9 维单位球面。

最后，令 $K^7 = M^8 \cap S^9$. 拓扑上可以证明 M^8 与 S^9 横截相交，于是 K^7 是一个 7 维光滑流形。进一步证明光滑流形 K^7 满足： K^7 拓扑同胚于标准球面 S^7 ，但是 $\lambda(K^7) \neq 0$ ，这就是一个 7 维怪球。

上述例子与米尔诺原始论文中的例子等价（微分同胚），但要简单和深刻的多。深入研究发现，超曲面 M^8 不是处处光滑的，它在原点 $O \in \mathbb{C}^5$ 有一个孤立奇点（各偏导数等于 0），正是这个奇点导致了 $K^7 = M^8 \cap S^9$ 的怪异（具有怪异微分结构）。这个例子也说明了超曲面奇点附近的拓扑结构非常奇妙，后来发展出了所谓的“奇点理论”。

上述工作是几何（拓扑）与分析结合的典范。进一步研究发现，标准球面 S^7 共有 28 种互不相容的微分结构，这 28 种怪球可表示如下：

$$\sum(6r-1, 3, 2, 2, 2): z_1^{6r-1} + z_2^3 + z_3^2 + z_4^2 + z_5^2 = 0 \cap S^9$$

其中， $r = 1, 2, 3, \dots, 28$. 显然 $K^7 = \sum(5, 3, 2, 2, 2)$.

在 7 维怪球上发现了怪异微分结构之后，人们继续研究其它维数的怪球有无微分结构。如果有，有多少种？目前这方面工作可见下表：

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
\mathcal{S}_n	1	1	1	?	1	1	28	2	2.4	6	992	1	3	2	2.8128	2	2.8	2.8

球面微分结构分类表

这里 $n = 3$ 情况是 Perelman 证明 Poincare 猜想的结果的推论，即对 3 维球面同胚与微分同胚等价。目前剩下的难题是：

4 维球面 S^4 上是否存在怪异微分结构？

但是，7 维球面球面比较已经比较复杂了，而且维数多少有些特殊（至少不太整齐）。比 n 维球面更简单的非平凡流形就是 \mathbb{R}^n 了，这方面的一个著名结论

是：

当 $n \neq 4$ 时，欧氏空间 \mathbf{R}^n 有惟一的微分结构。

这个结论没有出乎意料，想想也是很自然的事情，毕竟欧氏空间整体上算是比较“平坦”的了。不过当 $n = 4$ 时，却难住了大家，很长时间没有实质性的突破。或者说没有找到合适的“不变量”来分类。

关于 \mathbf{R}^4 空间的实质性突破来自数学家 Donaldson，他在研究 Yang-Mills 方程（来自数学物理领域）时发现了一组新的不变量，可以对四维流形进行细致的分类。利用此不变量，人们相继在 \mathbf{R}^4 上发现了若干怪异微分结构。特别地，利用反证法证明了：

\mathbf{R}^4 上存在不可数个互不相容的微分结构。

个人感觉 \mathbf{R}^4 的冲击力远超 7 维怪球的冲击力，加深了人们对空间的认识。

不过有一点挺可惜，怪异 \mathbf{R}^4 的构造例子中没有找到类似 Milnor 怪球

$$z_1^5 + z_2^3 + z_3^2 + z_4^2 + z_5^2 = 0 \cap S^9$$

这样简单的例子，已知的几个例子都需要很高的拓扑分割与粘合技巧。

评注：抛开分析这个框架，数学中的反例还是很多的，不过不可能都列举了。从简单到复杂，举了这么多例子，无非想强调不能简单地视“反例”仅仅为“反例”，起负面否定效应其实，若能换一个角度看反例，也许能打开另一扇数学之窗，这方面 7 维怪球就是代表。Milnor 关于怪球的文章如下：

ANNALS OF MATHEMATICS
Vol. 64, No. 2, September, 1956
Printed in U.S.A.

ON MANIFOLDS HOMEOMORPHIC TO THE 7-SPHERE

BY JOHN MILNOR¹

(Received June 14, 1956)

这篇文章长度只有 6 页，而他本人也因这篇文章的工作获得菲尔兹奖。印象中，获得菲尔兹的大牛中，获奖文章长度还没有低于 6 页的，这也算一个记录了。

附录：

打个比方，如果说本书正文相当于投资原理的话，附录则相当于具体案例分析。有一句老话，数学中的魔鬼隐藏在细节中。说的是，学数学，你若不亲自动手算一算，推一推，你是很难发现问题的。很多问题看起来很简单，但真正证明起来发现比想象中困难的多。泛泛而谈总是很难有收获的。

本书附录里收集了 7 个经典定理，个别文章是笔者的数学笔记。因为文章较长，都放到正文中显得过于累赘，于是采用有的放矢的方法。一些证明放到正文里，一些放到附录里。比如原文已经包括了数学大师柯尔莫哥洛夫的少年成名之作：

$L[0, 2\pi]$ 中几乎处处无界发散的 Fourier 级数反例。

文章已经很长了，所以同是著名定理，只能把素数定理的证明放到附录里。附录里的文章部分与正文重复，主要是例子太经典了。

这七篇文章可以作为正文思想的进一步阐述，现简单说明一下：

1. 微积分基本定理新讲。本文主要体现了数学的联系与统一性；
2. 多项式逼近连续函数的折线逼近证明。分析中逼近思想的综合体现；
3. 布劳威尔不动点定理。分析中逼近与变换思想的综合体现；
4. 处处连续处处不可导函数。经典的反例与构造法代表；
5. 误差与正态分布。逆向思维与形式推导的经典例子；
6. 重积分变量代换公式。逼近、变换和分解三大分析思想的综合运用；
7. 素数定理的简单解析证明笔记。逼近、变换和分解三大分析思想的综合运用，同时也是形式推导和硬算（指结论的发现过程）的代表例子，本身还是历史上的著名定理。

本书正文内容加上附录里的七篇文章，基本可以代表笔者对数学分析，乃至整个数学的理解。相信会对读者学习数学给予启示和帮助。

微积分基本定理新讲

SCIbird

本文主要讨论微积分基本定理成立的若干条件，核心是围绕下面的“微积分基本定理”（也称 N-L 公式）来展开：

设 $F(x) \in C[a, b]$ 是连续函数，满足 $F'(x) = f(x)$ 且 $f(x)$ 是 Lebesgue 可积的，则

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

若无特别说明，涉及到的积分都默认为 Lebesgue 积分。

从微积分基本定理可以看出，定理成立的前提是函数可积且存在原函数，缺一不可。由此形成两条探索路线：

1. 已知 $f(x)$ 是 Lebesgue 可积的，还需要什么条件保证原函数存在；
2. 已知 $f(x)$ 存在原函数 $F(x) \in C[a, b]$ ，还需要什么条件保证 $f(x)$ 是可积的。

先来看看路线 1，由原函数的存在性和导数介值定理可知，被积函数 $f(x)$ 无第一类间断点。所以排除了类似阶梯函数这样具有第一类间断点的函数。在此前提下，分类讨论：

- a. 函数 $f(x)$ 无第二类间断点，此时 $f(x)$ 是连续函数，变上限积分

$$\int_a^x f(t) dt$$

就是 $f(x)$ 的一个原函数。

- b. 函数 $f(x)$ 具有第二类间断点，此时若无其它附加条件，不能保证原函数存在。比如狄利克雷函数

$$D(x) = \begin{cases} 1, & x \in \mathbf{Q} \\ 0, & x \in \mathbf{R} - \mathbf{Q} \end{cases}$$

$D(x)$ 是 Lebesgue 可积的，所有间断点都是第二类的，但不存在原函数。反证假设存在原函数，根据导数介值定理， $D(x)$ 应该具有价值性，矛盾！

那么需要对第二类间断点附加什么条件才能保证原函数存在？这方面进展似乎不多，就笔者了解。于是人们退而求其次，主要沿着路线 2 来研究 N-L 公式。

路线 2, 已知 $f(x)$ 存在原函数 $F(x) \in C[a, b]$. 先看看几个 **Riemann** 积分的经典例子:

存在原函数, 但非 **Riemann** 可积的函数 $f(x)$, 反例如下:

$$F(x) = \begin{cases} x^2 \sin \frac{1}{x^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

这个函数的导函数在 $[-1, 1]$ 上处处存在

$$f(x) = F'(x) = \begin{cases} 2x \sin \frac{1}{x^2} - \frac{2}{x^2} \cos \frac{1}{x^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

但因为导函数 $f(x)$ 在 $[-1, 1]$ 上无界, 因而不是 **Riemann** 可积的。

Volterra 曾经构造了一个著名的反例 $F(x)$, 其导函数 $f(x) = F'(x)$ 处处存在且有界, 但却不是 **Riemann** 可积的。这个例子中, $f(x)$ 的间断点的集合测度非零, 所以不是 **Riemann** 可积的。反例的构造细节请参考《数学分析问题研究与评注》一书。

存在原函数, 但非 **Lebesgue** 可积的函数 $f(x)$, 反例如下:

$$F(x) = \begin{cases} x^2 \cos \frac{\pi}{x^2}, & x > 0 \\ 0, & x = 0 \end{cases}$$

其导函数

$$f(x) = F'(x) = \begin{cases} 2x \cos \frac{\pi}{x^2} + \frac{2\pi}{x} \sin \frac{\pi}{x^2}, & x > 0 \\ 0, & x = 0 \end{cases}$$

导函数 $f(x)$ 在 $[0, 1]$ 上不是 **Lebesgue** 可积的。

所以问题化为: 设 $F(x) \in C[a, b]$ 是连续函数, 且存在导函数 $f(x) = F'(x)$, 那么原函数 $F(x)$ 满足什么条件时, $f(x)$ 是 **Lebesgue** 可积的且 **N-L** 公式成立。路线 2 获得了巨大的突破, 概括为如下定理:

设 $F(x) \in C[a, b]$ 是连续函数, 且存在导函数 $f(x) = F'(x)$, 若 $F(x)$ 满足“绝对连续”条件, 则 **N-L** 公式成立

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

绝对连续也称**强连续**，是比一致连续和连续还要强的条件，定义如下：

设 $f:[a, b] \rightarrow \mathbf{R}$ 是实函数，如果 $\forall \varepsilon > 0, \exists \delta > 0$ ，使得当 $[a, b]$ 中任意有限个互不相交的开区间 (a_i, b_i) ， $i = 1, 2, \dots, n$ 满足

$$\sum_{i=1}^n (b_i - a_i) < \delta$$

时，就有

$$\sum_{i=1}^n |f(b_i) - f(a_i)| < \varepsilon$$

则称 f 为 $[a, b]$ 上的**绝对连续函数**。

如果我们放宽 “ $f(x) = F'(x)$ 处处成立” 这一条件，只要求几乎处处成立，那么有如下更一般的定理：

设 $F(x)$ 是 “**绝对连续**” 函数，则几乎处处存在 **Lebesgue** 可积函数 $f(x) = F'(x)$ ，且满足

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

换句话说，当 $F(x)$ 是 “**绝对连续**” 时， $F'(x)$ 几乎处处存在，且 **Lebesgue** 可积。这一定理的证明请参考《实变函数》教材。

因此，站在原函数角度看，绝对连续条件解决了导函数可积问题，且 **N-L** 公式成立，路线 2 可以告一段落。不过，绝对连续这一条件用起来不太方便，能否用其它条件代替，以便于应用。这方面有一个实用定理：

设 $F(x) \in C[a, b]$ 是连续函数，满足 $F'(x) = f(x)$ 且 $f(x)$ 是 **Lebesgue** 可积的，则

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

这个定理见于一些《实变函数》教材中，实际应用足够了。但是，我们还有更强的版本，下面的定理允许 $F'(x)$ 有一些例外点。

“最强微积分基本定理”

设 $F(x) \in C[a, b]$ 是连续函数， $f(x) \in L[a, b]$ 是 **Lebesgue** 可积的，若至多除去一个可数集之外，有 $F'(x) = f(x)$ ，则

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

证明：不妨设集合 D 为可数集，使得 $F'(x) = f(x)$, $x \in \mathcal{A} = [a, b] - D$ ，先证明 $\forall u, v (u < v) \in [a, b]$ ，下列积分不等式

$$|F(u) - F(v)| \leq \int_u^v |f(x)| dx$$

成立。不失一般性，可以假设 $[u, v] = [a, b]$ ，这样做的合理性可以从后面的证明看出。同时可假设 $F(x)$ 在区间 $[a, b]$ 的端点处可导，否则做一闭区间序列 $[a_n, b_n] \rightarrow [a, b]$ 来逼近。

$\forall \varepsilon > 0$ 并固定，记 $c_i = i\varepsilon / (b - a)$, $i = 0, 1, 2, \dots$ ，定义

$$E_i = \{x \in \mathcal{A} : c_{i-1} \leq |f(x)| < c_i\} , i \in \mathbf{N}$$

因为 $f(x)$ 与 $|f(x)|$ 都是 Lebesgue 可积的，所以集合 E_i 可测。又 \mathcal{A} 是集合 E_i 的无交并集，所以集合 \mathcal{A} 的测度为

$$m(\mathcal{A}) = b - a = \sum_{i=1}^{\infty} m(E_i)$$

以及不等式

$$c_{i-1} m(E_i) \leq \int_{E_i} |f(x)| dx \leq c_i m(E_i) , i \in \mathbf{N}$$

由此得到

$$0 \leq c_i m(E_i) - \int_{E_i} |f(x)| dx \leq \frac{\varepsilon}{b-a} m(E_i)$$

对于 $i \in \mathbf{N}$ ，存在有界开集 $G_i \subset \mathbf{R}$ ，使得

$$G_i \supset E_i \text{ 且 } m(G_i) \leq m(E_i) + \frac{\varepsilon}{c_i 2^i} , i \in \mathbf{N}$$

因为 D 为可数集，不妨设 $D = \{u_j : j \in \mathbf{N}\}$ 有了上面铺垫之后，下面定义函数 $H, M : [a, b] \rightarrow \mathbf{R}$ 且 $H(a) = M(a) = 0$ ，以及

$$H(x) = \sum_{i=1}^{\infty} c_i m(G_i \cap [a, x]) , M(x) = \sum_{u_j \in [a, x]} \frac{\varepsilon}{2^j}$$

其中 $a < x \leq b$ 。显然，函数 H, M 都是单调递增的。记上确界

$$\zeta = \sup\{x \in [a, b] : |F(x) - F(a)| \leq H(x) + M(x)\}$$

显然 a 属于上述集合（说明非空），所以上确界 ζ 存在。如果 $\zeta = b$ ，则结论得证。现假设 $\zeta < b$ ，下面分类讨论：

- (1) 若 $\zeta \in \mathcal{A}$ ，则存在 E_k ，使得 $\zeta \in E_k$ 对某个 k 成立，故 $|f(x)| < c_k$ 。由此根据导数定义，存在 $x_1 \in (\zeta, b)$ ，使得

$$[\zeta, x_1] \subset G_k \text{ 且 } |F(x_1) - F(\zeta)| < c_k(x_1 - \zeta)$$

因为 $F(x)$ 是连续函数, 所以 $|F(\zeta) - F(a)| \leq H(\zeta) + M(\zeta)$. 由此得到

$$\begin{aligned} |F(x_1) - F(a)| &\leq |F(x_1) - F(\zeta)| + |F(\zeta) - F(a)| \\ &< c_k(x_1 - \zeta) + H(\zeta) + M(\zeta) \end{aligned}$$

根据 $H(x)$ 定义, 有 $H(\zeta) + c_k(x_1 - \zeta) \leq H(x_1)$ 和 $M(\zeta) \leq M(x_1)$. 由此

$$|F(x_1) - F(a)| \leq H(x_1) + M(x_1)$$

这与 ζ 是上确界的最大性矛盾! (上面利用了 $[a, x_1] = [a, \zeta] \cup [\zeta, x_1]$)

(2) 若 $\zeta \in D$, 则对某个 $m \in \mathbb{N}$ 有 $\zeta = u_m$. 因为 $F(x)$ 是连续函数, 所以存在 $x_2 \in (\zeta, b)$ 使得

$$|F(x_2) - F(\zeta)| < \frac{\varepsilon}{2^m}$$

于是

$$\begin{aligned} |F(x_2) - F(a)| &\leq |F(x_2) - F(\zeta)| + |F(\zeta) - F(a)| \\ &< \frac{\varepsilon}{2^m} + H(\zeta) + M(\zeta) \end{aligned}$$

又 $M(\zeta) + (\frac{1}{2})^m \varepsilon \leq M(x_2)$, 所以 $|F(x_2) - F(a)| \leq H(x_2) + M(x_2)$, 同理, 矛盾!

综合 (1) (2) 讨论知 $\zeta = b$ 成立。

再利用不等式 $m(G_i \cap [a, x]) \leq m(G_i) \leq m(E_i) + c_i^{-1}(\frac{1}{2})^i \varepsilon$, 得到

$$|F(b) - F(a)| \leq H(b) + M(b) \leq \int_a^b |f(x)| dx + 3\varepsilon$$

因为 ε 是任意取定的, 所以

$$|F(b) - F(a)| \leq \int_a^b |f(x)| dx$$

最后证明函数 $F(x)$ 是绝对连续的。实际上 $f(x) \in L[a, b]$, 根据 Lebesgue 积分的绝对连续性, $\forall \varepsilon > 0, \exists \delta > 0$, 使得对任意可测集 $E \subset [a, b]$, $m(E) < \delta$, 则有不等式

$$\int_E |f(x)| dx < \varepsilon$$

根据上述结论, 任取 $[a, b]$ 中有限个互不相交的开区间 (a_i, b_i) , 满足

$$\sum_{i=1}^n (b_i - a_i) < \delta$$

令 $E = \bigcup (a_i, b_i)$, 则 $m(E) < \delta$, 于是

$$\sum_{i=1}^n |F(b_i) - F(a_i)| \leq \sum_{i=1}^n \int_{a_i}^{b_i} |f(x)| dx = \int_E |f(x)| dx < \varepsilon$$

这就证明了 $F(x)$ 是绝对连续函数，从而牛顿-莱布尼茨公式成立。

评注：上面的证明构思还是技巧性特别强的，特别是函数 H, M 的构造非常巧妙，同时也不太好想，所以定理的结论比证明更重要。这里的取上确界的 \sup 方法非常具有一般性，如用确界存在原理证明连续函数介值定理，该方法在处理一维连续函数问题时非常实用。

上述定理的证明引自下述文献，推荐大家读一读。

Fundamental Theorem of Calculus for Lebesgue Integration

J. J. Koliha

上面的定理之所以说是最强形式的微积分基本定理，是因为不可导集 D 只能加强为可数集。如果不可导集 D 为不可数零测集，则微积分基本定理不一定成立。这方面有一个著名的反例，即康托函数反例，这是一个单调递增函数，常见于《实变函数》教材。一般地，对单调递增函数 $f(x)$ ，只能取到不等式

$$\int_a^b f'(x) dx \leq f(b) - f(a)$$

微积分基本定理是自己高中时遇到的第一个深刻的数学分析公式，而且实用性也特别强，所以单独写一篇文章。多数情况下在 Riemann 积分框架内就够用了，证明也简单的多。

设 $f(x) \in R[a, b]$ ，且存在连续函数 $F(x) \in C[a, b]$ 满足 $F'(x) = f(x)$ ，即 $f(x)$ 存在原函数 $F(x)$ ，则

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

因为以上公式指出了微分运算与积分运算之间的互逆关系，所以被称作微积分基本定理。证明不难，直接利用 Riemann 积分的定义即可，任取区间 $[a, b]$ 的一个分割 $\{x_i\}$ ，并记 $x_0 = a, x_n = b$ ，则由拉格朗日中值定理，得

$$F(b) - F(a) = \sum_{k=1}^n F(x_k) - F(x_{k-1}) = \sum_{k=1}^n f(\xi_k) \Delta x_k$$

注意到等式右边即是 Riemann 和，取极限并利用极限的唯一性，得

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

这个证明并不复杂，一些工科教材也讲述定积分 **Riemann** 和的定义，但不知道为什么多数工科微积分教材没有收录进去上面的证明。

不过笔者认为，微积分基本定理最自然的解释还是物理的例子---距离&速度。

设直线上单向连续运动质点 t 时刻对应的距离函数为 $S(t)$ ，问在 $[t_1, t_2]$ 这段时间内质点运行的距离？将 $[t_1, t_2]$ 分成很多小时间段，每小段时间如此之短，对应速度可视为不变，则

$$S(t_2) - S(t_1) \approx \sum_i v(t_i) \Delta t_i \rightarrow \int_{t_1}^{t_2} v(t) dt$$

不难理解 $S'(t) = v(t)$ 关系，即

$$v(t) := \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} = S'(t)$$

以上一种比较自然地理解牛顿-莱布尼茨公式的思路。

多项式逼近连续函数的折线逼近证明

魏尔斯特拉斯的多项式逼近连续函数定理断言：

闭区间 $[a, b]$ 上的连续函数可由多项式一致逼近。

打个比方，这一定理的意义好比用有理数可以逼近任意实数。定理的证明已经有好几种，从初等方法到高等方法都有，不过笔者还是觉得勒贝格的折线逼近证明法最自然。本文就介绍这一方法，内容摘自张筑生老师的《数学分析新讲》第三册。

“折线逼近法”证明的大体思路是：闭区间上的连续函数可以用折线函数逼近（比较直观）。而折线函数可以用函数 x 和 $|x|$ 进行线性组合而得到。考虑到 $|x| = \sqrt{x^2}$ ，经过一些代换，利用二项级数展开，得到了一致收敛的幂级数，这说明 $|x|$ 可以用多项式一致逼近。

根据二项级数展开

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$$

当 $\alpha > 0$ 时，上述级数在闭区间 $[-1, 1]$ 上一致收敛于函数 $(1+x)^\alpha$ 。

于是在闭区间 $[0, 2]$ 上 \sqrt{x} 可以展开成一致收敛的幂级数

$$\sqrt{x} = [1 + (x-1)]^{\frac{1}{2}} = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} (x-1)^n$$

这样我们证明了：

引理 1：连续函数 $\varphi(x) = \sqrt{x}$ 在闭区间 $[0, 1]$ 上可以用多项式一致逼近。

引理 2：在任意闭区间 $[-c, c]$ 上，连续函数 $\psi(x) = |x|$ 可以用多项式一致逼近。

证明：设多项式序列 $\{p_n(x)\}$ 在闭区间 $[0, 1]$ 上一致收敛于函数 $\varphi(x) = \sqrt{x}$ 。注意到等式关系 $\sqrt{x^2} = |x|$ ，于是我们记

$$q_n(x) = cp_n\left(\frac{x^2}{c^2}\right)$$

则多项式序列 $\{q_n(x)\}$ 在闭区间 $[0, 1]$ 上一致收敛于函数

$$c\varphi\left(\frac{x^2}{c^2}\right) = c\sqrt{\frac{x^2}{c^2}} = |x|.$$

定义折线函数

$$\lambda(x) = \frac{|x|+x}{2} = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$$

由引理 2 可知, 连续函数 $\lambda(x)$ 可在任意任意闭区间 $[-c, c]$ 上用多项式一致逼近。

上面定义的折线函数 $\lambda(x)$ 是最简单和基础的折线函数, 任意的折线函数都可以用它们的线性组合来表示。

引理 3: 定义在闭区间 $[a, b]$ 上的任意折线函数 $\Lambda(x)$ 都可以表示成

$$\Lambda(x) = c + c_0\lambda(x-x_0) + \cdots + c_m\lambda(x-x_m)$$

因而定义在闭区间上的任意折线函数都可以用多项式一致逼近。

证明: 设折线函数 $\Lambda(x)$ 的转折点为

$$x_0, x_1, \cdots, x_m, x_{m+1}$$

这里

$$a = x_0 < x_1 < \cdots < x_m < x_{m+1} = b$$

为了叙述方便, 我们把闭区间 $[a, b]$ 的两个端点 $a = x_0$ 和 $b = x_{m+1}$ 也都看作转折点。对于任意一组实数 c, c_0, c_1, \cdots, c_m , 显然

$$c + c_0\lambda(x-x_0) + \cdots + c_m\lambda(x-x_m)$$

也是以 $x_0, x_1, \cdots, x_m, x_{m+1}$ 为转折点的折线函数。我们知道, 任意折线函数由它在转折点的函数值完全确定。为了使

$$c + c_0\lambda(x-x_0) + \cdots + c_m\lambda(x-x_m) = \Lambda(x), \forall x \in [a, b]$$

只需取 c, c_0, c_1, \cdots, c_m 满足以下条件下列线性方程组:

$$\begin{cases} c = \Lambda(x_0) \\ c + c_0(x_1 - x_0) = \Lambda(x_1) \\ \cdots \cdots \\ c + c_0(x_m - x_0) + \cdots + c_{m-1}(x_m - x_{m-1}) = \Lambda(x_m) \\ c + c_0(x_{m+1} - x_0) + \cdots + c_{m-1}(x_{m+1} - x_m) = \Lambda(x_{m+1}) \end{cases}$$

即

$$c + \sum_{i=0}^m c_i \lambda(x_k - x_i) = \Lambda(x_k)$$

其中, $k=0, 1, \dots, m+1$.

以上线性方程组可以惟一确定系数 c, c_0, c_1, \dots, c_m , 引理 3 成立。

我们只须证明任意连续函数可由折线函数一致逼近, 由此可以推出魏尔斯特拉斯定理:

闭区间 $[a, b]$ 上的连续函数可由多项式一致逼近。

定理证明: 因为 $f(x)$ 在闭区间 $[a, b]$ 上一致连续, 所以 $\forall \varepsilon > 0, \exists \delta > 0$, 只要 $x, y \in [a, b], |x - y| < \delta$, 就有

$$|f(x) - f(y)| < \varepsilon$$

我们用分点把闭区间 $[a, b]$ 分成 $m+1$ 段,

$$a = x_0 < x_1 < \dots < x_m < x_{m+1} = b$$

使得

$$x_{k+1} - x_k < \delta, k = 0, 1, \dots, m$$

然后定义折线函数 $\Lambda(x)$ 如下:

$$\Lambda(x) = f(x_k) + \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}(x - x_k)$$

其中, $x \in [x_k, x_{k+1}], k = 0, 1, \dots, m$.

下面证明

$$|f(x) - \Lambda(x)| < \varepsilon, \forall x \in [a, b]$$

为此, 我们记

$$\alpha(x) = \frac{x_{k+1} - x}{x_{k+1} - x_k}, \beta(x) = \frac{x - x_k}{x_{k+1} - x_k}$$

在闭区间 $[x_k, x_{k+1}]$ 上显然有

$$\alpha(x) \geq 0, \beta(x) \geq 0, \alpha(x) + \beta(x) = 1$$

将 $\Lambda(x)$ 和 $f(x)$ 表示成

$$\Lambda(x) = \alpha(x)f(x_k) + \beta(x)f(x_{k+1})$$

$$f(x) = \alpha(x)f(x) + \beta(x)f(x)$$

于是, 在闭区间 $[x_k, x_{k+1}]$ 上, 有

$$|f(x) - \Lambda(x)| \leq \alpha(x)|f(x) - f(x_k)| + \beta(x)|f(x) - f(x_{k+1})| < \varepsilon$$

我们实际上在每一个小闭区间上证明了上述不等式, 所以证明了

$$|f(x) - \Lambda(x)| < \varepsilon, \forall x \in [a, b]$$

而前面已经证明了折线函数 $\Lambda(x)$ 可由多项式一致逼近, 于是定理成立。

评注：上述证明思路可概括为**多项式**→**折线函数**→**连续函数**，非常自然，以至于没有什么可点评的了，个人觉得这个证明比伯恩斯坦多项式的证明还要简单和自然。这个例子是分析中逼近思想的典型代表，充分体现了**用简单的逼近复杂的，用熟悉的逼近陌生的，这种思想**。

布劳威尔不动点定理

本文中，我们给出明布劳威尔不动点定理一个非常经典的解析证明，这个证明引自《数学分析新讲》。

布劳威尔不动点定理：

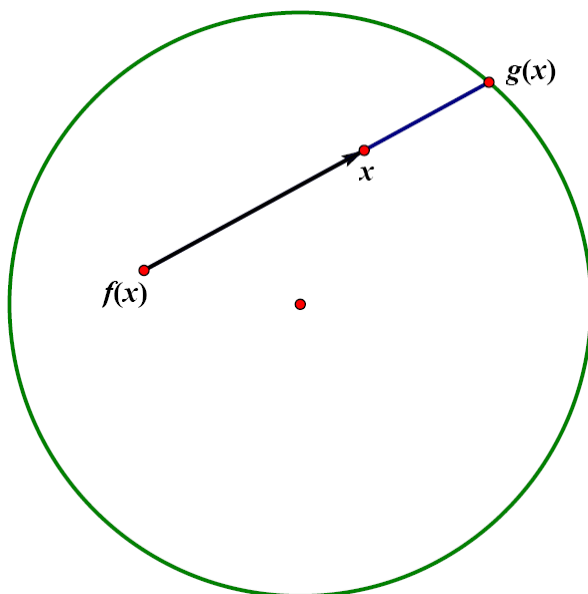
设 D^n 是 n 维单位闭球， $f: D^n \rightarrow D^n$ 是连续映射，则存在 $x_0 \in D^n$ ，满足 $f(x_0) = x_0$ 。

在布劳威尔不动点定理的条件中，单位闭球

$$D^n = \{(x_1, \dots, x_n) \in \mathbf{R}^n : x_1^2 + \dots + x_n^2 \leq 1\}$$

的凸性非常重要。比如把 D^n 换成单位球面 S^n ，则结论不再成立。反例是映射 $f(x) = -x$ ，显然映射 $f: S^n \rightarrow S^n$ 是连续映射，但没有不动点。

《新讲》给出的布劳威尔不动点定理的解析证明核心思想在于下面这张图



假设布劳威尔不动点定理不成立，那么可以做出一个映射 $g: D^n \rightarrow S^{n-1}$ 满足 $g(x) = x$ ， $x \in S^{n-1}$ 。这样的连续映射 g 称为收缩映射， S^{n-1} 称为 D^n 的收缩核。但是，可以证明这样的收缩核实际上不存在，从而不动点定理为真。

为利用分析工具，先假设 f 为光滑映射。《新讲》中首先以 $n=2$ 的情况为例，对于一般的 n 证明没有本质区别。

引理：记圆盘 $D = \{(x_1, x_2) \mid x_1^2 + x_2^2 \leq 1\}$ ，其边界为圆周 S ，则不存在光滑

映射 $g: D \rightarrow S$, 满足 $g(x) = x, x \in S$.

引理证明: 考虑映射

$$g = (g_1, g_2): D \rightarrow \mathbf{R}^2$$

以及微分形式 $\omega = g_1 dg_2$ 的积分

$$\int_{\partial D} \omega$$

一方面代入 $g = (g_1, g_2)$, 由 $g_1^2 + g_2^2 = 1$ 推出雅可比行列式

$$\frac{\partial(g_1, g_2)}{\partial(x_1, x_2)} = 0$$

代入斯托克斯公式, 得到 (此处 $S = \partial D$)

$$\int_{\partial D} \omega = \int_D d\omega = \iint_D \frac{\partial(g_1, g_2)}{\partial(x_1, x_2)} dx_1 \wedge dx_2 = 0$$

另一方面, 根据 $g(x) = x, x \in S$, 得

$$\int_{\partial D} \omega = \int_{\partial D} x_1 dx_2 = \iint_D dx_1 \wedge dx_2 = \pi > 0$$

这就产生了矛盾! 于是光滑映射 $g = (g_1, g_2)$ 不存在。

上述证明方法的思想是计算两次, 即对同一个对象用两种不同的方法计算, 结果应该是一样的。否则, 即产生矛盾。一般的情形证明没有本质区别。

设 $g = (g_1, g_2, \dots, g_n)$, 构造微分形式 $\omega = g_1 dg_2 \wedge \dots \wedge dg_n$.

一方面, $g(x) = x \in S^{n-1} = \partial D^n$, 所以 $\|g(x)\| = 1, \forall x \in S^{n-1}$, 从而

$$\begin{aligned} \int_{\partial D^n} \omega &= \int_{D^n} d\omega = \int_{D^n} d(g_1 dg_2 \wedge \dots \wedge dg_n) \\ &= \int_{D^n} \frac{\partial(g_1, \dots, g_n)}{\partial(x_1, \dots, x_n)} dx_1 \wedge \dots \wedge dx_n = 0 \end{aligned}$$

另一方面, 由 $g(x) = x, x \in S^{n-1}$, 直接代入, 得到

$$\int_{\partial D^n} \omega = \int_{D^n} d\omega = \int_{D^n} dx_1 \wedge \dots \wedge dx_n = \text{vol}(D^n) > 0$$

这就证明了不存在光滑的收缩映射 $g: D^n \rightarrow S^{n-1}$.

如果布劳威尔不动点定理不成立, 即 $f(x) \neq x$ 恒成立, 则可以做出一条由 $f(x)$ 出发到 x 的射线 (见上图), 交边界 S^{n-1} 于 $g(x)$, 这样定义的映射

$$g: D^n \rightarrow S^{n-1}$$

是光滑映射且满足 $g(x) = x$, $x \in S^{n-1}$, 这与上面的引理矛盾。于是证明了当 f 是光滑映射时, 存在不动点。

对于连续映射情况, 可以选择一个光滑映射序列来一致逼近。假设连续映射 $f: D^n \rightarrow D^n$ 没有不动点。

因为单位闭球是有界闭集 (紧致性), 所以连续函数 $\|f(x) - x\|$ 在 D^n 上一定取得正的最小值 $\mu = \min \|f(x) - x\| > 0$. 取定 $\varepsilon > 0$ 满足

$$0 < 3\varepsilon < \mu = \min \|f(x) - x\|$$

根据魏尔斯特拉斯逼近定理, 存在多项式向量函数 $p(x)$, 使得

$$\|f(x) - p(x)\| < \varepsilon$$

不过对于 $x \in D^n$ 可能有 $p(x) \notin D^n$ (映射到单位闭球外面)。但可以判定

$$\|p(x)\| \leq \|p(x) - f(x)\| + \|f(x)\| < 1 + \varepsilon$$

于是令

$$h(x) = \frac{1}{1+\varepsilon} p(x)$$

则得到光滑映射 $h: D^n \rightarrow D^n$.

再次利用含绝对值不等式, 得到

$$\begin{aligned} \|h(x) - f(x)\| &= \left\| \frac{1}{1+\varepsilon} p(x) - f(x) \right\| = \frac{1}{1+\varepsilon} \|p(x) - f(x) - \varepsilon f(x)\| \\ &\leq \frac{1}{1+\varepsilon} \|p(x) - f(x)\| + \frac{\varepsilon}{1+\varepsilon} \|f(x)\| < \frac{2\varepsilon}{1+\varepsilon} < 2\varepsilon \end{aligned}$$

由此得到估计

$$\|h(x) - x\| \geq \|x - f(x)\| - \|h(x) - f(x)\| \geq 3\varepsilon - 2\varepsilon = \varepsilon > 0$$

但是根据前面的结论, 光滑映射 h 具有不动点 $h(x) = x$, 矛盾! 定理成立。

评注: 上面的证明集变换与逼近思想于一体, 非常巧妙。光滑情形逼近连续情形也可以直接证明: 首先可以选择一个光滑映射序列 $g_n(x)$ 来一致逼近 $f(x)$ 。

记 $g_n(x_n) = x_n$, 因为单位闭球是有界闭集, 所以可以选出一个收敛子列 $x_{n_k} \rightarrow x_0 \in D^n$, 故 $g_{n_k}(x_{n_k}) \rightarrow x_0$ 。

再利用 f 连续性与 g_{n_k} 的一致收敛性, 可知

$$|f(x_0) - g_{n_k}(x_k)| \leq |f(x_0) - f(x_{n_k})| + |f(x_{n_k}) - g_{n_k}(x_{n_k})| \rightarrow 0$$

如此便证明了 $f(x_0) = x_0$ 成立。

《新讲》的证明是非常严格的，必须证明利用射线做出的映射 $g: D^n \rightarrow S^{n-1}$ 确实是光滑映射。映射 g 可以表示成

$$g(x) = f(x) + t(x - f(x))$$

其中 $t = t(x)$ 使得 $\|g(x)\|^2 = 1$ ，即使得

$$a(x)t^2 + 2b(x)t + c(x) = 0$$

这里

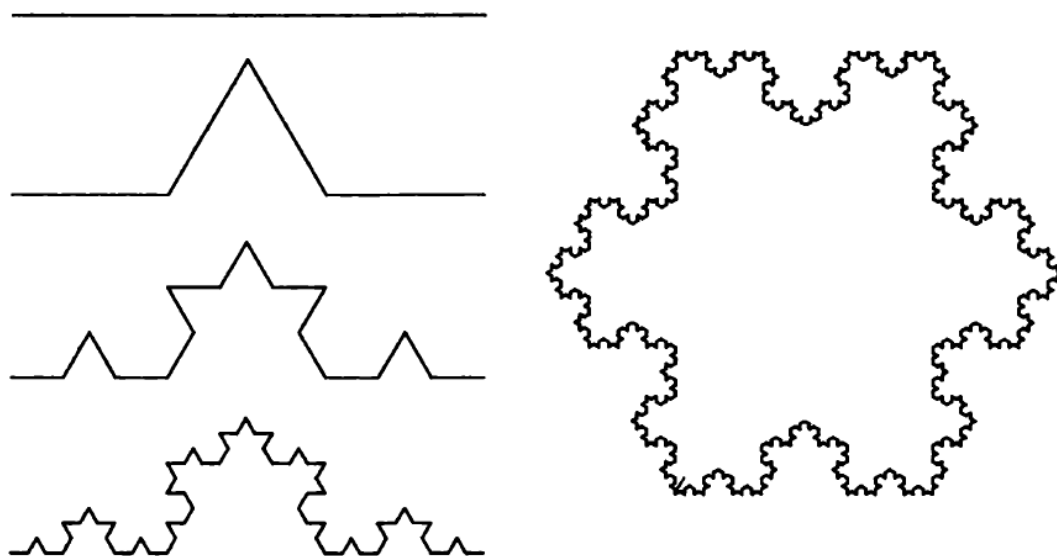
$$a(x) = \|x - f(x)\|^2 > 0, \quad b(x) = (x - f(x)) \cdot f(x) \quad (\text{内积})$$

$$c(x) = \|f(x)\|^2 - 1 \leq 0$$

由此出发可以证明，对任意的 $x \in D^n$ ，上面方程有惟一的正根 $t = t(x)$ 。再由二次方程的求根公式表达式可知 $t = t(x)$ 是光滑函数。

处处连续处处不可导函数

处处连续处处不可导函数的存在相当出乎意料，至少笔者当年听闻时感觉如此。不过随着知识的加深，才发现自己高中时代就见过一种处处没有切线的曲线，这就是分形几何之雪花曲线：

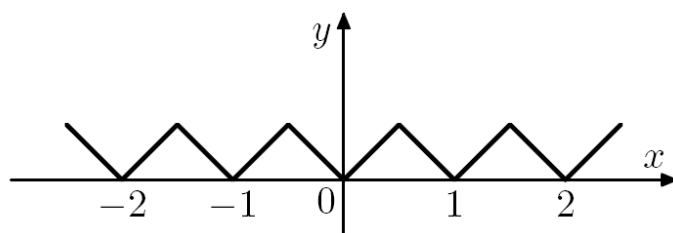


雪花曲线（也称科赫曲线）

上面这种曲线的极限情况就是处处没有切线的曲线。后来笔者在自学《数学分析新讲》时，第一次见到了处处连续处处不可导函数的表达式，以及严格的数学证明。书中采用锯齿函数叠加的构造式证明，很直观。

首先，定义以 1 为周期的函数

$$u_0(x) = |x|, \quad x \in \left[-\frac{1}{2}, \frac{1}{2}\right]$$



显然， $u_0(x+1) = u_0(x)$ ， $0 \leq u_0(x) \leq \frac{1}{2}$ 。

构造函数序列

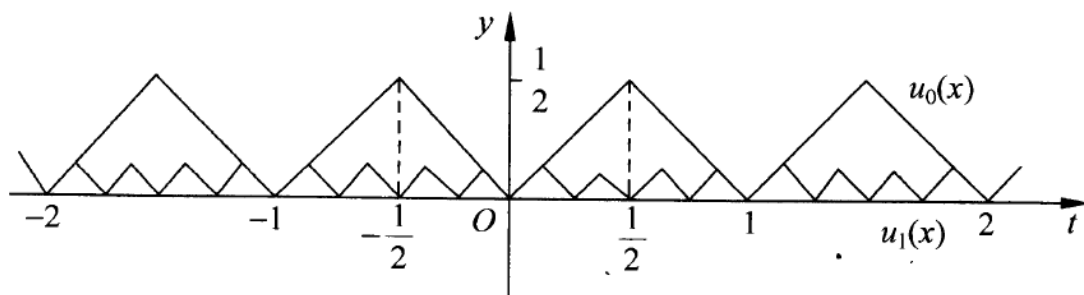
$$u_k(x) = \frac{1}{4^k} u_0(4^k x), \quad k = 1, 2, \dots$$

不难证明

$$u_k(x + \frac{1}{4^k}) = u_k(x)$$

所以 $u_k(x)$ 是以 $1/4^k$ 为周期的连续函数, 且满足

$$0 \leq u_k(x) \leq \frac{1}{2} \cdot \frac{1}{4^k}, \quad \forall x \in \mathbf{R}$$



现 $\forall m \in \mathbf{Z}$, 因为 $u_0(x)$ 在 $[\frac{m}{2}, \frac{m+1}{2}]$ 上是线性的, 所以 $u_k(x)$ 在 $[\frac{m}{2 \cdot 4^k}, \frac{m+1}{2 \cdot 4^k}]$ 上也是线性的。这些 $u_k(x)$ 都是锯齿函数, 且随着 k 的增大, 锯齿越来越密。

了解一些分形几何知识的读者可以发现, 这种自相似图形已经有了分形几何的味道了。把这些函数叠加到一起, 得到函数级数

$$f(x) = \sum_{k=0}^{\infty} u_k(x), \quad x \in \mathbf{R}$$

上面的级数具有优级数

$$\sum_{k=0}^{\infty} \frac{1}{2} \cdot \frac{1}{4^k} < +\infty$$

所以 $f(x)$ 在 \mathbf{R} 上一致收敛。又 $u_k(x)$ 处处连续, 故 $f(x)$ 在 \mathbf{R} 上处处连续。

下面证明 $f(x)$ 处处不可导。 $\forall c \in \mathbf{R}$. 对任意的自然数 n , 存在整数 m_n 使得

$$m_n \leq 2 \cdot 4^{n-1} c < m_n + 1$$

也就是

$$\frac{m_n}{2 \cdot 4^{n-1}} \leq c < \frac{m_n + 1}{2 \cdot 4^{n-1}}$$

对任意的自然数 n , 在区间

$$I_n = [\frac{m_n}{2 \cdot 4^{n-1}}, \frac{m_n + 1}{2 \cdot 4^{n-1}})$$

内可以选取一点 x_n , 使得

$$|x_n - c| = \frac{1}{4^n}$$

显然, $\lim x_n = c$.

考察差商

$$\frac{f(x_n) - f(c)}{x_n - c} = \sum_{k=0}^{\infty} \frac{u_k(x_n) - u_k(c)}{x_n - c}$$

当 $k \geq n$ 时, 函数 $u_k(x)$ 的周期 $\frac{1}{4^k}$ 能够整除 $|x_n - c| = \frac{1}{4^n}$, 因而

$$\frac{u_k(x_n) - u_k(c)}{x_n - c} = 0$$

当 $k \leq n-1$ 时, $I_n \subset I_k$ 且函数 $u_k(x)$ 在 I_k 上斜率等于 ± 1 . 根据记号约定

$$I_k = [\frac{m_k}{2 \cdot 4^{k-1}}, \frac{m_k + 1}{2 \cdot 4^{k-1}})$$

当 m_k 是奇数的时候, $u_k(x)$ 在 I_k 上斜率为 -1 ; 当 m_k 是偶数的时候, $u_k(x)$ 在 I_k 上斜率为 1 . 于是

$$\frac{f(x_n) - f(c)}{x_n - c} = \sum_{k=0}^{\infty} \frac{u_k(x_n) - u_k(c)}{x_n - c} = \sum_{k=0}^{n-1} \frac{u_k(x_n) - u_k(c)}{x_n - c} = \sum_{k=0}^{n-1} (\pm 1)$$

上式右端与 n 有同样的奇偶性, 从而当 $n \rightarrow \infty$ 时极限不存在. 由 $c \in \mathbf{R}$ 取法的任意性可知 $f(x)$ 处处连续处处不可导.

上述证明称作**构造性证明**, 属于直接法, 而数学上还有**存在性证明**, 属于间接法, 而且结论更惊人---处处连续处处不可导函数比可导函数多得多, 这个结论好比超越数比代数数多得多这个结论, 非常震撼!

存在性证明依赖贝尔的“纲定理”, 为此先介绍一下疏集概念. 我们讨论的空间默认为度量空间 X .

设 $S \subset X$, 其闭包记为 \bar{S} , 如果 \bar{S} 的内部是空集, 则称 S 为**疏集**. 若每一个 S_n 都是疏集, 则称

$$E = \bigcup_{n=1}^{\infty} S_n$$

为**第一纲点集合**. 也就是说, 可数个疏集的并集为第一纲点集合. 非第一纲点集合称为**第二纲点集合**.

(纲定理) 完备的度量空间必是第二纲的.

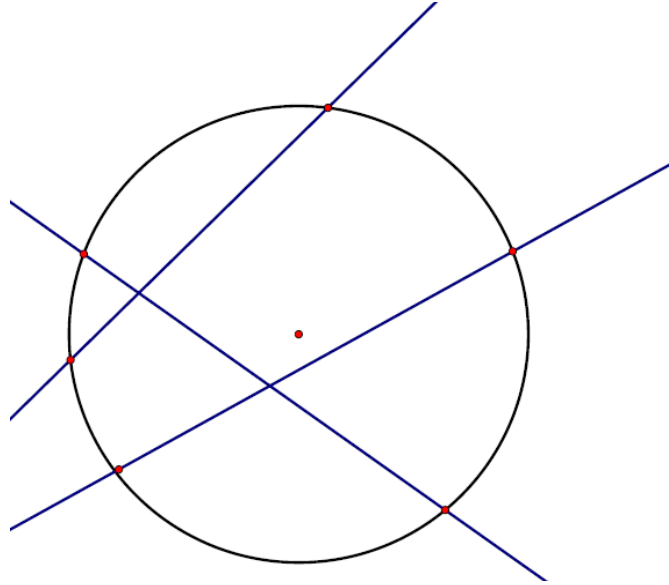
定理的证明采用反证法, 假设 X 是第一纲的. 由于 S_n 是疏集, 故 \bar{S}_n 不能包含任何开球. 然后利用疏集定义构造一个闭球套序列, 再在闭球套序列中取柯西列, 使得其极限点不属于任何 S_n , 这与度量空间 X 的完备性矛盾. 纲定理的

证明可参考任何一本《泛函分析》教材，这里从略。在理解纲定理的证明时，必须深刻领悟疏集和可数这两个条件的作用。

为了加深对纲定理的印象，考虑一个具有几何直观的特例：设完备度量空间 X 为平面， S_n 是直线，则不能用可数条直线覆盖住整个平面。显然直线 S_n 是疏集，假设可以用可数条直线覆盖住整个平面，即

$$X = \bigcup_{n=1}^{\infty} S_n$$

考虑一个单位圆 Γ 。因为任意直线 S_n 至多与单位圆 Γ 有两个交点，所以可数条直线至多与单位圆有可数个交点。



但是单位圆上有不可数个点，于是必有不在 X 中的点，矛盾！

现在可以开始存在性证明了。首先构造完备的度量空间 X ，记区间 $[a, b]$ 上所有连续函数之集合为 $C[a, b]$ ，引入 **max** 度量

$$\|f\| = \max_{a \leq x \leq b} |f(x)|$$

可以证明 $C[a, b]$ 在 **max** 度量下构成一个完备的度量空间，记为 X 。

构造集合

$$S_n = \left\{ f \in X : \exists x \in [a, b], \left| \frac{f(x+h) - f(x)}{h} \right| \leq n, \forall h > 0 \right\}$$

显然，如果 $\varphi \in X$ 至少在一点可导，则 φ 必属于某个 S_n 。

如果我们能证明每个 S_n 都是疏集，则根据纲定理，必存在

$$f \in X - \bigcup_{n=1}^{\infty} S_n$$

这就证明了处处连续处处不可导函数的存在性。

证明分两步：

1. S_n 是 X 中的闭集 (即 $\bar{S}_n = S_n$)

因为 X 是完备的, 所以可以取一个一致收敛列 $f_k \in S_n$, 其中

$$f_k \rightarrow f$$

为证明 S_n 是 X 中的闭集, 只需证明极限 $f \in S_n$ 即可。

根据 S_n 的定义, 存在 $x_k \in [a, b]$ 使得

$$\left| \frac{f_k(x_k + h) - f_k(x_k)}{h} \right| \leq n, \forall h > 0$$

注意到 $\{x_k\}$ 是有界数列, 因此存在收敛子列。不妨设 $x_k \rightarrow x_0$, 则 $x_0 \in [a, b]$ 。

对任意 $h \geq 0$, 有

$$|f_k(x_k + h) - f(x_0 + h)| \leq |f_k(x_k + h) - f(x_k + h)| + |f(x_k + h) - f(x_0 + h)|$$

由 f_k 的一致收敛性, 可知

$$f_k(x_k + h) - f(x_k + h) \rightarrow 0$$

由 f 的连续性, 可知

$$f(x_k + h) - f(x_0 + h) \rightarrow 0$$

于是, 得到

$$f_k(x_k + h) \rightarrow f(x_0 + h) \text{ 且 } f_k(x_k) \rightarrow f(x_0)$$

从而当 $k \rightarrow \infty$ 时,

$$\left| \frac{f(x_0 + h) - f(x_0)}{h} \right| \leq n, \forall h > 0$$

这说明 $f \in S_n$, 故 S_n 是 X 中的闭集。

2. S_n 是疏集。否则假设 S_n 包含某个开球

$$K = \{g \in X : \|g - f\| < \varepsilon\}$$

但是, 我们可以找到一个折线函数 $\varphi(x)$, 满足:

$$\|\varphi - f\| = \max_{a \leq x \leq b} |\varphi(x) - f(x)| < \varepsilon$$

并且折线 φ 的每段斜率绝对值都大于 n , 这说明 $\varphi \notin S_n$, 与 $\varphi \in K$ 矛盾!

综上所述, 可知 $\bigcup_{n=1}^{\infty} S_n$ 是第一纲的。因为 X 是完备的, 根据纲定理, 必有连续函数 ϕ , 不在任何 S_n 中, 这个连续函数 ϕ 处处连续处处不可导。

评注：上述存在性证明还是很巧妙的，其中 S_n 的构造是关键，可以这样想：处处连续处处不可导的否定形式是至少在一点可导，而导数定义为差商的极限，所以考虑差商情况自然包括了可导函数的情况（比如拉格朗日中值定理）。

直观上，任取正数 λ ，存在连续函数（如折线函数），其某点斜率的绝对值大于 λ 。所有连续函数斜率的绝对值范围可以是 $[0, +\infty)$ ，差商也是如此（如折线函数）。所以用自然数集合分割差商斜率绝对值取值范围是一种常见方法

$$\exists x \in [a, b], \left| \frac{f(x+h) - f(x)}{h} \right| \leq n, \forall h > 0$$

而且这种分割是可数的，这有点像 Lebesgue 积分定义中对 y 轴的分割。

用了构造 S_n 的想法之后，接下来就自然了。首先利用一致收敛性和连续性可知 S_n 是闭集。为了证明 S_n 是疏集，只需证明它不能包含任意开球，这不难做到，只需要指出存在每段斜率绝对值都大于 n 的函数即可。

请允许笔者用模糊的语言形容下，对处处连续处处不可导函数（也称 BT 函数）构造性和存在性证明的感受。关键的地方在于差商，导数可以看成是差商收敛的情况（当两点无限接近时）。直观上，若两点无限接近时，函数剧烈震荡，差商振幅不断上下震荡增大，则不能期望其收敛，这一点正是这些 BT 函数存在的关键。为了理解笔者这些模糊的描述，推荐大家读读分形几何方面的文章，只要你看看分形几何的图形，应该能明白笔者的意思了。

再说说疏集。直观上，疏集因为没有内部（比如单点集合），所以给人的感觉非常小，而完备的度量空间 X 通常比较大（在完备意义下），所以第一纲感觉比第二纲少多了（如上面的直线与平面例子）。第一纲的概念很像可数个零测集的并集还是零测集一样。当然，“多少”只是模糊的说法，无穷不像有限那样直观，反而很神奇。比如存在不可数的零测集，如康托三分集。因而一个 Riemann 可积函数可能有不可数个间断点。

第一纲概念的关键在于引入一种分类标准。像第一纲集、零测集、稠密集等，都是根据某类具体问题而引入的。用分析打个比方，一个数列的算术平均和收敛不能确保数列本身收敛，它本身可以发散。但是，如果我们研究的问题中间某个步骤只涉及到算术平均和，那么就可以适当扩大考察对象范围，将某些发散数列（只要算术平均和收敛即可）也考虑进来。

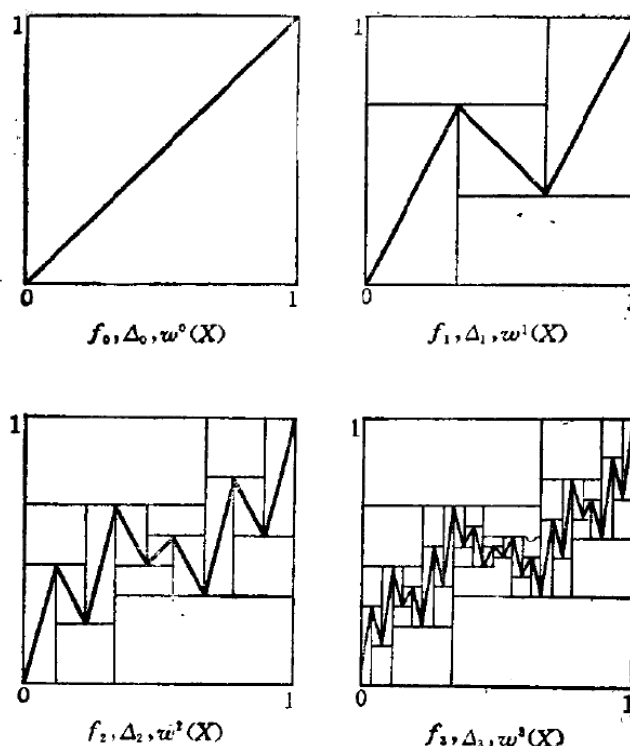
处处连续处处不可导函数反例最早由魏尔斯特拉斯给出，例子如下

$$f(x) = \sum_{k=0}^{\infty} b^k \cos(a^k \pi x), \quad x \in \mathbf{R}$$

其中 $0 < a < 1, ab \geq 1$. 这个定理的详细证明请参考齐民友的《重温微积分》。正文中那个例子由范德瓦尔登给出，想法是把三角函数换成更简单的锯齿函数，这是一种简化问题的例子。

正文的例子尽管每一项 $u_k(x)$ 图像都很直观（锯齿），但是叠加到一起就不那么直观了，你可以想象一堆周期和振幅各不相同的锯齿函数叠加到一起有多么曲折，但很难直观化。

出于直观化考虑，邹应在其《数学分析》一书中收录了美国数学月刊中的一个处处连续处处不可导函数反例。当然，例子的证明仍然很复杂，但函数图形却非常直观（如下图），有兴趣的读者可以去看一看。



误差与正态分布

SCIbird

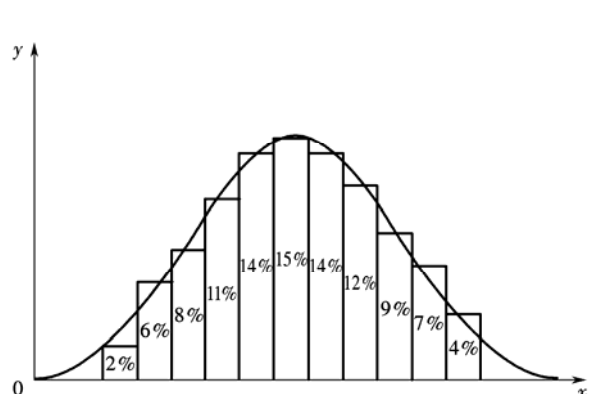
说明：本文介绍了高斯在正态分布方面的杰出贡献，叙述问题时采用现代符号和说法。根据数学史资料，高斯是在研究误差过程中导出正态分布的。本文写作主要参考了陈希孺院士所著的《数理统计学简史》一书。

先看一张图，这是一张德国马克钞票，头像对应人物就是数学大师高斯。



德国马克

高斯头像左侧对应那条曲线就是概率论和数理统计中大名鼎鼎的正态分布曲线。



正态分布曲线

定性上看，正态分布曲线形态如上图所示，曲线是轴对称的，中间高，而在两侧迅速递减。用数学语言来描述正态分布，正态分布曲线表达式（概率密度函数）：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ 其中 } -\infty < x < +\infty, \sigma > 0$$

其分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \text{ 其中 } -\infty < x < +\infty$$

这种分布称为正态分布，简记为 $N(\mu, \sigma^2)$ 。特别的，当 $\mu=0, \sigma=1$ 时，对应的正态分布 $N(0, 1)$ 称为标准正态分布。

从实际应用角度看，正态分布可能是最重要的概率分布了！不论在数理统计领域、自然科学和工程技术，还是经济和金融领域等等，到处能看到正态分布的身影。管理领域中的所谓“六西格玛 (6σ)”概念，最初出自正态分布，大意是说每一百万次产品采样数，抽样合格率为 99.99966%。不过经过这些年的发展，“六西格玛”管理已经发展成一套系统的程序（尽管不同公司可能不一样），早已经不单单局限在统计数字了。

正态分布重要的原因还体现在它是某些常见分布的极限分布，其体现形式是极限定理。

（隶莫弗-拉普拉斯）若 X 是 n 次伯努力实验中事件 A 出现的次数，即满足 $P\{X=k\} = C_n^k p^k q^{n-k}$, $0 < p < 1, q = 1-p$, 则对任意有限区间 $[a, b]$, 当 $n \rightarrow \infty$ 时，一致地有

$$\lim_{n \rightarrow \infty} P\left\{a < \frac{X - np}{\sqrt{npq}} \leq b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

该定理给出了当 n 较大时，如何近似计算二项分布。

注：伯努利试验：是在同样的条件下重复地、各次之间相互独立地进行的一种试验。这种试验中，每一次试验只有两种结果，即某事件 A 要么发生，要么不发生。并且每次发生的概率都是相同的。判断是否为独立重复试验的关键是每次试验事件 A 发生的概率不变，并且每次试验的结果同其他各次试验的结果无关，重复是指试验为一系列的试验，并非一次试验，而是多次，但要注意重复事件发生的概率相互之间没有影响。

比上面极限定理更深刻的是下面的林德伯格-列维中心极限定理：

（林德伯格-列维）设随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 是独立同分布的，期望 $EX_1 = \mu$ ，方差 $DX_1 = \sigma^2 > 0$ ，记 $S_n = X_1 + X_2 + \dots + X_n$ ，则对任意 x 都有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

中心极限定理还有其它形式，请读者参考相关教材。

从中心极限定理可以粗略地概括出：

在实际问题中，如果有大量独立的随机变量，而且每个随机变量相对其和的影响都比较小，则这些随机变量之和就“近似于”正态分布。之所以说是近似是因为正态分布是极限分布，而在实际问题中是无法达到数学中极限状态的。

注：以上方框中的那段话非常重要，它指出了正态分布应用的前提或界限，而当下社会，正态分布明显被滥用了。

上面的观点很重要的，因为下面提及高斯从研究误差导出正态分布的过程，而误差恰恰具有上段提到的那些性质（见方框内的文字）。一般来说，随机测量误差是出自大量不显著的来源的叠加，这一点对高斯的误差理论很重要——因为在高斯的推导中用到了这样一个假设，即“平均数是自然合理的（相当于默认这是一公理）”。可能是因为高斯把正态分布与误差研究紧密结合起来的缘故吧，后世人也称正态分布为高斯分布（尽管严格说来，隶莫弗和拉普拉斯之前也发现了正态分布）。

下面我们来说，高斯是如何从研究误差导出正态分布的。据史料分析，高斯应该是在研究天文学测量误差过程中发现了正态分布的。一个重要问题：对同一对象进行反复测量，随机测量误差服从怎样的概率分布？

我们用现代的说法来描述下高斯的思考过程：设随机误差的概率密度函数为 $f(x)$ ，真值为 θ ， n 个独立的测量值为 x_1, x_2, \dots, x_n ，高斯把后者的概率取为

$$L(\theta) = L(\theta; x_1, x_2, \dots, x_n) = \prod_{k=1}^n f(x_k - \theta)$$

高斯考虑把满足 $L(\hat{\theta}) = \max L(\theta)$ 的 $\hat{\theta}$ 作为 θ 的估计值，按现代说法 $L(\theta)$ 叫做似然函数，上面的这种方法叫做最大似然估计法。

接下来高斯假设（实际上相当于默认为“公理”）算术平均值

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$

即为应取的估计 $\hat{\theta} = \bar{x}$ ，然后设法寻找误差概率密度函数 $f(x)$ ，满足

$$L(\bar{x}) = \max L(\theta).$$

由对数性质得，

$$\ln L(\theta) = \sum_{k=1}^n \ln f(x_k - \theta)$$

对 θ 求导，得

$$\frac{d \ln L(\theta)}{d\theta} = - \sum_{k=1}^n \frac{f'(x_k - \theta)}{f(x_k - \theta)}$$

记 $g(x) = f'(x)/f(x)$ ，因为 $L(\bar{x}) = \max L(\theta)$ ，所以

$$\sum_{k=1}^n g(x_k - \bar{x}) = 0$$

我们来归纳考察下函数 $g(x)$ 具有哪些性质，先看 $n=2$ 情形，

令 $x_1 - \bar{x} = -(x_2 - \bar{x}) = x$ ，则 $g(x) + g(-x) = 0$ ，即

$$g(-x) = -g(x), g(0) = 0$$

对一般的正整数 n ，令 $n = m+1$ ，取 $x_1 = x_2 = \cdots = x_m = -x, x_n = mx$ ，则 $\bar{x} = 0$ ，

$$mg(-x) + g(mx) = 0$$

再利用 $g(-x) = -g(x)$ ，得到

$$g(mx) = mg(x)$$

我们对一切正整数 m 和实数 x ，推导出性质 $g(mx) = mg(x)$ ，进而对一切有理数 r ，有 $g(rx) = rg(x)$ 。这是因为有理数 r 可以表示成两整数之比 m/n ($m > 0$)，及 $mg(x) = g(mx) = g(nmx/n) = ng(mx/n) = ng(rx)$

取 $x=1$ ，则 $g(r) = g(1)r = cr$ 对一切有理数 r 成立。我们大概已经猜出来 $g(x) = cx$ 是线性函数这一结果了。

事实上，我们假定函数 $g(x)$ 连续，再结合有理数在实数中的稠密性（有理数可任意逼近实数），则 $g(x) = cx$ 对一切实数 x 成立。

注：上面求解函数 $g(x)$ 的过程是函数方程理论的经典解法（参考柯西爬坡法），问题相当于对任意正整数 n 和实数 x_1, x_2, \cdots, x_n ，求解函数方程 $\sum g(x_k - \bar{x}) = 0$ 。

根据 $g(x) = f'(x)/f(x)$ 定义，两边直接积分得

$$f(x) = ke^{\frac{1}{2}cx^2}$$

按概率密度函数定义，应有 $\int_{-\infty}^{+\infty} f(x) dx = 1$ ，故 $c < 0$ ，记 $c = -1/h^2$ ，则

$$f(x) = \frac{1}{\sqrt{2\pi}h} e^{-\frac{x^2}{2h^2}}$$

这就是正态分布 $N(0, h^2)$!

于是可知观测值的误差概率密度函数 (x_1, x_2, \dots, x_n) 为

$$\frac{1}{(\sqrt{2\pi}h)^n} \exp \left\{ -\frac{1}{2h^2} \sum_{k=1}^n (x_k - \theta)^2 \right\}$$

若想使得上式取得最大值, 必须使得 $\sum (x_k - \theta)^2$ 取得最小值, 这与最小二乘法估计发生了联系, 同时给出了最小二乘法合理性的一种解释。以上就是高斯的数学工作了, 至于对后世的影响, 陈希孺院士所著的《数理统计学简史》一书中有更详细的评价。

用一个问题来说明最小二乘法数学思想:

设对某未知量 θ 重复测量 n 次, 测量结果分别为 x_1, x_2, \dots, x_n , 若想利用这些测量值对知量 θ 做一估计, 怎样估计为好?

设真值为 a , 则测量值 x_k 与 a 的误差记为 $\varepsilon_k := x_k - a, k=1, 2, \dots, n$. 如果 a 确实为真值, 测量值 x_k 应该分布在 a 的附近, 从而误差绝对值 $|\varepsilon_1|, |\varepsilon_2|, \dots, |\varepsilon_n|$ 应该比较小。如何从整体上考虑呢? 直观上考虑 $L = |\varepsilon_1| + |\varepsilon_2| + \dots + |\varepsilon_n|$ 似乎不错, 但去掉绝对值符号是个难以操作的问题。退一步的话, 考虑误差的平方和是个相对容易的方法, 令

$$L(a) = |\varepsilon_1|^2 + |\varepsilon_2|^2 + \dots + |\varepsilon_n|^2 = \sum_{k=1}^n \varepsilon_k^2 = \sum_{k=1}^n (x_k - a)^2$$

直观上, 若 a 为真值, 则 $L(a)$ 应该尽可能小。视 a 为变量 x , 我们猜想使得 $L(x)$ 最小的那个值 \bar{x} 为所求估计值, 即 $L(\bar{x}) = \min L(x)$, 求解 $L'(\bar{x}) = 0$, 得

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$

即估计值取测量值的算术平均值 \bar{x} 时, $L(\bar{x}) = \min L(x)$ 。实践表明, \bar{x} 确实是真值 a 的一个很好的近似值。

寻找估计值, 使得误差平方和达到最小, 这种方法叫做最小二乘法。当然, 在最小二乘法出现之前, 人们已经凭经验利用算术平均值了作为真值的估计值来解决实际问题了, 很多例子表明这种方法很实用。这反过来增加了最小二乘法合理的可能性。

最小二乘法一般形式是： $L = \sum (A_k - \vartheta)^2$ ，其中 L 为目标函数， A_k 是观测值， ϑ 为理论值。理论值根据设定的模型计算，其中含有未知参数，其值使得“目标函数 L 达到最小值”。求和运算 Σ 也可用积分运算 \int 代替。

评注：在这篇文章里，高斯运用**逆向思维**，不像通常那样（像我们做概率习题中遇到的）根据某某分布求未知参数。高斯的想法是假设参数已知，并且就是**算术平均值**，反过来求一种概率分布，使得其最大似然估计是算术平均值。经过推导，高斯求出了这种分布，即后来大名鼎鼎的正态分布。推导过程本身并不复杂，但能够进行逆向思维，却不一般。最后，高斯偏爱算术平均值作为未知概率分布的最大似然估计也不是偶然的，这与他本人对最小二乘法的研究密切相关。

我们计算一下正态积分以确定系数 k ，这里采用泊松的解法。

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = ?$$

设正方形区域 $D = \{(x, y) : |x| \leq R, |y| \leq R\}$ ，考虑过 D 的四个顶点的外接圆 S 以及 D 的内切圆 S' ，对应半径分别为 $\sqrt{2}R$ 和 R 。

考察累次积分（注意两个积分对称）

$$\left(\int_{-R}^R e^{-x^2} dx \right)^2 = \int_{-R}^R e^{-x^2} dx \int_{-R}^R e^{-y^2} dy = \iint_D e^{-(x^2+y^2)} dx dy$$

注意到区域关系 $S' \subset D \subset S$ ，我们来估算下面的积分

$$\iint_{S'} e^{-(x^2+y^2)} dx dy < \iint_D e^{-(x^2+y^2)} dx dy < \iint_S e^{-(x^2+y^2)} dx dy$$

不等式左边积分用极坐标表示是

$$\iint_{S'} e^{-(x^2+y^2)} dx dy = \int_0^{2\pi} d\varphi \int_0^R e^{-r^2} r dr = (1 - e^{-R^2})\pi$$

同理，可计算出右边积分 $(1 - e^{-2R^2})\pi$ ，于是我们得到了不等式估计

$$(1 - e^{-R^2})\pi < \left(\int_{-R}^R e^{-x^2} dx \right)^2 < (1 - e^{-2R^2})\pi$$

令 $R \rightarrow +\infty$ ，得到

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

或者写成更常见的等价表示是（正态积分）

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1$$

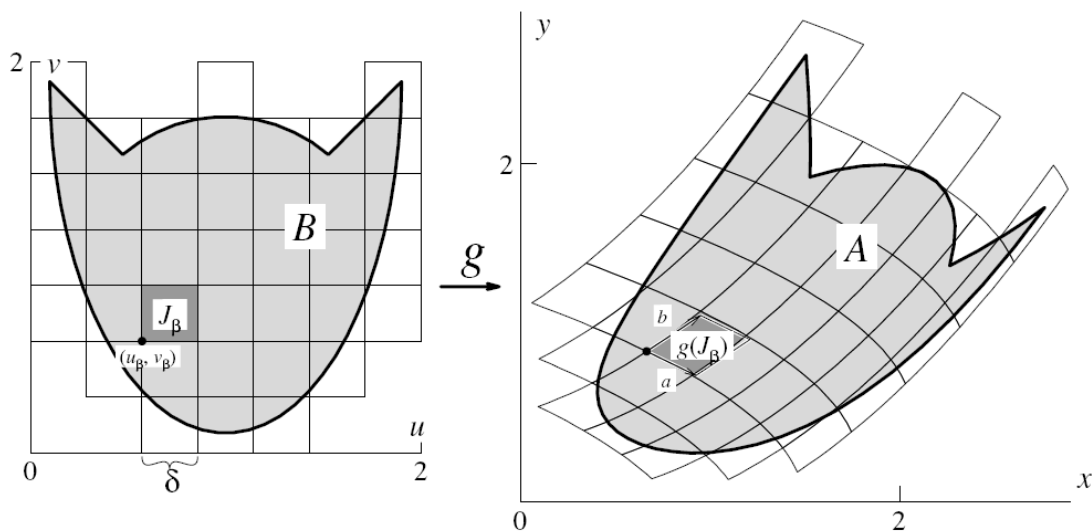
由此可知前面提及的高斯正态分布里面的常数 $k = \frac{1}{\sqrt{2\pi h}}$ 。

注：需要说明的是，正态分布虽然很有用，但时下有被滥用了，这一点在金融领域尤其明显。注意到正态曲线两侧下降速度特别快（你可能听说过某某金融危机发生的是“20 西格玛事件”，500 年难得一遇，你感到很无语），而金融领域相当多的事实与此不符，很多金融事件有着一一条“大尾巴”，所谓“肥尾现象”。不要小瞧“肥尾”，很多人因为它的出现或者破产，或者大发横财。

重积分变量代换公式

本文涉及的 n 元重积分变量代换定理的证明出自《数学分析新讲》第二册，限于篇幅，这里只给出了《新讲》中的核心证明。

由于积分的 **Riemann** 和定义可以看做是“体积线性组合”（然后再求极限），所以解决在变换 φ 下的体积变化关系是解决问题的关键。考察一个小立方体 C 在变换 φ 下 $\varphi(C)$ 的体积就是自然思路了。遗憾的是由于 $\varphi(C)$ 一般是非线性变换，所以 $\varphi(C)$ 多半要发生扭曲变形。直观上，如果立方体 C 充分小，变换 φ 又比较光滑，那么 $\varphi(C)$ 扭曲程度就不那么厉害。



借用上图说明一下（这里 $g = \varphi$ ）

由微分思想可知，一个映射 φ 的作用在一点 x_0 邻域范围内可以近似看做是一个局部线性映射 $L(x) = \varphi(x_0) + \varphi'(x_0)(x - x_0)$ 。

稍微精确一些描述，取定小立方体 C 的中心点 x_0 ，那么当 C 充分小时，线性变换 $L(x) = \varphi(x_0) + \varphi'(x_0)(x - x_0)$ 与变换 $\varphi(x)$ 相差高阶无穷小量。也就是说处理问题时， $\varphi(x)$ 局部上可由线性变换近似代替，微分的精华思想也在于此。

根据线性代数中的一个结论，对可逆线性变换 A ，有体积关系

$$\mu(A(C)) = |\det A| \mu(C)$$

这里用符号 μ 表示若尔当体积。

令线性变换 $A = L$ ，则 $\det A = \det \varphi'(x_0)$ 或写成 $\det A = \det D\varphi(x_0)$ 。

由此得到体积近似关系: $\mu(\varphi(C)) \approx |\det \varphi'(x_0)| \mu(C)$. 进而

$$f(x)\mu(\varphi(C)) \approx f(\varphi(t))|\det \varphi'(t_0)|\mu(C)$$

其中 $x = \varphi(t)$. 根据积分的 Riemann 和定义, 不难猜出重积分变量代换公式

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt$$

此处 x, t 都是向量, $D\varphi(t) = \varphi'(t)$ 表示雅可比矩阵 (导数的推广). 上述想法表述为下面的重积分变量换元公式:

设 $\Omega \subset \mathbf{R}^m$ 是一个开集,

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射, $E \subset \Omega$ 是一个闭若尔当可测集. 如果

(1) 雅可比行列式 $\det D\varphi(t) \neq 0, \forall t \in \text{int } E$;

(2) φ 在 $\text{int } E$ 中是单射.

那么 $\varphi(E)$ 也是一个闭若尔当可测集, 并且对于任何在 $\varphi(E)$ 上连续的函数 $f(x)$ 都有

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt.$$

这个定理的证明有些复杂, 细节见《数学分析新讲》第二册, 基本想法是“采用把局部微分同胚分解成简单微分同胚的复合, 然后再转化为累次积分来证明”. 张老师把它拆成若干引理, 每个引理相对容易证明一些, 同时使得证明框架结构更清晰.

为方便讨论, 引入 \max 度量 $\rho(x, y) = |x - y| = \max |x^i - y^i|$, 这里 x^i 是向量 x 的第 i 个坐标分量. 这种 \max 度量比一般的欧氏度量 $\sqrt{\quad}$ 要方便一些, 不过在有限维空间中二者是等价的. 显然, 在 \max 度量下的邻域 $U_\rho(x, \delta)$ 都是开方块.

对非空集合 $S \subset \mathbf{R}^m$, 定义度量

$$\rho(x, S) = \inf_{y \in S} \rho(x, y)$$

引理 1: 设 $\Omega \subset \mathbf{R}^m$ 是一个开集,

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射, 则对任何紧致集 $K \subset \Omega$, 存在实数 $\lambda = \lambda(K)$, 使得

$$|\varphi(x) - \varphi(y)| \leq \lambda |x - y|, \forall x, y \in K$$

证明过程还是比较有技巧的, 这一引理主要用于不等式估计.

引理 2: 设 $\Omega \subset \mathbf{R}^m$ 是一个开集,

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射。如果 Λ 是 \mathbf{R}^m 中的一个零集, 它的闭包 $\text{Cl } \Lambda \subset \Omega$, 那么 $\varphi(\Lambda)$ 也是一个零集。

设 $\Omega \subset \mathbf{R}^m$ 是一个开集, $\varphi: \Omega \rightarrow \mathbf{R}^m$ 是一个连续可微映射, $E \subset \Omega$ 是一个若尔当可测集。如果要考察集合 $\varphi(E)$ 的若尔当可测性, 就需要了解边界 $\text{Bd } \varphi(E)$ 是否为零集。请不要误以为

$$\text{Bd } \varphi(E) = \varphi(\text{Bd } E)$$

下面的反例说明这等式一般并不成立。

例 考察 \mathbf{R}^2 中的点集

$$E = \{(x, y) \mid x^2 + y^2 \leq 1, y \geq 0\}$$

和连续可微映射

$$\begin{aligned} \varphi: \mathbf{R}^2 &\longrightarrow \mathbf{R}^2 \\ (x, y) &\mapsto (x^2 - y^2, 2xy) \end{aligned}$$

(这是复映射 $w = z^2$ 所对应的实映射)。显然, φ 在 E 的内部是单一的, 且

$$\det D\varphi(x, y) = \begin{vmatrix} 2x & -2y \\ 2y & 2x \end{vmatrix} = 4(x^2 + y^2) > 0, \forall (x, y) \in \text{int } E$$

但即使这样的条件下, 仍然不能保证 $\text{Bd } \varphi(E) = \varphi(\text{Bd } E)$ 。

实际上, E 的一部分边界点经 φ 映射之后变成了 $\varphi(E)$ 的内点, 所以对本例的情形有

$$\text{Bd } \varphi(E) \neq \varphi(\text{Bd } E).$$

引理 3: 设 $\Omega \subset \mathbf{R}^m$ 是一个开集,

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个映射, $E \subset \Omega$ 是一个有界闭集。

(1) 如果 φ 是连续映射, 那么 $\varphi(E)$ 也是一个有界闭集;

(2) 如果 φ 是连续可微映射, 并且满足这样的条件

$$\det D\varphi(x) \neq 0, \forall x \in \text{int } E$$

那么

$$\text{Bd } \varphi(E) \subset \varphi(\text{Bd } E).$$

证明: (1) 我们指出 $\varphi(E)$ 是一个列紧集, 因而也是有界闭集。设 $\{v_n\}$ 是 $\varphi(E)$ 中的任意一个点列。因为 $v_n \in \varphi(E)$, 所以存在 $u_n \in E$, 使得 $\varphi(u_n) = v_n$ 。

又因为 E 是有界闭集，也是列紧集，所以从 $\{u_n\}$ 中抽出一个子序列 $\{u_{n_k}\}$ ，这个子序列收敛于 E 中的某点 x ：

$$\lim u_{n_k} = x \in E$$

于是由 φ 的连续性

$$v_{n_k} = \varphi(u_{n_k}) \rightarrow \varphi(x) \in \varphi(E)$$

我们证明了 $\varphi(E)$ 也是一个列紧集，也是有界闭集。

(2) 因为 $\varphi(E)$ 是一个有界闭集，所以边界 $\text{Bd } \varphi(E) \subset \varphi(E)$ 。

对于任何 $y \in \text{Bd } \varphi(E) \subset \varphi(E)$ ，存在 $x \in E$ ，使得 $\varphi(x) = y$ 。我们指出

$$x \notin \text{int } E$$

否则由逆映射定理就会得出 $y = \varphi(x) \in \text{int } \varphi(E)$ ，这就产生矛盾！于是，只能有

$$x \in \text{Bd } E, y \in \varphi(\text{Bd } E)$$

这样，我们证明了

$$\text{Bd } \varphi(E) \subset \varphi(\text{Bd } E).$$

推论： 设 $\Omega \subset \mathbf{R}^m$ 是一个开集，

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射， $E \subset \Omega$ 是一个闭若尔当可测集。如果

$$\det D\varphi(x) \neq 0, \forall x \in \text{int } E$$

那么 $\varphi(E)$ 也是一个闭若尔当可测集。

简单图形逼近

如果 \mathbf{R}^n 的子集 S 可以表示为有限个两两无公共内点的闭方块的并集，那么我们就说 S 是一个简单图形。任何简单图形当然都是闭若尔当可测集。

引理 4： 设 $\Omega \subset \mathbf{R}^m$ 是一个开集，

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射， $E \subset \Omega$ 是一个闭若尔当可测集。如果

$$\det D\varphi(t) \neq 0, \forall t \in \text{int } E$$

那么对任何 $\varepsilon > 0$ ，存在简单图形

$$S = S_\varepsilon \subset \text{int } E$$

使得

$$\mu(E \setminus S) < \varepsilon \quad \text{且} \quad \mu(\varphi(E) \setminus \varphi(S)) < \varepsilon.$$

评注：引理 4 的证明比较琐碎，主要是一系列体积的分割，但结论本身还是很直观的，就是“体积下和”逼近思想。这里已经有了 Riemann 和的简单函数逼近思想了，这在下面引理 5 中体现得比较明显。引理 5 说明，要证明重积分变量代换公式，只需要考虑积分区域最简单的情形。

引理 5：设 $\Omega \subset \mathbf{R}^m$ 是一个开集， $E \subset \Omega$ 是一个闭若尔当可测集，而

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射，且满足

(1) 雅可比行列式 $\det D\varphi(t) \neq 0, \forall t \in \text{int } E$;

(2) φ 在 $\text{int } E$ 中是单射。

如果函数 f 在集合 $\varphi(E)$ 上是连续，并且对任意闭方块 $\Pi \subset \text{int } E$ ，都有

$$\int_{\varphi(\Pi)} f(x) dx = \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt$$

那么就有

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt$$

证明：根据引理 4，对任何 $\varepsilon > 0$ ，存在简单图形

$$S_\varepsilon \subset \text{int } E$$

使得

$$\mu(E \setminus S_\varepsilon) < \varepsilon \quad \text{且} \quad \mu(\varphi(E) \setminus \varphi(S_\varepsilon)) < \varepsilon.$$

于是对于简单图形 S_ε ，应该有

$$\int_{\varphi(S_\varepsilon)} f(x) dx = \int_{S_\varepsilon} f(\varphi(t)) |\det D\varphi(t)| dt$$

而我们又有

$$\left| \int_{\varphi(E)} f(x) dx - \int_{\varphi(S_\varepsilon)} f(x) dx \right| = \left| \int_{\varphi(E) \setminus \varphi(S_\varepsilon)} f(x) dx \right| \leq B \mu(\varphi(E) \setminus \varphi(S_\varepsilon)) < B\varepsilon$$

其中， $B = \sup_{x \in \varphi(E)} |f(x)|$ 。同理，得到不等式

$$\begin{aligned} & \left| \int_E f(\varphi(t)) |\det D\varphi(t)| dt - \int_{S_\varepsilon} f(\varphi(t)) |\det D\varphi(t)| dt \right| \\ &= \left| \int_{E \setminus S_\varepsilon} f(\varphi(t)) |\det D\varphi(t)| dt \right| \leq C \mu(E \setminus S_\varepsilon) < C\varepsilon \end{aligned}$$

其中， $C = \sup_{t \in E} |f(\varphi(t)) \det D\varphi(t)|$ 。

令 $\varepsilon \rightarrow 0$ 取极限, 就得到

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt.$$

简单变换情形

做完上述“Riemann 和”的铺垫工作后, 下面才进入核心想法: “采用把局部微分同胚分解成简单微分同胚的复合”, 然后再化为累次积分。为此, 先定义简单变换概念。

设 $\varphi^h(t^1, \dots, t^m): \Omega \rightarrow \mathbf{R}$ 是一个连续可微函数, 我们把如下形状的变换称为简单变换:

$$\begin{cases} x^i = t^i, i \neq h \\ x^h = \varphi^h(t^1, \dots, t^m) \end{cases}$$

换句话说, 简单变换是这样一种连续可微映射

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

它(至多)只改变 $t = (t^1, \dots, t^m) \in \Omega$ 的一个坐标。

我们先对简单变换证明重积分变量代换公式成立。

引理 6: 设 $\Omega \subset \mathbf{R}^m$ 是一个开集, $E \subset \Omega$ 是一个闭若尔当可测集, 而

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个简单变换, 且满足

- (1) 雅可比行列式 $\det D\varphi(t) \neq 0, \forall t \in \text{int } E$;
- (2) φ 在 $\text{int } E$ 中是单射。

如果函数 f 在集合 $\varphi(E)$ 上是连续, 那么

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt$$

证明: 必要时给变量重新编号, 可设简单变换 $x = \varphi(t)$ 具有这样的形式

$$\begin{cases} x^i = t^i, i = 1, 2, \dots, m-1 \\ x^m = \psi(t^1, \dots, t^m) \end{cases}$$

根据引理 5, 只须对任意闭方块 $\Pi \subset \text{int } E$, 证明公式

$$\int_{\varphi(\Pi)} f(x) dx = \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt.$$

不妨设在 Π 上有

$$\det D\varphi(t) = \frac{\partial \psi}{\partial t^m} > 0$$

(另一种情形可类似讨论)。把 Π 写成

$$\Pi = \Pi' \times \Pi'' \subset \mathbf{R}^{m-1} \times \mathbf{R}$$

这里

$$\Pi' = [\alpha^1, \beta^1] \times \cdots \times [\alpha^{m-1}, \beta^{m-1}]$$

$$\Pi'' = [\alpha, \beta]$$

于是, 我们有

$$\varphi(\Pi) = \left\{ (x', x^m) \left| \begin{array}{l} x' \in \Pi' \\ \psi(x', \alpha) \leq x^m \leq \psi(x', \beta) \end{array} \right. \right\}$$

因而由累次积分公式, 得到

$$\begin{aligned} \int_{\varphi(\Pi)} f(x) dx &= \int_{\Pi'} dx' \int_{\psi(x', \alpha)}^{\psi(x', \beta)} f(x', x^m) dx^m \\ &= \int_{\Pi'} dx' \int_{\alpha}^{\beta} f(x', \psi(x', t^m)) \frac{\partial \psi(x', t^m)}{\partial t^m} dt^m \\ &= \int_{\Pi'} dt' \int_{\alpha}^{\beta} f(t', \psi(t', t^m)) \frac{\partial \psi(t', t^m)}{\partial t^m} dt^m \\ &= \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt \end{aligned}$$

于是我们对简单变换证明了重积分变量代换公式成立。

对于一般的连续可微变换, 我们把它分解成一些列简单变换的复合。需要注意的是, 这种分解通常只在局部情形成立, 整体分解一般不存在。这里采用的思想就是“把局部微分同胚分解成简单微分同胚的复合”。

引理 7: 设 $\Omega \subset \mathbf{R}^m$ 是一个开集,

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射, $\tau \in \Omega$. 如果 $\det D\varphi(\tau) \neq 0$, 那么存在 $\delta > 0$, 使得重积分变量代换公式对于包含在邻域 $U_\rho(\tau, \delta)$ 之中的任何闭若尔当可测集成立。这就是说, 对任何闭若尔当可测集 $E \subset U_\rho(\tau, \delta)$ 和任何在 $\varphi(E)$ 上连续的函数 f , 都有

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt.$$

证明: 因为

$$\det D\varphi(\tau) = \frac{\partial(\varphi^1, \dots, \varphi^m)}{\partial(t^1, \dots, t^m)}(\tau) \neq 0$$

所以这行列式前 $m-1$ 行至少有一个不等于 0 的 $m-1$ 阶主子式。我们可以给变量 t^1, \dots, t^m 重新编号, 使得 $\det D\varphi(\tau)$ 的 $m-1$ 阶主子式

$$\frac{\partial(\varphi^1, \dots, \varphi^{m-1})}{\partial(t^1, \dots, t^{m-1})}(\tau) \neq 0$$

仿此, 用数学归纳法可以证明: 适当地给变量 t^1, \dots, t^m 编号, 可以使得 $\det D\varphi(\tau)$ 的各阶顺序主子式都不等于 0. 在下面的讨论中, 假定变量 t^1, \dots, t^m 已经按照这样的要求排列妥当。

我们定义 m 个变换,

$$\theta_k: \begin{cases} x^i = \varphi^i(t^1, \dots, t^m), & i \leq k \\ x^j = t^j, & j > k \end{cases}$$

这里 $k=1, 2, \dots, m$.

容易看出, $\det D\theta_k(\tau)$ 与 $\det D\varphi(\tau)$ 的第 k 个顺序主子式相等, 因而

$$\det D\theta_k(\tau) \neq 0, \quad k=1, 2, \dots, m$$

我们可以选取 $\delta > 0$ 充分小, 使得这 m 个变换 $\theta_1, \dots, \theta_m$ 在开集 $U_\rho(\tau, \delta)$ 之上都是微分同胚 (这里用到了逆映射定理)。再令

$$\begin{cases} \psi_1 = \theta_1 \\ \psi_k = \theta_k \circ \theta_{k-1}^{-1}, & k=2, \dots, m \end{cases}$$

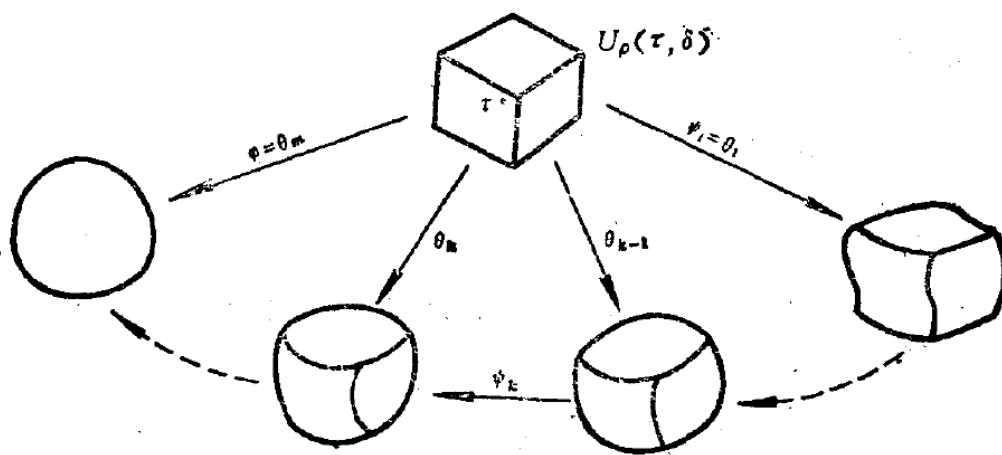
由此可知, $\psi_1 = \theta_1$ 是定义在开集 $U_\rho(\tau, \delta)$ 之上的简单变换, $\psi_k = \theta_k \circ \theta_{k-1}^{-1}$ 是定义在 $\theta_{k-1}(U_\rho(\tau, \delta))$ 之上的简单变换, 并且在 $U_\rho(\tau, \delta)$ 之上有

$$\varphi = \psi_m \circ \psi_{m-1} \circ \dots \circ \psi_1$$

这样我们已将 φ 局部地分解为简单变换的复合, 在此基础上逐次应用引理 6 得到

$$\begin{aligned} \int_{\varphi(E)} f(x) dx &= \int_{\psi_m \circ \psi_{m-1} \circ \dots \circ \psi_1(E)} f(x) dx \\ &= \int_{\psi_{m-1} \circ \dots \circ \psi_1(E)} f(\psi_m(u)) |\det D\psi_m(u)| du \\ &= \dots \\ &= \int_E f(\varphi(t)) |\det D\varphi(t)| dt \end{aligned}$$

这样我们就证明了引理 7.



简单变换分解 $\varphi = \psi_m \circ \psi_{m-1} \circ \cdots \circ \psi_1$ 示意图

到这里，我们可以直接证明重积分变量代换公式了。

(重积分变量代换定理) 设 $\Omega \subset \mathbf{R}^m$ 是一个开集，

$$\varphi: \Omega \rightarrow \mathbf{R}^m$$

是一个连续可微映射， $E \subset \Omega$ 是一个闭若尔当可测集。如果

- (1) 雅可比行列式 $\det D\varphi(t) \neq 0, \forall t \in \text{int } E$;
- (2) φ 在 $\text{int } E$ 中是单射。

那么 $\varphi(E)$ 也是一个闭若尔当可测集，并且对于任何在 $\varphi(E)$ 上连续的函数 $f(x)$ 都有

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt.$$

证明：根据引理 5，只需要对任意闭方块 $\Pi \subset \text{int } E$ 证明以下变量代换公式

$$\int_{\varphi(\Pi)} f(x) dx = \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt$$

成立即可。

为此，我们来考察

$$\Delta(\Pi) = \left| \int_{\varphi(\Pi)} f(x) dx - \int_{\Pi} f(\varphi(t)) |\det D\varphi(t)| dt \right|$$

对于任意一个 $\tau \in E$ ，存在一个相应的 $\delta = \delta(\tau) > 0$ ，使得 $U_\rho(\tau, \delta)$ 满足引理 7 的要求。开集族

$$\left\{ U_\rho\left(\tau, \frac{\delta(\tau)}{2}\right) \mid \tau \in E \right\}$$

覆盖了有界闭集 E (紧致性)。于是存在这开集族中的有限个开集

$$U_\rho(\tau_1, \frac{\delta_1}{2}), \dots, U_\rho(\tau_q, \frac{\delta_q}{2})$$

使得

$$E \subset \bigcup_{h=1}^q U_\rho(\tau_h, \frac{\delta_h}{2})$$

我们记

$$\eta = \min\{\frac{\delta_1}{2}, \dots, \frac{\delta_q}{2}\}$$

设闭方块 $\Pi \subset \text{int } E$. 我们可以把 Π 分割成两两无公共内点的闭子方块

$$\Pi_1, \dots, \Pi_r$$

使得各闭子方块的棱长都小于 η :

$$\ell(\Pi_k) < \eta, \quad k = 1, 2, \dots, r$$

如此, 每一个 Π_k 必定与某个

$$U_\rho(\tau_h, \frac{\delta_h}{2})$$

相交。因为 $\ell(\Pi_k) < \eta < \frac{1}{2}\delta_h$, 所以 Π_k 完全包含在 $U_\rho(\tau_h, \delta_h)$ 之中。于是,

$$\Delta(\Pi_k) = 0, \quad k = 1, 2, \dots, r.$$

又闭方块 Π 是两两无公共内点的闭子方块 Π_1, \dots, Π_r 的并集, 根据 $\Delta(\Pi)$ 的定义, 以及积分的有限可加性, 容易证明

$$0 \leq \Delta(\Pi) \leq \Delta(\Pi_1) + \dots + \Delta(\Pi_r)$$

由此可知 $\Delta(\Pi) = 0$, 所以我们证明了重积分变量代换公式:

$$\int_{\varphi(E)} f(x) dx = \int_E f(\varphi(t)) |\det D\varphi(t)| dt.$$

评注: 这个定理的具体点评, 可以参考本书第 2 章和第 3 章内容。逆映射定理如下, 设 $\Omega \subset \mathbf{R}^m$ 是一个开集, $\varphi: \Omega \rightarrow \mathbf{R}^m$ 是一个连续可微映射, $a \in \Omega$. 如果映射 φ 在 a 点对应的雅可比矩阵行列式满足

$$\det D\varphi(a) \neq 0$$

那么 φ 在 a 点是局部微分同胚。也就是存在包含 a 点的邻域 U 和包含 $b = \varphi(a)$ 点的邻域 V , 使得映射 $\varphi: U \rightarrow V$ 是微分同胚。

素数定理的简单解析证明笔记

SCIbird

作为《我的数学分析积木》一书收录的最后一个著名定理，笔者选择了素数定理。一方面，素数定理非常有名，是与费马大定理齐名的。另一方面，尽管定理本身属于数论范畴，但其证明是典型的分析方法。素数定理的解析证明融逼近、变换和分解三大分析基本思想与一身，非常具有代表性。

遵循全书的探索、猜想与证明风格，介绍了猜测素数定理的一种思路，同时给出完整的解析证明。主要参考文献是：华罗庚的《数论导引》和柯朗的《什么是数学》这两本书，解析证明引自下面的重要文献（引自《美国数学月刊》）：

Newman's Short Proof of the Prime Number Theorem

D. Zagier

Dedicated to the Prime Number Theorem on the occasion of its 100th birthday

与《数论导引》相比，这篇英文文献大幅简化了素数定理的解析证明。

我们知道素数 $2, 3, 5, 7, \dots$ 有无穷多个。古人曾经意图寻找素数的分布规律（如通项公式），但都以失败告终。素数的分布似乎毫无规律，总的来看很稀疏，偶尔有集聚现象，如存在相邻的素数，其差为 2，这对素数称为孪生素数。这方面有一个著名的猜想，即孪生素数猜想，由希尔伯特提出，内容为猜测存在无穷多对孪生素数。孪生素数猜想尚没有得到完全证明，但已有了阶段性成果。特别是最近传奇华人数学家张益唐取得了突破性结果，他的论文显示对于某一个小于 7 千万的数字 N ，存在无穷多的素数对，他们之间的差小于 N 。重要的是张的方法可以改进，进一步缩小 N 值，据称 Tao 等人最近热衷于此工作，且 N 值已经被缩小到 500 以内，大家可以在网上搜搜相关信息。

约定 p 始终代表素数， z, s 表示复数。记 $\pi(x)$ 表示不超过 x 的素数的个数，所谓素数定理，指下面极限表达式

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x / \ln x} = 1 \quad \text{或者} \quad \pi(x) \sim \frac{x}{\ln x}$$

这个结果是惊人的，尽管素数分布似乎没什么规律，但其个数的分布却有渐进规律。为了叙述方便，先引进两个重要的函数：

函数 $\text{Li } x = \int_2^x \frac{dt}{\ln t}$ (这是一个非初等积分) 由高斯引入的, 利用罗比达法则可知

$$\lim_{x \rightarrow \infty} \frac{\text{Li } x}{x / \ln x} = 1$$

函数 $\text{li } x$ 按如下方式定义

$$\text{li } x = \lim_{\eta \rightarrow 0} \left(\int_0^{1-\eta} + \int_{1+\eta}^x \right) \frac{dt}{\ln t}$$

两者之间的关系是: $\text{Li } x = \text{li } x - \text{li } 2$, 其中 $\text{li } 2 \approx 1.04$.

与前人思路不同, 高斯不是简单研究 $\pi(x)$ 的数值分布特点, 而是考察比值 $\pi(x)/x$ 的分布特点, 猜测是否有简单函数 $\varphi(x)$ 来近似代替。

按数学家迪厄多内的说法, 高斯考虑 $(x, x+1000)$ 内的所有素数, 通过大量的计算, 高斯从数值表中观察到, 当 x 很大时, 有下面的近似性质:

$$\frac{\pi(x+1000) - \pi(x)}{1000} \text{ 正比于 } \frac{1}{\ln x}$$

高斯猜想 $\text{Li } x$ 可以作为 $\pi(x)$ 的渐进表达式, 数值计算也支持这一猜想。

x	$\pi(x)$	$\frac{x}{\log x}$	$\text{li } x$	$\frac{\pi(x)}{\text{li } x}$	$\frac{\pi(x)}{x}$
1 000	168	145	178	0.94...	0.1680
10 000	1 229	1 086	1 246	0.98...	0.1229
50 000	5 133	4 621	5 167	0.993...	0.1026
100 000	9 592	8 686	9 630	0.996...	0.0959
500 000	41 538	38 103	41 606	0.9983...	0.0830
1 000 000	78 498	72 382	78 628	0.9983...	0.0785
2 000 000	148 933	137 848	149 055	0.9991...	0.0745
5 000 000	348 513	324 149	348 638	0.9996...	0.0697
10 000 000	664 579	620 417	664 918	0.9994...	0.0665
20 000 000	1 270 607	1 189 676	1 270 905	0.9997...	0.0635
90 000 000	5 216 954	4 913 897	5 217 810	0.99983...	0.0580
100 000 000	5 761 455	5 428 613	5 762 209	0.99986...	0.0576
1 000 000 000	50 847 478	48 254 630	50 849 235	0.99996...	0.0508

表格出自华罗庚的《数论导引》

高斯这一深刻的发现一方面包含辛勤的计算劳动, 同时体现了高斯的敏锐数学直觉。笔者当初阅读文献时第一感觉是高斯似乎在寻找素数分布的某种概率分布 (笔者下意识想到了高斯与正态分布), 正确地说, 高斯是在寻找素数的某种密度函数分布 (近似表达式)。

同一时期, 法国数学家勒让德 (此人数学上经常与高斯撞车) 也发现了素数分布近似关系 $\pi(x) \sim x/(\ln x - B)$, 其中 B 是一个常数。数学史上一般认为, 勒

让德和高斯是素数定理的“发现者”（他们没有给出数学证明）。

有没有比数值计算更有逻辑一些的近似推导，来猜测出素数定理的表达式呢？数学家柯朗在《什么是数学》一书中，给出一种从统计方法角度来猜测素数定理的探索途径。总的方针是寻找：一个简单连续函数 $\varphi(x)$ ，使得对较大的数 a 与 b 有近似关系

$$\pi(b) - \pi(a) \approx \int_a^b \varphi(x) dx$$

当然，如果对较大的数 x 有近似表达式

$$\pi(x) \approx \int_2^x \varphi(t) dt$$

就更好了！严格说，满足上述要求的简单函数 $\varphi(x)$ 未必存在，至少不能预先知道。不过从数学和谐角度看，可以假设这样的 $\varphi(x)$ 存在，下面我们通过形式推导来猜一猜 $\varphi(x)$ 的表达式。

初等数论中有一个关于阶乘 $n!$ 的定理，即素数 p 是 $n!$ 的一个素因子，则 p 作为因子的次数“约为”（当整数 n 很大时）

$$\alpha_p = \left[\frac{n}{p} \right] + \left[\frac{n}{p^2} \right] + \left[\frac{n}{p^3} \right] + \cdots \approx \frac{n}{p} + \frac{n}{p^2} + \frac{n}{p^3} + \cdots = \frac{n}{p-1}$$

这里 $[x]$ 表示不超过 x 的最大整数。

可以证明，满足 $p \leq n$ 的所有素数均为阶乘 $n!$ 的因子，反之亦然。于是

$$n! = \prod_{p \leq n} p^{\alpha_p} \approx \prod_{p \leq n} p^{\frac{n}{p-1}}$$

根据数学分析中的斯特林公式 $\ln n!$ 近似等于 $n \ln n$ ，故

$$\ln n \approx \sum_{p \leq n} \frac{\ln p}{p-1}$$

用 x 代替 n ，得到近似表达式

$$\ln x \approx \sum_{p \leq x} \frac{\ln p}{p-1}$$

因而对上式右边进行近似（估计）是一个突破口。

考虑区间 $[2, x)$ ，其内的素数按照从小到大顺序排列为

$$2 = p_1 < p_2 < \cdots < p_m$$

因为 $\ln x / (x-1)$ 是减函数，所以在区间 $[p_i, p_{i+1})$ 上有

$$\frac{\ln p_{i+1}}{p_{i+1}-1} \int_{p_i}^{p_{i+1}} \varphi(t) dt \leq \int_{p_i}^{p_{i+1}} \varphi(t) \frac{\ln t}{t-1} dt \leq \frac{\ln p_i}{p_i-1} \int_{p_i}^{p_{i+1}} \varphi(t) dt$$

根据假设, $\varphi(x)$ 应该使得 $\int_{p_i}^{p_{i+1}} \varphi(t) dt \approx 1$ (对较大的素数 p_i)。对上面不等式

在区间 $[2, x)$ 内的素数求和, 得到“近似”表达式

$$\int_2^x \varphi(t) \frac{\ln t}{t-1} dt \approx \sum_{p \leq x} \frac{\ln p}{p-1}$$

这个近似关系自然不是很严格, 但在形式推导之中还是有启迪作用的 (至少比没有方向强)。下一步, “大胆假设”

$$\ln x = \int_2^x \varphi(t) \frac{\ln t}{t-1} dt$$

等式两边对 x 求导, 得到

$$\varphi(x) = \frac{x-1}{x \ln x}$$

当 x 较大时, $(x-1)/x \approx 1$ 。因此猜测: 当 x 较大时,

$$\text{Li } x = \int_2^x \frac{dt}{\ln t}$$

作为 $\pi(x)$ 的近似表达式是一种可能。

有了猜测结果, 可以通过数值计算来验证一下。幸运的是, 数值计算表明近似效果不错, 这应该是形式推导与数值计算结合的一个典范。又因为

$$\lim_{x \rightarrow \infty} \frac{\text{Li } x}{x / \ln x} = 1$$

所以, 我们大胆猜测

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x / \ln x} = 1$$

这就是素数定理了。不过, 站在数值计算角度看, $x / \ln x$ 的精确性不及 $\text{Li } x$, 高斯当年的眼神还是很准的。

有了探索和猜想, 下面需要严格的数学证明了。素数定理的第一个证明来自阿达玛和普桑, 证明方法比较高深 (用到了复变函数论)。后来塞尔伯格和厄尔多斯给出了纯初等证明 (一堆繁琐的不等式估计, 思路比较复杂, 启发性不大), 证明细节可参考华老的《数论导引》。后人们不断简化素数定理的证明, 已经有好几种证明了。本书收录的这个解析证明非常具有代表性, 其融逼近、变换和分解三大分析基本思想于一体。

先定义三个重要的函数

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad \phi(s) = \sum_p \frac{\ln p}{p^s}, \quad \vartheta(x) = \sum_{p \leq x} \ln p$$

第一个函数 $\zeta(s)$ 就是传说中的黎曼 ζ 函数！这方面有太多熟知的重要结论了，任何一本解析数论教材都会用相当的篇幅来介绍黎曼 ζ 函数的若干基本性质。

函数 $\zeta(s)$ 能与素数发生联系，关键在于下面的无穷乘积（欧拉恒等式）

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \frac{1}{1-p^{-s}}, \quad \operatorname{Re}(s) > 1$$

考虑等比级数

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots$$

令 $x = p^{-s}$ ，利用整数的素数分解定理，有

$$\begin{aligned} \prod_p \frac{1}{1-p^{-s}} &= (1 + 2^{-s} + 2^{-2s} + \cdots)(1 + 3^{-s} + 3^{-2s} + \cdots) \cdots \\ &= 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^s} = \zeta(s) \end{aligned}$$

当 $\operatorname{Re}(s) > 1$ 时， $\zeta(s)$ 定义了一个解析函数。特别地，利用欧拉恒等式可知，当 $\operatorname{Re}(s) > 1$ 时，有 $\zeta(s) \neq 0$ 。

解析函数论中已经证明了， $\zeta(s)$ 可以解析延拓到整个复平面上，得到一个亚纯函数，它只在 $s=1$ 处有一个一阶极点。不过，为证明素数定理，只需要 $\zeta(s)$ 能解析延拓到 $\operatorname{Re}(s) > 0$ 即可，为此考虑下面的等式

$$\zeta(s) - \frac{1}{s-1} = \sum_{n=1}^{\infty} \frac{1}{n^s} - \int_1^{\infty} \frac{1}{x^s} dx = \sum_{n=1}^{\infty} \int_n^{n+1} \left(\frac{1}{n^s} - \frac{1}{x^s} \right) dx$$

我们证明等式右边的级数当 $\operatorname{Re}(s) > 0$ 时是绝对收敛的。为此考虑

$$\left| \int_n^{n+1} \left(\frac{1}{n^s} - \frac{1}{x^s} \right) dx \right| = \left| s \int_n^{n+1} dx \int_n^x \frac{1}{u^{s+1}} du \right| \leq \max_{n \leq u \leq n+1} \left| \frac{s}{u^{s+1}} \right| = \frac{|s|}{n^{\operatorname{Re}(s)+1}}$$

由比较判别法可知函数 $\zeta(s)$ 能解析延拓到 $\operatorname{Re}(s) > 0$ 且在 $s=1$ 处有一个一阶极点。

当初阿达玛和普桑证明素数定理时，关键用到了函数 $\zeta(s)$ 在 $\operatorname{Re}(s)=1$ 上没有零点这一结论。

引理 1: 函数 $\zeta(s)$ 在 $\operatorname{Re}(s) \geq 1$ 上无零点, 即 $\zeta(s) \neq 0$.

证明: 引理 1 的证明比较有技巧, 被多数解析数论教材所收录。关键在于下面的著名不等式: $\forall x, y \in \mathbf{R}$ 且 $x > 1$, 则

$$|\zeta^3(x)\zeta^4(x+iy)\zeta^2(x+2iy)| \geq 1$$

根据欧拉恒等式, 只需要对每一个素数 p , 证明不等式

$$\left| \left(1 - \frac{1}{p^x}\right)^3 \left(1 - \frac{1}{p^{x+iy}}\right)^4 \left(1 - \frac{1}{p^{x+2iy}}\right)^2 \right| \leq 1$$

记 $r = 1/p^x, e^{i\theta} = 1/p^{iy}$, 则 $0 < r < 1$. 上面不等式等价于

$$|(1-r)^3(1-re^{i\theta})^4(1-re^{2i\theta})^2| \leq 1$$

于是只需要对 $0 < r < 1$, 证明

$$|(1-re^{i\theta})^4(1-re^{2i\theta})^2| \leq \frac{1}{(1-r)^3}$$

固定 r 不动。直接计算, 得到

$$|(1-re^{i\theta})^4(1-re^{2i\theta})^2| = (1+r^2-2r\cos\theta)^2(1+r^2-2r\cos 2\theta)$$

再令 $a = \cos\theta$, 则 $\cos 2\theta = 2\cos^2\theta - 1 = 2a^2 - 1$, 利用算术-几何均值不等式

$$\begin{aligned} |(1-re^{i\theta})^4(1-re^{2i\theta})^2| &= (1+r^2-2ra)^2(1+2r+r^2-4ra^2) \\ &\leq \left[\frac{(1+r^2-2ra) + (1+r^2-2ra) + (1+2r+r^2-4ra^2)}{3} \right]^3 \\ &= \left[\frac{3(1+r+r^2)-r(1+2a)^2}{3} \right]^3 \leq (1+r+r^2)^3 \\ &\leq (1+r+r^2+\cdots)^3 = \left(\sum_{n=0}^{\infty} r^n \right)^3 = \frac{1}{(1-r)^3} \end{aligned}$$

为证明引理 1 成立, 反证假设存在 y_0 , 使得 $\zeta(1+iy_0) = 0$. 前面论证了 $\zeta(s)$ 函数在 $s = 1+2iy_0$ 处解析, 且有惟一的极点 $s = 1$ (一阶)。所以

$$\lim_{x \rightarrow 1+} \zeta(x)^3 \zeta(x+iy_0)^4 \zeta(x+2iy_0)^2 = 0$$

与上面的不等式矛盾! 引理 1 成立。

注: 严格的证明还要讨论 $3(1+r+r^2)-r(1+2a)^2$ 的符号, 然后才能放缩。推导过程中也利用了等比级数求和公式。为避免算术-几何均值不等式, 也可以设

$$h(a) = (1+r^2-2ra)^2(1+2r+r^2-4ra^2)$$

然后用导数法求出 $\max h(a)$, 这都是数学分析中常见的方法, 只是计算相对繁琐而已。

$\zeta(s)$ 函数的零点分布非常重要，大名鼎鼎的黎曼假设（悬赏 100 万美元）被认为是数学中的第一难题：

$\zeta(s)$ 函数在 $0 < \operatorname{Re}(s) < 1$ 内的零点全部在直线 $\operatorname{Re}(s) = \frac{1}{2}$ 上。

黎曼假设有一个与素数定理相关的等价命题：

“黎曼假设”等价于 $\pi(x) = \operatorname{Li} x + O(\sqrt{x} \ln x)$

$\zeta(s)$ 函数之后，下一个重要的“桥梁函数”是 $\phi(s)$ ，我们有如下结论：

引理 2: $\phi(s) - \frac{1}{s-1}$ 在 $\operatorname{Re}(s) \geq 1$ 上是解析的。

证明：利用 $\zeta(s)$ 函数的无穷乘积表达式，可知

$$\begin{aligned} -\frac{\zeta'(s)}{\zeta(s)} &= -\left(\ln\left(\prod_p \frac{1}{1-p^{-s}}\right)\right)' \\ &= \sum_p \left(\ln(1-p^{-s})\right)' = \sum_p \frac{p^{-s} \ln p}{1-p^{-s}} = \sum_p \frac{\ln p}{p^s-1} \end{aligned}$$

然后利用裂项技巧

$$\frac{1}{p^s-1} = \frac{1}{p^s} - \frac{1}{p^s(p^s-1)}$$

得到

$$-\frac{\zeta'(s)}{\zeta(s)} = \phi(s) + \sum_p \frac{\ln p}{p^s(p^s-1)}$$

等式中的无穷级数在 $\operatorname{Re}(s) > \frac{1}{2}$ 时是收敛的。

于是 $\phi(s)$ 是 $\operatorname{Re}(s) \geq 1$ 上的亚纯函数，它的极点只能是 $s=1$ 和 $\zeta(s)$ 的零点。但是，根据引理 1， $\zeta(s)$ 在 $\operatorname{Re}(s) \geq 1$ 上满足 $\zeta(s) \neq 0$ ，所以 $\phi(s)$ 的极点是 $s=1$ ，为一阶极点。由此可知，

$\phi(s) - \frac{1}{s-1}$ 在 $\operatorname{Re}(s) \geq 1$ 上是解析的。

注：以上的变换虽然有些眼花缭乱，但都是简单的基本计算，只要亲自动手算一算，不难理解。关于 $\zeta(s)$ 函数更多的性质，读者可参考解析数论方面的教材。

接下来我们的重点是研究 $\vartheta(x)$ 函数，我们将证明

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x / \ln x} = \lim_{x \rightarrow \infty} \frac{\vartheta(x)}{x} = 1$$

即通过 $\vartheta(x)$ 函数来证明素数定理。

引理 3: $\vartheta(x) \sim O(x)$. (称 $f \sim O(g)$, 如果存在常数 c 满足 $|f| \leq cg$)

证明: 这里的关键地方是利用了二项式定理展开,

$$2^{2n} = (1+1)^{2n} = C_{2n}^0 + C_{2n}^1 + \cdots + C_{2n}^{2n} \geq C_{2n}^n$$

注意到

$$C_{2n}^n = \frac{(2n)!}{(n!)^2} \geq \prod_{n < p \leq 2n} p = e^{\vartheta(2n) - \vartheta(n)}$$

由此可知

$$\vartheta(2n) - \vartheta(n) \leq 2n \ln 2$$

利用归纳法, 不难得到

$$\vartheta(2^n) < 2^{n+1} \ln 2$$

$\forall x \in \mathbf{R} (x \geq 1)$, 存在正整数 m 满足 $2^{m-1} \leq x < 2^m$, 故

$$\vartheta(x) \leq \vartheta(2^m) \leq 2^{m+1} \ln 2 = (4 \ln 2) 2^{m-1} \leq (4 \ln 2) x$$

由此可知 $\vartheta(x) \sim O(x)$.

本文后面将利用变换的方法证明更强的结论 $\vartheta(x) \sim x$. 方法是利用所谓的“Tauber 型定理”, 这方面内容可参考 Rudin 的《泛函分析》第 9 章 Tauber 理论. 证明想法非常巧妙: 通过研究 $f(t)$ 在某个线性变换 $L[f(t)]$ 下的渐进性质, 来研究函数 $f(t)$ 的无穷远处的渐进性质, 这里选择的线性变换是 Laplace 变换.

引理 4: 广义积分 $\int_1^\infty \frac{\vartheta(x) - x}{x^2} dx$ 收敛。

引理 4 的证明比较繁琐, 我们后面会细说. 假定引理 4 成立, 则利用反证法和广义积分收敛的柯西准则可证明 $\vartheta(x) \sim x$.

若存在常数 $\lambda > 1$, 使得对充分大的 x 有 $\vartheta(x) \geq \lambda x$ 成立. 注意到 ϑ 函数是不减的, 于是

$$\int_x^{\lambda x} \frac{\vartheta(t) - t}{t^2} dt \geq \int_x^{\lambda x} \frac{\lambda x - t}{t^2} dt = \int_1^\lambda \frac{\lambda - u}{u^2} du > 0$$

与广义积分的柯西收敛准则矛盾! 类似地, 存在正的常数 $\lambda < 1$, 使得对充分大的 x 满足 $\vartheta(x) \leq \lambda x$, 同理可证

$$\int_{\lambda x}^x \frac{\vartheta(t) - t}{t^2} dt \leq \int_{\lambda x}^x \frac{\lambda x - t}{t^2} dt = \int_\lambda^1 \frac{\lambda - u}{u^2} du < 0$$

矛盾. 于是利用反证法证明了 $\vartheta(x) \sim x$.

有了 $\vartheta(x) \sim x$ 的结论, 我们利用简单的不等式放缩来证明素数定理。

(素数定理) $\lim_{x \rightarrow \infty} \frac{\pi(x)}{x / \ln x} = 1$

证明: 由不等式 $\vartheta(x) = \sum_{p \leq x} \ln p \leq \sum_{p \leq x} \ln x = \pi(x) \ln x$ 以及 $(\forall \varepsilon > 0)$

$$\begin{aligned} \vartheta(x) = \sum_{p \leq x} \ln p &\geq \sum_{x^{1-\varepsilon} \leq p \leq x} \ln p \geq \sum_{x^{1-\varepsilon} \leq p \leq x} (1-\varepsilon) \ln x \\ &= (1-\varepsilon) \ln x [\pi(x) - \pi(x^{1-\varepsilon})] \end{aligned}$$

又 $\pi(x^{1-\varepsilon}) \leq x^{1-\varepsilon}$, 所以

$$\vartheta(x) \geq (1-\varepsilon) \ln x [\pi(x) - x^{1-\varepsilon}]$$

进而得到不等式估计

$$\frac{\vartheta(x)}{x} \leq \frac{\pi(x) \ln x}{x} \leq \frac{1}{1-\varepsilon} \frac{\vartheta(x)}{x} + \frac{1}{1-\varepsilon} \frac{\ln x}{x^\varepsilon}$$

利用 $\vartheta(x) \sim x$ 两边取极限, 由夹逼定理以及 ε 可以任意小, 得

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x / \ln x} = 1$$

这就证明了素数定理。

上面的证明过程充分体现了分析中的逼近思想。我们接下来利用变换的思想证明引理 4, 这样就给出素数定理一个完整的数学证明。引理 4 的证明需要一点复变函数知识, 即 Laplace 变换与柯西积分公式。

引理 5: 对 $\vartheta(x)$ 函数, 有如下 Laplace 变换 ($\operatorname{Re}(s) > 1$)

$$\frac{\phi(s)}{s} = \int_0^{+\infty} \vartheta(e^t) e^{-st} dt$$

证明: 若引理 5 成立。令 $f(t) = \vartheta(e^t) e^{-t} - 1$ 及 $g(z) = \phi(z+1)/(z+1) - 1/z$, 则

$$g(z) = \int_0^{+\infty} f(t) e^{-zt} dt \quad (\operatorname{Re}(z) > 0)$$

再做变换 $x = e^t$, 当 $z \rightarrow 0$ 时就得到我们所需证明的广义积分, 即

$$\int_1^\infty \frac{\vartheta(x) - x}{x^2} dx = \int_0^{+\infty} f(t) dt$$

这就是我们为什么要考虑引理 5 中变换的原因。

将素数按由小到大顺序依次记为 p_1, p_2, p_3, \dots , 不妨设 $p_0 = 0$. 于是, 当 $\ln p_n < t < \ln p_{n+1}$ 时, $\vartheta(e^t)$ 为常数 (实际上 $\vartheta(x)$ 是阶梯函数)。

直接计算, 得到

$$\int_{\ln p_n}^{\ln p_{n+1}} \vartheta(e^t) e^{-st} dt = \vartheta(p_n) \int_{\ln p_n}^{\ln p_{n+1}} e^{-st} dt = \frac{1}{s} \vartheta(p_n) \left(\frac{1}{p_n^s} - \frac{1}{p_{n+1}^s} \right)$$

再利用 $\vartheta(p_n) - \vartheta(p_{n-1}) = \ln p_n$ ，求和得到

$$\begin{aligned} \int_0^{+\infty} \vartheta(e^t) e^{-st} dt &= \sum_{n=1}^{\infty} \int_{\ln p_n}^{\ln p_{n+1}} \vartheta(e^t) e^{-st} dt = \frac{1}{s} \sum_{n=1}^{\infty} \vartheta(p_n) \left(\frac{1}{p_n^s} - \frac{1}{p_{n+1}^s} \right) \\ &= \frac{1}{s} \sum_{n=1}^{\infty} \frac{\vartheta(p_n) - \vartheta(p_{n-1})}{p_n^s} = \frac{1}{s} \sum_{n=1}^{\infty} \frac{\ln p_n}{p_n^s} = \frac{1}{s} \phi(s) \end{aligned}$$

于是引理 5 成立。

令 $f(t) = \vartheta(e^t) e^{-t} - 1$ 及 $g(z) = \phi(z+1)/(z+1) - 1/z$ ，做变换 $x = e^t$ ，得到

$$\int_1^{\infty} \frac{\vartheta(x) - x}{x^2} dx = \int_0^{+\infty} f(t) dt$$

为了证明这个广义积分的存在性，考察极限过程 $\lim_{z \rightarrow 0} g(z)$ 的思路就自然了。

引理 6: 设 $t \geq 0$ ， $f(t)$ 是有界的局部可积函数，定义函数

$$g(z) = \int_0^{+\infty} f(t) e^{-zt} dt \quad (\operatorname{Re}(z) > 0)$$

若函数 $g(z)$ 可以解析延拓到 $\operatorname{Re}(z) \geq 0$ ，则广义积分收敛且满足

$$g(0) = \lim_{z \rightarrow 0+} \int_0^{+\infty} f(t) e^{-zt} dt = \int_0^{+\infty} f(t) dt$$

证明：令 $g(z) = \phi(z+1)/(z+1) - 1/z$ ，由引理 2 可知 $g(z)$ 可以解析延拓到 $\operatorname{Re}(z) \geq 0$ ，于是引理 4 可以从引理 6 推出来。后者的证明需要用到复变函数里的柯西定理，通过巧妙选择被积函数和积分围道，可以证明上述积分号下求极限定理成立。

对 $T > 0$ ，定义函数

$$g_T(z) = \int_0^T f(t) e^{-zt} dt$$

则 $g_T(z)$ 是复平面上的解析函数。于是我们只需证明

$$\lim_{T \rightarrow \infty} g_T(0) = g(0)$$

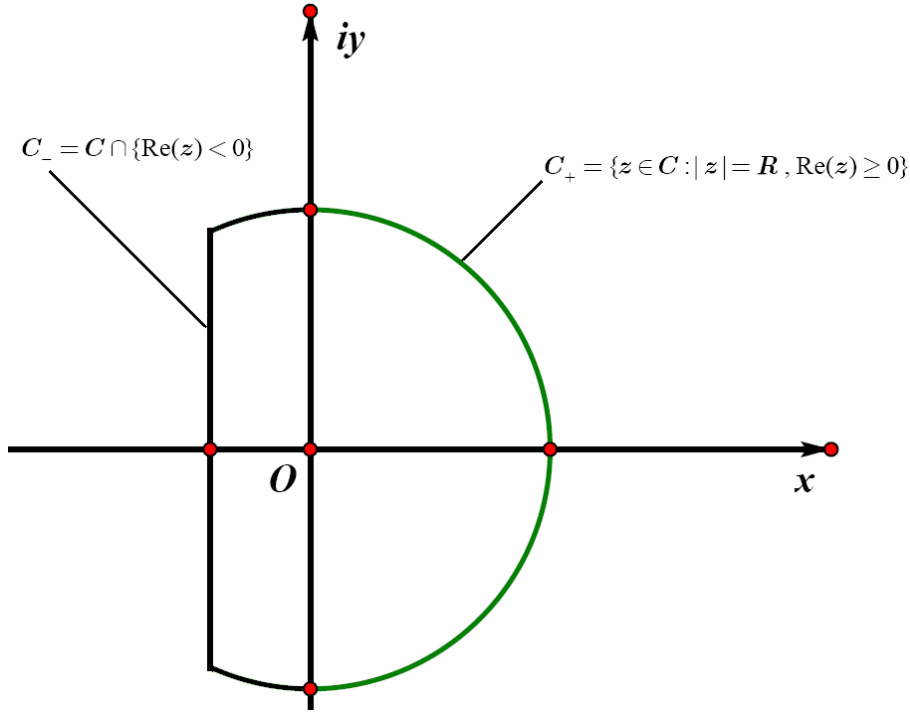
取定充分大的正实数 R 以及充分小的正实数 $\delta > 0$ (与 R 有关)，记曲线 C 为区域 $D = \{z \in \mathbb{C} : |z| \leq R, \operatorname{Re}(z) > -\delta\}$ 的边界 (即 $C = \partial D$)，满足 $g(z)$ 在 D 的内部以及边界 C 上是解析的。

记 $F(z) := (g(z) - g_T(z)) e^{zT} \left(1 + \frac{z^2}{R^2} \right)$ ，由柯西积分公式可知

$$\begin{aligned}
g(0) - g_T(0) &= F(0) = \frac{1}{2\pi i} \int_C \frac{F(z)}{z} dz \\
&= \frac{1}{2\pi i} \int_C (g(z) - g_T(z)) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z}
\end{aligned}$$

接下来要对上面的围道积分进行细致的分段估计，这种技巧在分析中非常常见。

记曲线 $C_+ = \{z \in C : |z| = R, \operatorname{Re}(z) \geq 0\}$ 和 $C_- = C \cap \{\operatorname{Re}(z) < 0\}$ ，则 $C = C_+ \cup C_-$



下面在这两段曲线上分别进行估计。记 $B = \max_{t \geq 0} |f(t)|$ ，则

$$|g(z) - g_T(z)| = \left| \int_T^\infty f(t) e^{-zt} dt \right| \leq B \int_T^\infty |e^{-zt}| dt = \frac{B e^{-\operatorname{Re}(z)T}}{\operatorname{Re}(z)}$$

另外，不难计算出

$$\left| e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{1}{z} \right| = e^{\operatorname{Re}(z)T} \cdot \frac{2\operatorname{Re}(z)}{R^2}$$

合并到一起，得到不等式估计

$$|F(z)| \leq \frac{2B}{R^2}, \quad z \in C_+$$

所以

$$\left| \frac{1}{2\pi i} \int_{C_+} (g(z) - g_T(z)) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} \right| \leq \frac{B}{R}$$

在 $C_- = C \cap \{\operatorname{Re}(z) < 0\}$ 上进行积分要更细致地处理，这里对 $g_T(z)$ 和 $g(z)$ 分开单独讨论。

因为函数 $g_T(z)$ 在全平面上解析, 所以在曲线 C_- 上的积分值只依赖于端点。我们可以进行积分途径形变, 即用半圆 $C'_- = \{z \in C : |z| = R, \operatorname{Re}(z) < 0\}$ 取代曲线 C_- , 而积分不变, 即

$$\frac{1}{2\pi i} \int_{C_-} g_T(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} = \frac{1}{2\pi i} \int_{C'_-} g_T(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z}$$

于是可以重复上面的不等式估计

$$|g_T(z)| = \left| \int_0^T f(t) e^{-zt} dt \right| \leq B \int_{-\infty}^T |e^{-zt}| dt = \frac{B e^{-\operatorname{Re}(z)T}}{|\operatorname{Re}(z)|}, \operatorname{Re}(z) < 0$$

同样有

$$\left| \frac{1}{2\pi i} \int_{C_-} g_T(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} \right| \leq \frac{B}{R}$$

最后只需要对积分

$$\frac{1}{2\pi i} \int_{C_-} g(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z}$$

进行估计了。利用 Lebesgue 控制收敛定理 (积分号下取极限), 可知

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi i} \int_{C_-} g(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} = 0$$

综上所述,

$$\limsup_{T \rightarrow \infty} |g(0) - g_T(0)| = \limsup_{T \rightarrow \infty} \left| \int_{C_+} + \int_{C_-} \right| \leq \frac{2B}{R}$$

再令 $R \rightarrow \infty$, 就得到了

$$g(0) = \lim_{z \rightarrow 0+} \int_0^{+\infty} f(t) e^{-zt} dt = \int_0^{+\infty} f(t) dt$$

于是, 广义积分收敛, 引理 6 成立, 从而引理 4 也成立, 这就给素数定理一个严格的解析证明。

评注: 纵观素数定理整个证明过程, 跌宕起伏, 逼近、变换与分解三大分析技巧反复融合, 交叉运用。证明过程也体现了复变函数这门数学工具的强大力量。

高斯最初猜想用 $\operatorname{Li} x$ 作为 $\pi(x)$ 的近似表达式, 不过 $\operatorname{Li} x$ 是非初等函数, 人们对此不太熟悉。于是在极限意义下, 人们找到了其等价无穷大函数 $x / \ln x$, 这是一个初等函数。根据素数定理, 粗略地说, 素数在区间 $[1, x]$ 中的分布密度

大约是 $1/\ln x$ ，这也是高斯当初的洞察力所在。

当然，很多人可能仍觉得这个解析证明较难，但考虑到素数定理的证明是一个“菲尔兹奖级别”的数学工作，上面的证明已经化简的很多了，相信对认真学过数学分析和复变函数的读者来说，难度不大。

最后再补充素数定理的一些内容。大量数值计算表明（例如首页的表格），似乎有 $\pi(x) - \text{li } x < 0$ 成立，很多人都认为这个猜想成立。但是数学家李特尔伍德证明：

存在无数多个 x ，使得 $\pi(x) - \text{li } x > 0$ ；存在无数多个 x ，使得 $\pi(x) - \text{li } x < 0$ 。
即 $\pi(x) - \text{li } x$ 的符号是剧烈振荡的。

后来有人推断之所以误认为上述猜想是正确的，是因为使得 $\pi(x) - \text{li } x > 0$ 成立的 x 是一个天文数字。这个例子表明数学证明的优越性，你不可能通过数值计算验证所有的情况，但可以给出一个否定形式的证明。

出乎意料，黎曼假设等价于 $\pi(x) = \text{Li } x + O(\sqrt{x} \ln x)$ ，如果该表达式成立，则可认为这是素数定理误差估计的最佳形式了（几乎难以改进）。

当初很多数学家认为素数定理不太可能有初等证明，一个重要原因是阿达玛等人的证明之中利用了黎曼 $\zeta(s)$ 函数零点的性质，不少人觉得素数定理的证明难以规避 $\zeta(s)$ 函数，因此不太可能有初等证明。后来，塞尔伯格和厄尔多斯分别给出了素数定理的初等证明（不用微积分），证明涉及一大堆繁琐的不等式估计，思路晦涩，启发性不大，华老等不太推荐。

应该指出，数论中有很多难题，难度不亚于素数定理，但难度不是数学的全部，特别是面对浩瀚的数学海洋时，必须学会取舍。笔者不太擅长数论，也不是数论粉丝，之所以关注素数定理，与数学偶像高斯有关，在整理《高斯经典文章及相关数学工作汇编》一书过程中，通过进一步了解素数定理，被其深刻优美的特点所吸引。特别是后来偶然看到这个简短的解析证明之后，更是拍案叫绝，因此有了本文。

本文没有反复精雕细琢，错误难免，读者需细心才行（最好亲自动手推导一下）。写作风格上，延续了猜想、案例验证（数值计算）与数学证明为一体的

style, 这样的写法能够使得读者阅读时有一个循序渐进的思考过程, 同时也能感受下数学研究的思维痕迹。

x	$\pi(x)$	$\frac{x}{\log(x)}$	Ratio
10^1	4	4.34	0.9217
10^2	25	21.71	1.1515
10^3	168	144.76	1.1605
10^4	1229	1085.74	1.1319
10^5	9592	8685.89	1.1043
10^6	78498	72382.41	1.0845
10^7	664579	620420.69	1.0712
10^8	5761455	5428681.02	1.0613
10^9	50847534	48254942.43	1.0537
10^{10}	455052511	434294481.90	1.0478
10^{11}	4118054813	3948131653.67	1.0430
10^{12}	37607912018	36191206825.27	1.0391
10^{13}	346065536839	334072678387.12	1.0359
10^{14}	3204941750802	3102103442166.08	1.0331
10^{15}	29844570422669	28952965460216.79	1.0308
10^{16}	279238341033925	271434051189532.38	1.0288
10^{17}	2623557157654233	2554673422960304.87	1.0270
10^{18}	24739954287740860	24127471216847323.76	1.0254
10^{19}	234057667276344607	228576043106974646.13	1.0240
10^{20}	2220819602560918840	2171472409516259138.26	1.0227
10^{21}	21127269486018731928	20680689614440563221.48	1.0216
10^{22}	201467286689315906290	197406582683296285295.97	1.0206

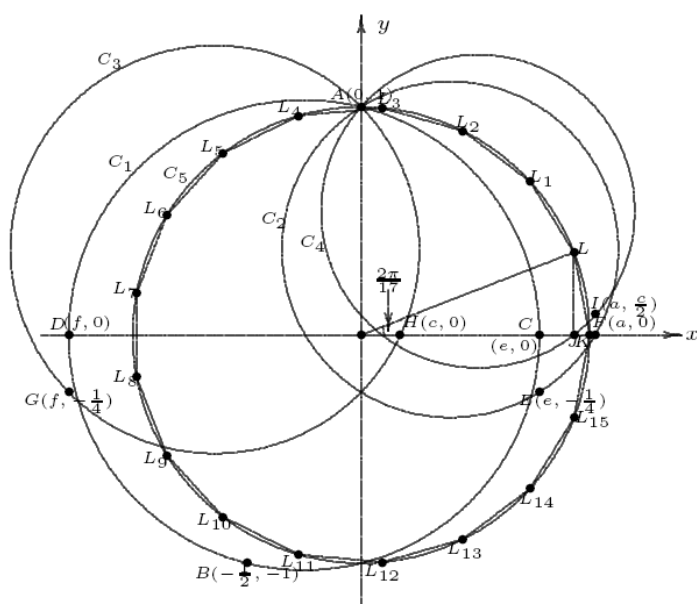
素数定理数值验算表（摘自网上）

后 记

就数学分析这门学科来讲，精华部分基本都包括在内了。就数学来讲，还有许多精彩的方法没有包括进来，比如极端化思想（最大最小原理），复变函数中证明黎曼映射定理就用到了这种思想。其它方法，如奇偶分析法，染色法，矩阵法等等也是很有特色的，可惜不能写进本书。

此前自己整理过好几本书，个别书也贴到网上去，如

高斯经典文章及相关数学工作汇编



不过那终究不是自己写的书，本书是自己真正意义上写的数学书。

写书是梳理思路的好方法，据称某位数学家有个习惯，哪一块内容他不懂，他就苦读这方面文献，然后就此写一本书，这不尝为一个可取的方法。实际上，现代社会很多人都要写分析报告，基本上围绕问题是什么，原因是什么，应该怎么办来展开。本书也采取如下思路。

写书费力不讨好，也是一件很难坚持的事情。但我在清华见过的很多牛人都有整理笔记的习惯，高手之所以是高手，一个好的学习习惯贡献很大。看看牛人们那一堆厚厚的学习笔记就知道了。

这本书公布到网上后，希望能够降低数学的神秘感，给新手们一种启示，给老手们一种借鉴（比如他们可以在本书中挑错），提供一些具有可操作性的方法。至于这些方法是否合理，智者见智，本书就不讨论了。

千里之行，始于足下。祝有缘者得此书者好运，希望能真正帮助你提高数学水平。

数学海洋如市场海洋一样浩瀚，相比之下，我们每一个人都太渺小了。

SCIbird