# 严格化潮流下的积分理论发展

杨睿骐

Qiuzhen College, Tsinghua University

2024 Spring

# 前言

在分析学的概念当中，“积分”无疑是最重要的概念之一，同时它也可以称作一面镜子，从侧面反映着分析学逐步严格化的进程。从最初依赖几何直观的牛顿与莱布尼茨的积分定义，到柯西迈出尝试的脚步而对“连续函数”做出的积分定义，再到黎曼、达布、若尔当、勒贝格对可积函数类的一步步拓宽，标志了分析学逐步脱离几何学与代数学的桎梏，化身成一个独立、严谨、且占有重要地位的学科的进程。在这过程中，有许多数学家穷其一生为此做出贡献，他们犯下一个又一个错误，但是他们用错误堆积成的，是一条通往真理的道路。本文要叙述的便是这段激动人心的历史。

本文分两章内容叙述，主体内容是第二章所述的积分概念的变迁史，而第一章主要叙述的是十九世纪分析学在法国由柯西所引领的严格化进程，作为第二章的内容——积分理论的发展——的一个大背景，主要起铺垫的作用。

同时我要特别感谢我的“分析 1”的授课教师归斌老师。他的课堂中穿插的有关数学史的部分很大程度上激起了我的兴趣，事实上我撰写本文的想法也是萌生于一次普通的“分析 1”课堂上，也就是“积分”一节的第一课。虽然这堂课并不是我所上过的关于积分这一概念的第一堂课，但却是印象最深刻的。归斌老师在教学概念的过程中不忘将其发展的历史渊源介绍给我们，使我们懂得许多数学发展的规律与哲理。在以往的学习经历中我们都是先学习黎曼积分，再学习傅里叶级数以至偏微分方程，但归斌老师在课堂上介绍由黎曼积分到勒贝格积分的转变的时候提到，实际上是先有偏微分方程问题引出傅里叶级数，然后人们才开始考虑积分的定义问题的，这颠覆了我的原有认知，也促使了这篇文章的出现。最后也感谢归斌老师在选题与参考文献方面给予我的帮助。

本文的主要参考文献是：Hans Niels Jahnke 所著 *A History of Analysis* 一书中：第六章，“The Foundation of Analysis in the 19th Century”，作者 Jesper Lützen；第九章，“Theory of Measure and Integration from Riemann to Lebesgue”，作者 Thomas Hochkirchen；以及 Thomas Hawkins 所著 *Lebesgue's Theory of Integration——It's Origins and Development* 的第二章与第三章；以及“分析 1”课程讲义：https://binguimath.github.io/Files/2023_Analysis.pdf，214-217。本文中所有数学家的图片均出自维基百科。

杨睿骐<br>2024 年 3 月 9 日

# 目录

# 第一章　分析学的严格化进程

## 1.1　对分析学严格化的动机

现代的分析学可以说拥有一个令人还算比较满意的基础，而这一切都源于前人在十九世纪对分析学严格化所作出的努力。严格化的过程历经曲折，因为它的目标是建立一个一致的标准，所以它并不仅仅是简单地澄清几个基本概念或者修正两三个基本定理的证明就能够告捷的，它会“牵一发而动全身”，从细节出发慢慢渗透入分析学的所有领域。严格化的过程甚至可以被称作创新的过程，在这个过程中还会伴随着数学中全新的领域的产生，这样的例子比比皆是，分析学的一个重要的基础——点集拓扑，就延伸出了许多全新的概念：逐点收敛、一致收敛、一致连续、紧性、完备性等等。

在当时，对分析学的严格化并不被认为是最亟待解决的问题，大部分数学家们的工作方向都集中在将前人得到的分析理论运用到一些技术性的问题上。值得一提的是，**数学理论中新的、技术性的定理的发展往往促使人们意识到巩固这个理论的基础的重要性**，从而为它的严格化贡献关键的动机。从这个角度来看，傅里叶级数这个概念对于分析的严格化尤为重要，因为它对分析学原有的概念例如函数、积分、收敛性、连续性等提出了质疑与挑战。同样地，微分方程、位势论、椭圆函数等领域也为分析学严格化的动机做出了贡献。

授课也是对分析学严格化的一个主要的动机。一些数学家们发现，当他们为学生上课介绍分析学时，往往因为一些最基础定义的叙述的模棱两可陷入一种尴尬两难的境地，因此他们决定寻求更加严格的叙述。这便是柯西与魏尔斯特拉斯的改革，以及戴德金与梅雷等人对实数的构造的直接原因。数学从科学中逐渐解放的进程也让人们愈发认识到，分析学的基础需要做出修正。在十八世纪甚至十九世纪初的法国，分析学的发展与理论物理的发展紧密相连。理论物理的研究需要频繁地使用分析学上的结论，有些结论在尚未被严谨地证明的情况下就已经被应用而推出一些优美的结果。这些物理学中优美的结果就反过来为一些分析学结论的正确性提供佐证。具体例如物理意义上一些微分方程的解以及级数的和。但当到了十九世纪上半叶，尤其是在德国，大学逐渐成

为数学研究的的中心。结合当时的新人文主义运动，便促使纯数学，作为一个独立的领域，发展起来。所以，为包括分析学在内的数学建立一个独立于应用的坚实的理论基础就变得重要了。

同时，分析学从几何学中的逐渐分离也推进了它的严格化。自欧几里得开始，几何学一直以来被认为是数学中基础最牢固的部分，即使后来自然数的概念被扩展到有理数和超越数，大多数数学家仍然认为欧几里得的理论大体是正确的，从而对其中数的概念的扩展而辩解。但是在十九世纪这一风向发生了改变。人们在欧几里得的观点中找到越来越多的罅隙，非传统的几何学也在撼动着欧几里得的权威。而一直以来，分析学的基础定理都是建立在几何直观上的，人们指出，它需要一个更坚实的基础。一些数学家在试图证明中值定理时便意识到了这一点。

在十八世纪，一些数学家们试图把分析的基础建立在代数而不是几何上。而这一道路在十九世纪便被抛弃了。取而代之的是由自然数和算数学建立的坚实基础。在 1870 年左右，人们高喊“算术化”的口号，从自然数构造有理数再到实数与复数，分析学的基础就由这些完全绕开几何的定义中奠定了。即使帕斯（Pasch）、皮耶里（Pieri）与希尔伯特等人几乎在同一时期建立了几何学的公理化基础，几何学也再也没重拾其在分析学的基础中的地位。恰恰相反的是，希尔伯特证明了，如果算数学是一致的，那么几何学是一致的。

我们可以将分析学的严格化大致分为两个时期：由柯西引领的以法国为中心的时期，与由魏尔斯特拉斯引领的以德国为中心的时期。这也从侧面反映了十九世纪普遍认同的发展方向，即在十九世纪上半叶由法国引领，而在下半叶由德国引领。本文主要着眼于第一个时期，也就是由柯西所引领的时期。

## 1.2 柯西及其对分析基本概念的定义

### 1.2.1 函数

对“函数”的概念的变革，可以认为是分析学严格化的过程中的一条线索，在十八世纪，分析学的严格化仍然处于“几何”到“代数”的转变，所以人们对于函数的理解也是处于由“曲线”到“解析式”的阶段，当时普遍认同的定义出自于欧拉：

**定义 1.2.1** (欧拉). 关于一个变量的**函数**是一个**解析表达式**，它由变量、数字或者常量以任何方式组合而成。

图 1.1: 欧拉

在现代人的思维之下去理解，这个定义多少有些含糊，但以当时的人们的思维，实际上它就定义了所有的“解析函数”，即可以被它的泰勒级数逼近的函数。也就是说，当时的人们认为所有的函数，或者说“好”的函数都是可以被它的泰勒级数逼近的。一切似乎都运转地非常和谐，直到十九世纪，人们开始系统地研究偏微分方程。

最简单的例子是**波方程**：

$$\partial_x^2 f(x,t) - \partial_t^2 f(x,t) = 0 \tag{1.2.1}$$

这个方程描述了一根弦上的振动：在 $t$ 时刻这根弦的形状就是函数 $x \in [a,b] \mapsto f(x,t) \in \mathbb{R}$ 的图像。（这里我们假设弦的两个端点就是 $(a,0)$ 与 $(b,0)$）

利用变量代换：

$$u = x+t, v = x-t \tag{1.2.2}$$

我们得到

$$\partial_x^2 - \partial_t^2 = 4\partial_u\partial_v \tag{1.2.3}$$

由此我们不难得到波方程的通解：

$$f(x,t) = \frac{1}{2}(g(x+t)+g(x-t)) + \frac{1}{2}\int_{x-t}^{x+t} h(s)\mathrm{d}s \tag{1.2.4}$$

此处 $g(x) = f(x,0), h(x) = \partial_t f(x,0)$，特别地，如果我们假设 $h=0$（即在 $t=0$ 时弦不发生振动），那么解就形如：

$$f(x,t) = \frac{1}{2}(g(x+t)+g(x-t)) \tag{1.2.5}$$

如果我们假设弦的两端被固定在 $(0,0)$ 和 $(1,0)$ 两个点上，那么我们就有 $f(0,t)=f(1,t)=0$，对任意 $t\in\mathbb{R}$，将条件转移到 $g$ 上，我们就有 $g(t)=-g(-t)=-g(2-t)$，这表明 $g$ 是一个以 2 为周期的周期函数。

周期函数这个概念，在当时函数的定义下其实是不被接受的。我们都知道，如果 $f$ 是一个解析函数的话，那么 $f(x)=\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(x-a)^n$，所以，$f$ 的取值就被 $f(a),f'(a),f''(a),\cdots$ 所决定了，从而就被 $f|_I$（其中 $I$ 为任意非空开区间）决定了。但周期函数不能被其在一个小区间上的取值所决定，例如一个在 $[0,1]$ 区间上拥有解析表达式 $f(x)=x$ 的函数，它在实轴上也应当拥有同样的解析表达式，所以它就一定不能是周期函数 $f(x)=x-n,\forall x\in(n,n+1]$。退一万步讲，周期函数的导数甚至可以有不连续点。

这些发现足以促使人们对函数的概念进行“清洗”了。

历史上有许多数学家给出了他们对于函数概念的定义。柯西的概念影响力较大，包括傅里叶等人在内，他们都站在柯西一边，希望摆脱所谓“表达式”的桎梏，但是摸着良心讲，他们都或多或少在证明一些定理的时候，在心中还是假装这些“函数”是有一个表达式的，无论是方便思考还是何种原因。随着时间流逝，狄利克雷在这一层面上迈出了至关重要的一步，他对于一类抽象定义的函数，证明了傅里叶级数的收敛性，这才可谓摆脱了表达式的困扰。因此在很长一段时间，“现代”的函数定义是以狄利克雷的名字命名的。

下面我们介绍当时的柯西体系中的分析概念的定义：

### 1.2.2 变量、极限、无穷小量

**定义 1.2.2.** 一个**变量**是指可以逐次取到若干个不同的值的量；相反地，一个**常量**是指只能取到一个确定的值的量。

**定义 1.2.3.** 当同一变量逐次所取的值无限趋近于一个固定值，最终使得它的值与该定值之间的差要多小有多小，那么最后这个定值就称为其他所有值的**极限**。

**定义 1.2.4.** 当同一变量逐次所取的值的绝对值无限减小，使得比任意给定正数的值还要小，这个变量就被称作**无穷小量**。

图 1.2: 柯西

**定义 1.2.5** (柯西). 对于这样一些互相联系的变量：只要知道了其中特定一个变量的值，那么我们就能得到其他所有变量的值，人们通常想象这些变量是用其中那个特定的变量来表达的，此时，这个量就叫做**自变量**，而用这个自变量来表达的其他的变量就被称作这个自变量的**函数**。

### 1.2.3 连续性

**定义 1.2.6** (柯西). 令 $f(x)$ 为一个关于 $x$ 的函数，假设对于每一个在两个给定值之间的 $x$，这个函数总是对应唯一的有限值。若我们首先取一个在这两个定值之间的 $x$，并对其施加一个无穷小的增加量 $\alpha$，那么这个函数将增加

$$f(x+\alpha)-f(x) \tag{1.2.6}$$

式 (1.2.6) 是一个同时关于 $\alpha$ 和 $x$ 的变量。若对每一个在这两个给定值之间的 $x$，式 (1.2.6) 的绝对值随着 $\alpha$ 的减小无限趋近于 0，那么我们称函数 $f(x)$，在这两个给定值之间，为一个关于 $x$ 的**连续函数**。我们可以进一步地定义函数 $f(x)$ **在** $x$ **的领域中连续**，只要该函数在一个包含 $x$ 的两个定值之间连续，而无论这两个定值之间的距离有多近。

在柯西所著《分析教程》中最独特，或许也是最核心的定义莫过于他对于函数连续性的定义了 (见定义 1.2.6), 柯西对连续性的定义与当时广泛认可的欧拉的定义有着显著的差异：

**定义 1.2.7** (欧拉). 一条曲线为连续的，当且仅当它的“本质 (nature)”可以被一个关于 $x$ 的函数所表达，同时如果一条曲线它的不同部分的“本质”需要用到不同的函数来表达，那么这条曲线就是间断的。

通过对比不难发现，欧拉的版本本质上是代数的、全局的，而柯西的版本则在本质上，用现代的语言来说，是“拓扑”的、局部的。柯西从全局迈向局部的这一步与他对“代数的一般性”的反对是不无关系的。但欧拉的定义在任何接受傅里叶想法的人看来都是可疑的，因为事实上，一个欧拉不连续函数 (例如 $|x|$) 的傅里叶级数展开就给出了一个“解析表达式”(用欧拉的话来说):

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\alpha)d\alpha \\ &\quad + \frac{1}{\pi}\sum_{n=1}^{\infty}\left(\cos nx\int_{-\pi}^{\pi} f(\alpha)\cos(n\alpha)d\alpha + \sin nx\int_{-\pi}^{\pi} f(\alpha)\sin(n\alpha)d\alpha\right) \end{aligned} \tag{1.2.7}$$

后来柯西自己也给出了一个例子:

$$f(x) = \begin{cases} x & \text{for} \quad x \leqslant 0 \\ -x & \text{for} \quad x \geqslant 0 \end{cases} = \sqrt{x^2} = \frac{2}{\pi}\int_0^{\infty} \frac{x^2 dt}{t^2 + x^2} \tag{1.2.8}$$

这表明，一个简单的记号的改变或许就能将一个欧拉连续函数变为一个欧拉不连续函数。所以，除非我们承认一个函数的连续性依赖于写下它的方式，否则我们就必须指出欧拉的概念是有歧义的。傅里叶本人并没有在这条道路上继续发展下去，但柯西选择继续深研。那么柯西是从哪里得到他的非传统的定义呢？

法国数学家阿博加斯特（Arbogast）在对偏微分方程解的性质进行偏几何意义上的研究时，已经研究了具有“跳跃”的函数，并将其命名为“不连续函数”(discontiguous)。此外，欧拉的连续性定义虽然广为流传，但其实并没有得到一致的使用。例如，拉格朗日对连续函数的描述方式就不同，类似于柯西后来的定义。在其于 1797 年的论文中，他试图证明，对于充分小的 $h$，函数 $f(x+h)-f(x)$ 的幂级数展开满足其任意一项都大于后面所有项之和这一性质。由此拉格朗日给出了其关于连续性的定义，并且，傅里叶也曾使用过柯西定义中的一些性质。

在柯西本人对于定积分的研究中，他还发现，连续性在微积分基本定理的证明中也非常重要，函数有没有“跳跃”看起来会比所谓的欧拉连续性更接近问题的本质。

拉格朗日的连续性是定义在每一个点上的，但柯西在《分析教程》中并没有这样定义，柯西将不连续性看作在点上的性质，而将连续性视为区间上的性质，这些也都体现了柯西对连续性这一概念的哲学意义上的理解。

### 1.2.4 级数的收敛性，与连续性的关系

**定义 1.2.8.** 一个**级数**是一个由变量构成的无穷序列

$$u_0, u_1, u_2, u_3, \text{etc}\ldots\ldots \tag{1.2.9}$$

它们互相之间以一个特定的规律联系起来。这些变量本身被视作是该级数的不同项。令

$$s_n = u_0 + u_1 + u_2 + \cdots + u_{n-1} \tag{1.2.10}$$

为前 $n$ 项的和，$n$ 为任意正整数。若部分和 $s_n$ 在 $n$ 趋于无穷时趋近于一个确定的极限 $S$，那么称该级数**收敛**，并将该极限称作该级数的**和**。相反地，如果在 $n$ 趋于无穷时该级数没有极限，那么称该级数**发散**，此时认为该级数的和不存在。

在十七世纪，级数"收敛"的概念最初有多种含义，一种是级数的项趋于零，另一种则是部分和 $s_n$ 趋近于一个固定的极限。当柯西在他的《分析教程》中选择后一种定义时，他并不是特别具有革命性。柯西的亮点在于，他在几个证明中相当严格地使用了 $\epsilon - N$ 语言来刻画这一收敛性，并且特别地，在于他在教科书中对发散级数没有和这一观点的一贯认知。在十八世纪，数学家们可以自由地使用发散级数，欧拉甚至试图正式定义它们的和。所以柯西很清楚，当他声称发散级数没有和时，他将会震惊整个数学界。

这一基本主张使得柯西有必要在试图找到级数的和之前首先验证级数的收敛性。为此，他证明了几个收敛性的判定法则。第一个也是最基本的一个是著名的柯西准则。他首先明确指出，如果一个级数收敛，它的第 $n$ 项必须随着 $n$ 的增大收敛到零，

> ……但这个条件并不充分，同时我们还一定有，随着 $n$ 不断增大，如下不同的和式：
>
> $$\begin{aligned} &u_n + u_{n+1}, \\ &u_n + u_{n+1} + u_{n+2}, \\ &\text{etc} \cdots \end{aligned}$$
>
> 即，下述项
>
> $$u_n, \quad u_{n+1}, \quad u_{n+2}, \quad \text{etc} \cdots$$
>
> 从第一项开始往后选取任意有限项之和，都将会小于一个固定值。反过来，如果这些条件满足了，那么我们就可以确保级数的收敛性。

从而柯西得出，收敛级数一定是柯西级数（即部分和序列是一个柯西序列），但当考虑逆问题，即柯西级数是否一定收敛的问题时，他只能摊开两手无能为力。在现代的处理方式中，这一逆问题是靠实数的完备性解决的（甚至后来这被用作是空间完备性的定义），它必须要么被假定为公理，要么从实数的构造中获得。但由于当时分析学公理尚未完善，柯西的著作中没有包含任何一条出路。这种对完备性的缺失是一个普遍的空白，在柯西的分析体系中，特别是在他对中值定理的证明和对连续函数积分存在性的证明中，这种缺失都有所体现。

柯西对收敛性判定法则的处理在其他方面堪称典范。他没有明确说明比较判别法，但他将其用于特定序列，并在这种情况下通过运用柯西准则证明了这一点。然后，他通过与几何级数的类比来证明了根判别法。最后，他从根判别法推导出了比值判别法。他还建立了其他判别方法。这些收敛性判别法并不是第一次出现，许多数学家们已经在使用它们了，但柯西给了它们一个严格的证明，强调了它们在收敛判别方面的基石作用。

柯西的分析观点中最著名的问题就是如下联系收敛性与连续性的*柯西定理*：

**定理 1.2.9.** *对于一个函数项级数，如果它的每一项都在一个定值的邻域上连续，并且级数在这个邻域上收敛，那么收敛到的函数也在这一特定的邻域上连续。*

这一"柯西定理"实际上需要函数列的一致收敛性才能得以成立，而柯西并没有强调这一概念，尽管当时许多数学家试图为柯西辩解，但柯西本人后来承认，自己确实有考虑的疏忽。而柯西的这个定理虽然是有谬误的，但是却引出一个非常重要的问题，也就是一致收敛性的问题，这个问题我们将在后文中详细展开。

### 1.2.5 积分，连续函数可积性的证明

**定义 1.2.10.** 设函数 $y = f(x)$ 关于 $x$ 在两定值 $x = x_0, x = X$ 之间连续，我们将 $x$ 在这两点之间取到的新值记作 $x_1, x_2, ..., x_{n-1}$ 并且假设它们是递增的。我们可以利用这些值将差值 $X - x_0$ 分割成一些小元素

$$x_1 - x_0,\ x_2 - x_1,\ x_3 - x_2,\ \cdots,\ X - x_{n-1} \tag{1.2.11}$$

并且保证它们都拥有相同的符号。这一步过后，我们将每个小元素乘上该元素的左端点对应的 $f$ 的函数值，即让 $x_1 - x_0$ 乘上 $f(x_0)$, 让 $x_2 - x_1$ 乘上 $f(x_1)$, ..., 最后让 $X - x_{n-1}$ 乘上 $f(x_{n-1})$；并且令

$$S = (x_1 - x_0)f(x_0) + (x_2 - x_1)f(x_1) + \cdots + (X - x_{n-1})f(x_{n-1}) \tag{1.2.12}$$

表示这样获得的乘积的和。那么 $S$ 的值自然会受两方面的影响，首先是用来分割 $X-x_0$ 的元素的数量 $n$，其次是这些元素的值，或说，采用的分割方式。

一个重要的观察是，如果这些小元素的绝对值变得很小，并且 $n$ 变得很大的时候，采用不同分割方式所带来的 $S$ 的差异也将变得逐渐可以被忽略。因此，当这些元素变得无穷小的时候，分割方式对 $S$ 的影响将非常细微，此时如果我们进一步让小元素的绝对值随着它们数量的增加而减少的话，$S$ 的值最终会变为一个常数。换句话说，它最后会变成一个确定的极限，并且只依赖于函数 $f(x)$ 与变量 $x$ 的两个界 $x_0, X$。这个极限就是我们所说的**定积分**。

柯西在他对定积分的定义上彻底与他的前辈们决裂开来 (见定义 1.2.10)，莱布尼茨曾将积分考虑为无穷小量的和，不过自伯努利家族对积分理论的推进以来，人们便习惯于将积分视作求微分的逆过程，这使得不定积分变为最初级的概念，并且使得积分计算的过程成为微分计算的附属品。傅里叶是第一个对此提出异议的人。他意识到，为了计算任意函数的傅里叶系数，人们不能再依赖于求微分的计算了，因为对不能给出解析表达式的函数进行微分是不一定是有意义的。因此他将目光集中在定积分 $\int_a^b f(x)dx$ 上（将积分的上下限放在积分符号的上标和下标上事实上是傅里叶的主意）并且强调它表示该曲线与 $x$ 轴之间的部分的面积。

柯西在考虑定积分的时候跟随了傅里叶的脚步，不同的是，柯西将定积分定义为一种“左和”的极限，而没有像傅里叶一样依赖于对于“面积”的模糊的定义。柯西的定义足够清晰、明确，使得他能够“证明”对于一个连续函数，他的积分存在。

### 柯西的“证明”

柯西的证明堪称一个杰作：首先对区间 $[a, b]$ 的分划

$$a < x_1 < x_2 < \cdots < x_{n-1} < b \tag{1.2.13}$$

柯西定义了如下“左和”：

$$S_1 = (x_1 - a)f(a) + (x_2 - x_1)f(x_1) + \cdots + (b - x_{n-1})f(x_{n-1}). \tag{1.2.14}$$

根据柯西在其著作《分析教程》中记录的一个定理，左和等于 $(b-a)M$，此处 $M$ 是在 $f(a), f(x_1), \cdots, f(x_{n-1})$ 之间的一个中间值。由于 $f$ 是连续的，由中值定理，一定有：

$$M = f(a + \theta(b-a)), \quad 0 \leqslant \theta \leqslant 1. \tag{1.2.15}$$

从而：

$$S_1 = (b-a)f(a+\theta(b-a)), \tag{1.2.16}$$

下一步柯西对分划进行加细，并考虑新的左和 $S_2$ 与 $S_1$ 的关系。由区间 $[a, x_1]$ 中小区间对左和 $S_2$ 的贡献为:

$$(x_1-a)f(a+\theta_0(x_1-a)), \quad 0 \leqslant \theta_0 \leqslant 1. \tag{1.2.17}$$

从而应用同样的结论，我们可以得到：

$$\begin{aligned} S_2 = &(x_1-a)f(a+\theta_0(x_1-a)) \\ &+(x_2-x_1)f(x_1+\theta_1(x_2-x_1)) \\ &+\cdots+(b-x_{n-1})f(x_{n-1}++\theta_{n-1}(b-x_{n-1})). \end{aligned} \tag{1.2.18}$$

记：

$$\begin{aligned} &f(a+\theta_0(x_1-a)) = f(a) \pm \epsilon_0, \\ &f(x_1+\theta_1(x_2-x_1)) = f(x_1) \pm \epsilon_1, \\ &\qquad\qquad\vdots \\ &f(x_{n-1}++\theta_{n-1}(b-x_{n-1})) = f(x_{n-1}) \pm \epsilon_{n-1}, \end{aligned} \tag{1.2.19}$$

于是柯西将 $S_2$ 重写为

$$\begin{aligned} S_2 &= (x_1-a)f(a)+(x_2-x_1)f(x_1)+\cdots+(b-x_{n-1})f(x_{n-1}) \\ &\pm \epsilon_0(x_1-a) \pm \epsilon_1(x_2-x_1) \pm \cdots \pm \epsilon_{n-1}(b-x_{n-1}) \\ &= S_1+(b-a)M(\pm\epsilon_0, \pm\epsilon_1, \cdots, \pm\epsilon_{n-1}), \end{aligned} \tag{1.2.20}$$

其中 $M(\pm\epsilon_0, \pm\epsilon_1, \cdots, \pm\epsilon_{n-1})$ 是 $\pm\epsilon_i$ 之间的一个中间值。柯西继续道：

> 我们不妨假设分划区间 $x_1-a, x_2-x_1, \cdots, b-x_{n-1}$ 的长度非常小，从而每个 $\epsilon_i$ 都会非常接近于 0，从而和式中的项 $(b-a)M(\pm\epsilon_0, \pm\epsilon_1, \cdots, \pm\epsilon_{n-1})$ 也会非常小。由此可见，当我们把一个原本非常细的划分再加细的时候，和式 $S$ 的值将不会有显著的改变。

接着，柯西通过构造两个划分的公共加细来比较两个划分对应的左和。由此他得出，当分划变得越来越细的时候，左和之间的差距会越来越小，从而左和本身会趋于一个定值，这个定值就被定义成定积分。我们注意到，最后一步实际上使用了空间完备性的假设；另一个需要强调的点是，柯西在得到对于细的分划，$M(\pm\epsilon_0, \pm\epsilon_1, \cdots, \pm\epsilon_{n-1})$ 的值非常小的结论的时候，实际上将连续性的条件当作了一致连续性来使用。

柯西还证明了微积分基本定理，从而将自己所定义的新的积分概念同牛顿与莱布尼茨的旧积分概念联系了起来。

为什么柯西要像这样修改积分概念的定义呢？他又是从哪里得到灵感的？我们在第 1.2.3 节提到，柯西之前已经发现定积分的值可能会与原函数端点值之差不一样，并且他与泊松（Poisson）都发现，复数域上的积分的值可能会依赖“积分路径”的选取。种种迹象或许都表明，积分的逆过程看起来并不是一个好的积分定义方式。欧拉及其同时代的数学家已经开始用左和来逼近积分了。而拉克鲁瓦（Lacroix）与泊松则致力于证明左和趋近于积分的值。我们可以在他们的论文中发现许多柯西的命题的前身。很有可能柯西的讨论就建立在这些来源上。但柯西的处理方式清晰得多，唯一需要的是一个清晰的对于连续性的定义；而前人的结论则不加论述地使用了 $f$ 的一二阶导数的存在性与函数的单调性。并且，最重要的是，柯西将这一看待方式从一个数值逼近过程上升到了积分的定义层面。

### 1.2.6 柯西概念间的紧密结构

从柯西对于二项式定理

$$(1+x)^{\mu} = 1 + \frac{\mu}{1}x + \frac{(\mu)(\mu-1)}{2!}x^2 + \cdots \tag{1.2.21}$$

的证明中我们可以直观地感受到柯西的概念之间的紧密的结构。二项式定理是所谓“代数分析”的重要基础之一，另外一个则是代数基本定理。接下来我们欣赏一下柯西对二项式定理的证明：

首先由比值判别法可知该级数在区间 $(-1,1)$ 上收敛。对于给定的 $x$，定义该级数的和为 $\varphi(\mu)$。接下来只需要证明 $\varphi(\mu) = (1+x)^{\mu}$ 即可。由此我们来研究这样一个简洁的函数方程:

$$\psi(\mu)\cdot\psi(\mu\prime) = \psi(\mu+\mu\prime). \tag{1.2.22}$$

在柯西所著《分析教程》的第五章中，他证明了该方程的连续解全部形如 $\psi(\mu) = A^{\mu}$(此处 $A$ 为常数)。他的证明成为了一种典范。从式 (1.2.22) 中柯西首先直接得出，对任意 $m \in \mathbb{N}$

$$\psi(m) = \psi(1)^m \tag{1.2.23}$$

随后，同样直接地，对任意 $m \in \mathbb{Q}_+$，该式也成立。而由于任意实数都可以表示为有理数的极限，而函数 $\psi(1)^{\mu}$ 关于 $\mu$ 是一个连续函数，于是柯西得到

$$\psi(\mu) = \psi(1)^{\mu} \tag{1.2.24}$$

对任意 $\mu \in \mathbb{R}_+$ 都成立，从而它对任意 $\mu \in \mathbb{R}$ 也成立。从有理数过渡到实数的一步是最关键的。也就是在这一步，柯西对于连续性的定义成为了至关重要的概念。尽管从欧拉开始，像这样的函数方程已经被许许多多数学家研究过，但柯西是第一个用令人满意的方式处理从有理数过渡到实数的过程的。

接着柯西证明了式 (1.2.21) 的右边是满足式 (1.2.22) 的，这是柯西关于两个级数的乘法的一个定理的直接应用：

**定理 1.2.11.** 令 $u_0, u_1, \cdots$，与 $v_0, v_1, \cdots$ 代表两个绝对收敛的级数，分别收敛到 $s, s'$，那么级数

$$\left(\sum_{i=0}^{k} u_i v_{k-i}\right)_{k\in\mathbb{Z}_{\geqslant 0}} \tag{1.2.25}$$

是一个收敛到 $s, s'$ 的级数。

由于 $\psi(\mu)$ 是一列关于 $\mu$ 的连续函数构成的级数的和。柯西定理表明，$\psi(\mu)$ 也是连续的。除了在幂级数上的应用之外，这是柯西使用有问题的柯西定理的唯一一处关键位置。于是由之前的论述可得，$\psi(\mu)$ 必须等于 $\psi(1)^\mu$ 或者说 $(1+x)^\mu$。这就完成了柯西对于二项式定理的证明。

虽然柯西的概念有着空前的影响，已经相对来说紧密的结构，但是其中也不乏存在一些问题。傅里叶级数的收敛性尚无从下手，“柯西定理”的谬误也让一致收敛的问题变得突出。在这些问题之下，人们开始对积分的概念进行革新，以解决当时横亘在分析学天空上的主要问题。这也是下一节我们主要讨论的方向。

# 第二章　积分理论的发展

## 2.1　黎曼积分的诞生与发展

### 2.1.1　进一步发展积分理论的动机

现在的我们很容易就能理解，严格化定义定积分 $\int_a^b f(x)\mathrm{d}x$ 的重要性。我们最初所学习的积分的定义就是黎曼积分，也就是将积分定义为黎曼和的极限，而黎曼和直观地代表着一些长方形面积的和。我们觉得这个定义是再自然不过的了，因为我们知道，积分可以直观地理解为函数图像与 $x$ 轴之间的区域的面积，并且这个面积可以用一些长方形的面积的和来逼近。但是在历史上，黎曼积分的出现晚于牛顿与莱布尼茨所定义的积分上百年之久，为什么过了如此之久，人们才发觉原本对于积分的定义是不够严密的呢？

一切的源头实际上是对于“函数”的定义的革新。在十八世纪，“函数”的公认定义出自于欧拉（见定义 1.2.1），那时的人们普遍认为函数可以被其泰勒级数所逼近，从而我们就可以很方便地利用导数的逆运算来定义积分，即对

$$f(x)=\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(x-a)^n \tag{2.1.1}$$

定义

$$F(x)=\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{(n+1)!}(x-a)^{n+1} \tag{2.1.2}$$

并且定义

$$\int_a^b f(x)=F(b)-F(a) \tag{2.1.3}$$

虽然那个时候的人们已经知道了积分与面积之间的联系。但是那个时候，人们认为这只是积分理论的一个应用而已，并没有以此来作为积分的定义，因为这么做会让积分的定义变得非常复杂。

在第 1.2.1 节中我们提到，偏微分方程理论的发展促使了数学家们对“函数”的概念进行了一番清洗。此番清洗过后，原先的积分定义也就站不住脚了。事实上，偏微分方程在傅里叶级数的引入和黎曼积分的定义中也扮演了一个比较重要的角色。

#### 偏微分方程的推动


图 2.1: 傅里叶

傅里叶级数的收敛性便是傅里叶在研究**热方程**

$$\partial_t f(x,t) - \partial_x^2 f(x,t) = 0 \tag{2.1.4}$$

时遇到的问题。此处 $x$ 在一个闭区间（设为 $[-\pi,\pi]$）上定义，代表一根细长的棍子。而 $f(x,t)$ 则代表 $t$ 时刻下棍子上的 $x$ 位置处的温度。傅里叶在解热方程的时候运用的主要是**分离变量**的方法：首先假设 $f(x,t)=u(x)v(t)$，此时热方程表明 $u''(x)v(t)=u(x)v'(t)$，从而

$$\frac{u''(x)}{u(x)} = \frac{v'(t)}{v(t)} \tag{2.1.5}$$

式 (2.1.5) 的左侧与 $t$ 无关，右侧与 $x$ 无关。从而式 (2.1.5) 的两侧都应该等于一个常数 $-\lambda$。所以此时方程的解就应该形如

$$u(x) = e^{\mathbf{i}\sqrt{\lambda}x}, v(t) = e^{-\lambda t} \tag{2.1.6}$$

如果进一步假设棍子的两端点 $-\pi, \pi$ 处的温度相等，那么 $\sqrt{\lambda}$ 必须是一个整数 $n \in \mathbb{Z}$。所以 $f(x,t) = e^{\mathrm{i}nx-n^2t}$。对这个结果取无限维线性组合，傅里叶就得到了通解：

$$f(x,t) = \sum_{n=-\infty}^{+\infty} a_n e^{\mathrm{i}nx-n^2t} \tag{2.1.7}$$

初始温度 $g(x) = f(x,0)$ 是一个**傅里叶级数**：

$$\sum_{n=-\infty}^{+\infty} a_n e^{\mathrm{i}nx} \tag{2.1.8}$$

特别的，这是一个以 $2\pi$ 为周期的周期函数。

我们想要利用式 (2.1.7) 得到在初始条件 $f(x,0) = g(x)$ 下的一般解 $f(x,t)$，我们必须首先由 $g$ 来确定所有傅里叶系数 $a_n$ 的值。事实上，由于

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\mathrm{i}kx} \cdot e^{-\mathrm{i}nx}\mathrm{d}x = \delta_{k,n} \tag{2.1.9}$$

我们不难得到:

$$a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(x)e^{-\mathrm{i}nx}\mathrm{d}x \tag{2.1.10}$$

这样，一个与积分理论的发展有关的本质性的问题便跃然纸上：**式 (2.1.10) 中的积分在 $g$ 不是解析函数的时候，该如何定义呢？**

**积分初探**

很有趣的一点是，人们对可积函数这一概念由连续函数到黎曼可积再到勒贝格可积进行扩展的举动，与研究哪些函数拥有傅里叶展开、哪些函数是“好的”的过程是齐头并进的。

与傅里叶几乎同时代的数学家柯西，给出了这个问题的部分回答，在第 1.2 节中我们提到，柯西在分析学严格化的进程中是一个先驱者，他的观点自成体系，在分析学的发展中建立了新的标准。对连续函数的定义无疑是柯西的体系中分析学的起点，而他对连续性的定义，空前地，并不依赖于函数是否拥有一个具体的解析表达式。这与傅里叶级数遇到的问题几乎不谋而合。

在柯西对于积分的定义（见定义 1.2.10）中，柯西对于区间 $[a,b]$ 的给定的分划 $a = x_0 < x_1 < \cdots < X = b$，柯西定义了如下的和：

$$S = (x_1 - x_0)f(x_0) + (x_2 - x_1)f(x_1) + \cdots + (X - x_{n-1})f(x_{n-1}) \tag{2.1.11}$$

柯西接下来进一步说明了，如果 $f$ 是一个连续函数（见定义 1.2.6），那么只要令分划 $P, P'$ 足够“细”，它们所对应的和 $S, S'$ 之间的差就可以无穷小。因此，对于“连续”函数，定积分 $\int_a^b f(x)\mathrm{d}x$ 就可以被良好地定义成上述“柯西和”的极限。

所以，柯西给出的这个问题的部分回答就是，至少对于那些“连续”的函数 $f$，定积分 $\int_a^b f(x)\mathrm{d}x$ 是良定的。但傅里叶的想法需要一个更一般的答案，就是，对任意的函数 $f$，我们是否可以定义 $\int_a^b f(x)\mathrm{d}x$ 呢？

狄利克雷给出了这个问题的进一步回答。狄利克雷从 1822 年开始学习数学，但因为在那个时候，德国的大学中，除了高斯以外，就没有几个非常出众的数学家了。所以他选择去到巴黎，当时毋庸置疑的世界数学的中心，进行学习。在他的早年，狄利克雷对傅里叶有关热传导的工作进行了研究（傅里叶的文章刚好是狄利克雷来到巴黎的那一年发表的），同时对柯西的著作《分析教程》进行了研读。狄利克雷私下里还同傅里叶有过交流。

图 2.2: 狄利克雷

狄利克雷将这两者的工作融合了起来。发表在了 1829 年的克雷尔杂志（*Creelles Journal*）上。他受傅里叶的启发去研究同样的问题，即在何等假设下，一个函数可以被写成傅里叶展开的形式。不同的是，他采用了柯西的严格化体系进行研究。他证明了，一个在 $(-\pi, \pi)$ 上定义的函数 $f$ 可以被写成傅里叶展开的充分条件是：

> $f$ 单调，只有有限多个不连续点，并且拥有性质 $f(x) = \frac{1}{2}(f(x_-)+f(x_+)), \forall x \in (-\pi, \pi)$。

有趣的是，此处关于函数连续性的条件其实在证明中的作用不大，它唯一的作用便

是保证傅里叶系数的定义式中的积分是有意义的。狄利克雷在 1829 年的文章中构造了一个性质很差的函数，即**狄利克雷函数**：

$$f:\mathbb{R}\to\mathbb{R}\,,\quad f(x)=\begin{cases}0 & x\in\mathbb{Q}\\ 1 & x\in\mathbb{R}\setminus\mathbb{Q}\end{cases} \tag{2.1.12}$$

它确实属于所谓“任意”函数一列，但按照柯西对于积分的定义，这个函数的定积分 $\int_a^b f(x)\mathrm{d}x$ 是没有意义的，因为我们总是既能找到一个任意细的划分 $a=x_0<x_1<\cdots<x_n=b$ 使得所有 $x_i$ 都是有理数（此时柯西和为 0），又能找到另一个所有 $x_i$ 都为无理数的划分（此时柯西和为 1）。狄利克雷深知并不是所有函数的积分都是有意义的，所以在他后来的著作中，他还是谨慎地将其讨论的对象限制在了连续函数上面。

> 连续性的假设到底可以被减弱到什么程度呢？一个可积函数最多能有多少个不连续点呢？我们到底能不能让一个有限区间上的可积函数拥有无穷多个不连续点呢？

1826 年，狄利克雷在为振兴德国自然科学研究而奔走的亚历山大 • 冯 • 洪堡 (Alexander von Humboldt) 的影响下，返回德国，同时也将心中的这个疑问一齐带回。1828 年，在洪堡的帮助下，狄利克雷来到学术气息浓厚的柏林，在柏林军事学院任教，开启了他在柏林长达 27 年的教学与研究生涯。他不知道的是，在若干年后，他会迎来一个非常特别的学生，将为他心中的疑问拨开一层云雾。他就是年轻的波恩哈德 • 黎曼……

### 2.1.2 黎曼积分

黎曼于 1847 年至 1849 年在柏林学习，在 1851 年完成他的博士论文以后，他便直接开始准备其德语国家教授资格考试（Habilitation）的关于三角级数的论文。狄利克雷在 1852 年的秋天留在了哥廷根大学，他在当时还是给予了黎曼不小的支持。在戴德金的文章中曾经引用了黎曼写给其父亲的一封信，信中描述了黎曼希望狄利克雷对其论文提出建议是，狄利克雷的反应：

> 前几天早上，狄利克雷与我共度了大约两个小时，他给了我很多关于我的教授资格考试的完整的建议，让我的工作变得简单多了；如果不是他，有些东西我可能要到图书馆查阅很久才能找到。

图 2.3: 黎曼

在 1853 年的 12 月，黎曼提交了他关于函数的三角级数的可表示性的文章 *Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe*，在其中，他以一种新的方式定义了积分，也就是我们后来所熟知的黎曼积分。他的工作花了很长时间才出名，直到 1868 年，也就是黎曼去世以后两年，这篇文章才被出版。

黎曼的文章分为两个部分，在第一个部分中，黎曼回溯了历史上对于任意（由图像给出的）函数以及它们的三角级数的可表示性的研究，并且对一些尚未被考虑的情况进行了研究。而在两个部分之间，黎曼认为有必要关于定积分的定义以及它何时有意义写一个简短的说明。作为其论文的某种“副产品”，黎曼用五六页的篇幅介绍了他关于积分定义的想法，其中包括黎曼积分的定义，一个可积性的判定准则，以及一个有趣的例子。

### 黎曼积分的定义

所以，$\int_a^b f(x)dx$ 意味着什么？

与柯西一样，黎曼也从一个划分 $a = x_0 < x_1 < \cdots < x_n = b$ 入手，但与柯西不同的是，他构造了一个不一样的和：

$$S = \sum_{i=1}^{n} \delta_i \cdot f(x_{i-1} + \delta_i \epsilon_i), \tag{2.1.13}$$

此处 $\delta_i = x_i - x_{i-1}$ $(i = 1, 2, \cdots, n)$，$\epsilon_i$ 为 0 到 1 之间的有理数。很明显 $S$ 的值取决于划分的选择与 $\epsilon_i$ 的选择。定义了这个和之后，黎曼积分的定义便呼之欲出：

**定义 2.1.1** (黎曼). 如果式 (2.1.13) 中定义的和，无论 $\delta$ 与 $\epsilon$ 怎样选取，只要所有 $\delta$ 趋于无穷小，都无限趋近于一个确定的极限值 $A$，那么这个极限就被称为 $\int_a^b f(x)\mathrm{d}x$。

也就是在这里，黎曼超越了柯西。柯西只能将积分定义在连续函数上，而黎曼引入了一类新的函数——黎曼可积函数。

黎曼的定义中存在着一个问题，就是如果黎曼和有若干个不同的极限的话，表达式 $\int_a^b f(x)\mathrm{d}x$ 也会失去意义。黎曼指出了一个可能出现的例子，即若函数在某一点附近的值趋于无穷大——此时函数是无界的——上述问题就有可能发生。所以黎曼将函数的有界性作为其“严格意义上”可积的必要条件，对于是否还有情况使得出现很多个不同的极限值，黎曼认为它的任意性太大了，因此很难做出一个一般性的适当的定义。

接下来黎曼准备好回答“什么样的函数可积”这个问题了。不过首先让我们研究“在严格意义上”可积的函数（也就是排除在今天人们所说的“反常积分”的情况），此时函数是有界的。为了简化记号，我们采用一些在黎曼的著作中没有出现的稍显现代化的记号：

令偏序集 $P=\{a=x_0<x_1<\cdots<x_n=b\}$ 表示一个区间 $[a,b]$ 的划分，令 $\delta(P)=\max\{(x_i-x_{i-1}):1\leqslant i\leqslant n\}$ 表示 $P$ 的**范数**，进一步地，令 $\mathcal{P}$ 表示区间 $[a,b]$ 的所有划分构成的集合，最后令 $\mathcal{P}_d:=\{P\in\mathcal{P}:\delta(P)<d\}$ 表示所有范数小于 $d$ 的划分构成的集合。

接下来我们可以简要地叙述黎曼的想法了。对于一个给定的划分，黎曼考虑了 $f$ 在区间 $[x_{i-1},x_i]$ 上振荡的最大幅度，也就是“在此区间上函数最大值与最小值的差”，将其记作 $D_i$（这里黎曼的叙述在现代人看来有些不准确，因为他没搞清楚“最大/小值”与“上/下确界”间的区别）。黎曼接着写下了下述等价关系：

$$\int_a^b f(x)\mathrm{d}x \ \text{ exists} \Leftrightarrow \Omega(P):=\sum_{i=1}^{n} D_i(x_i-x_{i-1}) \stackrel{\delta(P)\to 0}{\longrightarrow} 0. \tag{2.1.14}$$

在此情况下，$\Delta(P):=\sup\{\Omega(P):P\in\mathcal{P}_d\}$ 在 $d\to 0$ 时也趋于 0。现在对 $\sigma>0$ 令 $l(P,\sigma)$ 表示在划分 $P\in\mathcal{P}$ 的区间中，振荡幅度 $D_i$ 大于 $\sigma$ 的区间的长度的总和。在此记号下我们有：

$$\sigma l(P,\sigma)\leqslant\Omega(P)\leqslant\Delta(d)\Rightarrow l(P,\sigma)\leqslant\frac{\Delta(d)}{\sigma}. \tag{2.1.15}$$

所以对于一个给定的 $\sigma>0$，适当选取划分 $P$ 让 $d$ 足够小，就可以令 $l(P,\sigma)$ 无限趋近于 0。于是黎曼由此得到了下述定理：

**定理 2.1.2.** *假设在区间 $[a,b]$ 上定义的有界函数 $f$ 是可积的，那么对任意 $\epsilon>0$，对任意 $\sigma>0$，存在一个 $d>0$ 使得对任意划分 $P\in\mathcal{P}_d$，都有 $l(P,\sigma)<\epsilon$ 成立。*

这个定理说的就是，震荡幅度大于 $\sigma$ 的区间总长，无论 $\sigma$ 有多小，都可以通过适当选取划分来将其控制地任意小。事实上，黎曼也证明了这个条件是充分的，黎曼甚至还给出了一个对于有极点（poles）的函数适用的判定准则。

## 黎曼的例子

接下来让我们欣赏黎曼所构造的一个函数。这个函数打破了人们对于连续性与可积性之间的联系的刻板印象，因此引起了当时学术界的轰动。这是一个可积函数，但它具有高度的不连续性。事实上，它在任意两点之间都有着无穷多个不连续点！

记一个实数 $x$ 的小数部分为 $(x)$，我们定义：

$$\Phi(x) := \begin{cases} (x), & (x) < \dfrac{1}{2}, \\ 0, & (x) = \dfrac{1}{2}, \\ (x) - 1, & (x) > \dfrac{1}{2}, \end{cases} \quad x \in \mathbb{R}, \tag{2.1.16}$$

我们得到了一个在 $x \in \{\pm\frac{2k+1}{2} : k \in \mathbb{N}_+\}$ 上不连续且在其余区间均连续的函数 $\Phi$。在不连续点 $x$ 处，$\Phi(x) = 0$，但 $\Phi(x \pm 0) = \mp\frac{1}{2}$。并且我们有 $\Phi(\mathbb{R}) = (-\frac{1}{2}, \frac{1}{2})$，从而 $\Phi$ 有界。

进一步，对 $n \in \mathbb{N}$ 令 $\Phi_n(x) := \Phi(nx)$。因为 $\Phi_n(\pm\frac{2k+1}{2n}) = \Phi(\pm\frac{2k+1}{2})$，所以 $\Phi_n$ 在 $x$ 处不连续当且仅当 $x \in \{\pm\frac{2k+1}{2n} : k \in \mathbb{N}_+\}$。

这些全都定义完毕以后，黎曼将函数 $f$ 定义为：

$$f(x) := \sum_{n=1}^{\infty} \frac{\Phi_n(x)}{n^2}, \quad x \in \mathbb{R} \tag{2.1.17}$$

这个函数在 $x$ 处不连续当且仅当 $x = \frac{p}{2q}$，此处 $\gcd(p, 2q) = 1$，而在它的不连续点 $\frac{p}{2q}$ 处，我们有：

$$f(x \pm 0) - f(x) = \mp\frac{1}{2q^2} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \mp\frac{\pi^2}{16q^2} \tag{2.1.18}$$

从而在任意给定的区间中，$f$ 都有无穷多个不连续点，用现代的语言来说，就是 $f$ 的不连续点在 $\mathbb{R}$ 中稠密。

然而 $f$ 在任意闭区间上都是可积的：因为 $f$ 有界，黎曼对于可积性的判定准则是适用的。$f$ 只在 $\frac{p}{2q}$ 处有跳跃间断点，而我们容易估计它跳跃的高度：

$$f(\frac{p}{2q} - 0) - f(\frac{p}{2q} + 0) = \frac{\pi^2}{8q^2} \tag{2.1.19}$$

所以对于给定的 $\sigma > 0$，区间 $[a, b]$ 中满足 $\frac{\pi^2}{8q^2} > \sigma$ 的跳跃间断点的个数总是有限的。那么对于足够细的划分，这些间断点就可以被拥有任意小的总长度的区间覆盖。

黎曼积分的定义直到 1868 年才发表，但它的确引起了许多数学家的注意。魏尔斯特拉斯曾评价黎曼的定义是他所见过的思考地最一般化的。它的一般性甚至允许有着稠密不连续点集的函数可积。在当时许多数学家认为，黎曼积分或许已经将可积函数的范围扩展到了极致。

### 2.1.3 汉克尔（Hankel）与达布（Darboux）

#### 汉克尔

> "…人们发现，即使在纯粹的几何领域，直觉上认为直接而必然的东西也具有如此大的欺骗性，大到以至于不能再允许它声称自己是科学证明的程度。"
>
> (汉克尔 1870)

图 2.4: 汉克尔

德国数学家赫尔曼·汉克尔（Herman Hankel）是第一批对函数的新定义以及黎曼对于积分的定义做出回应的数学家之一。1860 年，他曾在哥廷根大学进行过一段学习，也就是在那里，他受到了他的学术导师——黎曼——的深刻影响。汉克尔在 1870 年发表了他关于无限振荡与不连续函数的论文 *Untersuchungen über dEe unendlich oft* *oscillirenden und unstetigen Funktionen*，这篇文章很快便拥有了广泛的影响力，并且于 1882 年在《数学年鉴》（*Mathematische Annalen*）中重印，公之于整个数学界。

汉克尔的工作中的一部分对测度与积分理论的发展起到了关键的影响。尽管汉克尔的工作已经在某种意义上显示出向集合论转变的倾向，但在其中包含了一个历史上很有趣的错误，这个错误后来成为了测度理论与思想的发展与被世人接受的一道障碍：即“小”集合在测度理论以及拓扑意义下的定义不够充分。

汉克尔在他的工作中还分析了狄利克雷与黎曼等人对于函数的抽象定义，并将其与最初欧拉的定义做出对比。粗略研究了从旧概念向新概念的转变后，他指出，后者尚不足以满足分析的需要，因为他认为这种函数没有任何一般性质，因此不同位置的函数值之间的所有关系都消失了。由此他认为，分析学中对新的函数概念的定义的第一步就是将有关旧概念的想法移除，并且着重去研究“在两变量之间可能的关系的多样性”，这一点在狄利克雷对函数的定义中就有所体现。

在给定的函数的抽象定义之下，汉克尔首先用现在我们所熟知的方式定义了连续性，然后转而去研究一些比连续性稍弱的条件。经过这样的分析过后，汉克尔提出了他对函数的分类：第一类函数，是所有的连续函数；第二类函数，性质与第一类函数非常相似，是在任意有限区间都只有有限多个不连续点的函数。对于欧拉，傅里叶，柯西等人来说，这两类函数就已经将他们眼中的“函数”类型穷尽了（换句话说，他们认为不属于这两类当中的“函数”已经逾越了分析学所能研究的范围）。但狄利克雷与黎曼等人呼吁人们注意在有限区间拥有无穷多不连续点的函数的可能性，这类函数便被汉克尔放入了第三类函数当中。进一步地，汉克尔将第三类函数又细分为两个小类：点态不连续函数（pointwise discontinuous functions）与完全不连续函数（totally discontinuous functions）。

$$
\text{函数}\begin{cases}
\text{1. 在所有点处均连续} \\[1em]
\text{2. 在任意有限区间都只有有限多个不连续点} \\[1em]
\text{3. 在有限区间拥有无穷多不连续点}\begin{cases}\text{3a. 点态不连续} \\ \text{3b. 完全不连续}\end{cases}
\end{cases}
$$

图 2.5: 汉克尔对函数的分类

汉克尔这样细分的依据来自于狄利克雷与黎曼所强调的可积函数与不可积函数的区别，不过汉克尔在其中添加了一些自己的重要想法，并利用一个函数在某个点 $x$ 处的“跳跃”（jump）的概念来将这两种函数进行区分。在第 2.1.2 节黎曼对定理 2.1.2的证明

中，黎曼就使用过这个概念，但汉克尔希望将其推广到任意函数，而不仅限于那些在 $x$ 处拥有左右极限的函数。为此，汉克尔首先给出了函数 $f$ 在点 $x$ 处的跳跃大于正实数 $\sigma$ 的定义，而后将函数 $f$ 在点 $x$ 处的**跳跃**定义为所有这样的 $\sigma$ 之中的“最大值”（更准确地用现代语言来说，应该是“上确界”），用数学的语言，即：

$$j_f(x) = \sup\{\sigma : \forall\epsilon > 0,\ \exists\delta,\ \text{使得}\ |\delta| < \epsilon, \text{且}\ |f(x+\delta) - f(x)| > \sigma\} \tag{2.1.20}$$

这个记号，就是我们后来所熟知的函数在某点处的**振幅**的雏形。

而后，汉克尔进一步定义

$$S_\sigma = \{x : j_f(x) > \sigma\} \tag{2.1.21}$$

若对一个不连续函数，对任意的 $\sigma$，$S_\sigma$ 中的点都“散落”（scattered，*zerstreut*）在函数的定义域中（即用现在的话来讲，在函数的定义域上无处稠密），那么就称这个函数是点态不连续的。汉克尔将这些函数看作是连接前两类函数（见图 2.5）与剩余的不连续函数之间的桥梁。而后者——完全不连续函数——则被定义为存在某个 $S_\sigma$“填满”（fill up，*erfüllen*）了某个区间（即用现在的话讲，在某区间上稠密）的函数。

进行完此番细分以后，汉克尔相信他已经在分析学所能够处理的函数与它无法处理的函数之间划清了一条界限：他认为，一个函数是黎曼可积的当且仅当它优于点态不连续函数（即属于图 2.5中的第 1、2 及 3a 类）。特别地，汉克尔认为他证明了如下结果：一个函数优于点态不连续函数，当且仅当对任意正实数 $\sigma$，可以找到一族总长度任意小的区间将所有振幅大于 $\sigma$ 的点覆盖住。我们不难看出，这个命题的充分性是正确的：如果对某个 $\sigma$，$S_\sigma$ 在一个区间 $I$ 上稠密，那么任何与 $I$ 相交的区间的振幅都会大于 $\sigma$，因此，在任何对区间 $[a, b]$ 的划分中，振幅大于 $\sigma$ 的区间的总长度总是大于等于区间 $I$ 的长度，因而这个函数不是黎曼可积的。问题出在这个命题的必要性部分，它实际上要求一个无处稠密集是“零测”的，而这个断言是错误的。

不幸的是，汉克尔在发表这篇论文的三年以后，也就是 1873 年便英年早逝，而在他逝世两年以后，第一个上述结论的反例才被举出。对于已经知道问题的答案的人，对错误的答案或许会一笑置之，但我们不妨花一些篇幅，带入汉克尔思考这个问题时的场景，从而了解汉克尔的思维，这或许是有参考的意义和价值的。

### 汉克尔的想法

在讨论汉克尔的“证明”之前，我们不妨了解一下，在汉克尔的脑海里，点态与完全不连续函数这样的抽象概念对应着怎样的具体例子，因为它们实际上反映了汉克尔在

多大程度上理解了他所定义的集合类型的多样性；同时一般来说，这一点在“理解”一段证明的过程中也是尤为重要的。

汉克尔想法的来源首先是对狄利克雷函数的推广。简要起见我们采用稍显现代的记号，令 $B\left(\frac{1}{2^n},\zeta^n\right)$ 表示以 $\frac{1}{2^n}$ 为中心，半径为 $\zeta^n$ 的开球。定义集合

$$S_\zeta = \bigcup_{n=1}^{+\infty} B\left(\frac{1}{2^n},\zeta^n\right) \tag{2.1.22}$$

并且在区间 $[0,1]$ 上定义函数 $f$：

$$f(x) = \begin{cases} 0, & x \in S_\zeta \cap \mathbb{Q} \\ 1, & x \in [0,1] \setminus (S_\zeta \cap \mathbb{Q}) \end{cases} \tag{2.1.23}$$

那么振幅为 1 的区间总长 $s = \zeta + \zeta^2 + \zeta^3 + \cdots = \frac{\zeta}{1-\zeta}$，换句话说，$f$ 是完全不连续函数。下面这个例子对于代入汉克尔关于点态不连续函数的理解来说十分关键：我们定义 $S = \left\{\frac{1}{2^n} : n \in \mathbb{Z}_+\right\}$，并进一步定义函数 $g$：

$$g(x) = \begin{cases} 0, & x \in S \\ 1, & x \in [0,1] \setminus S \end{cases} \tag{2.1.24}$$

那么 $g$ 在 $S$ 这个无穷点集上有着高度为 1 的跳跃，但是与 $f$ 不一样的是，没有任何一个区间会被不连续点“填满”，我们总是可以选取总长任意小的区间把这些点覆盖住：我们取长度为 $\epsilon^n$ 的区间将 $\frac{1}{2^n}$ 这个点覆盖住，那么区间的总长即为 $s = \frac{\epsilon}{1-\epsilon}$, 在 $\epsilon$ 趋于无穷小的时候，$s$ 也趋于无穷小。

我们把这个例子放在心中，现在来考虑汉克尔对于如下命题的“证明”：

**命题 2.1.3.** *若 $S_\sigma$ 无处稠密，那么它可以被总长任意小的区间覆盖。*

> *证明（汉克尔）.* 只需考虑 $S_\sigma$ 有无穷多个元素的情况。考虑如下过程：首先将函数的定义域划分成一些区间，其中每个跳跃大于 $\sigma$ 的不连续点恰好被其中一个区间覆盖，并假定选取了足够大的区间，使得它们填满了整个定义域。再将每个区间分成 $n$ 等分，将其压缩成其中包含那个对应的不连续点的那一份，那么它们的长度之和就被减小到了定义域长度的 $\frac{1}{n}$，选取足够大的 $n$，则这些区间的总长就可以任意小。 □

这个证明或许有许多不太明确的地方，让我们看看另一个汉克尔的错误示范，来搞明白汉克尔到底是哪里想错了。汉克尔“证明”了一个点态不连续函数的左右极限一定存在，他的证明的前几行是这样的：

证明（汉克尔）. 固定一点 $x=a$，在此点处的跳跃大于 $\sigma$; 接下来我们找到下一个跳跃大于 $\sigma$ 的点 $x=a+h$；那么在这两点之间的区间中，振幅将会小于 $2\sigma\cdots$ □

在这里我们可以更明显地看出，汉克尔实际上所理解的“散落”的集合，实际上与他真正所定义的东西大相径庭。在汉克尔的心里，“散落”的集合没有任何极限点。并且，他后半部分的证明可以推出，对于 $\sigma'<\sigma$，$S_\sigma$ 中的点不能是 $S'_\sigma$ 的极限点。所有这些都指向一个原因，那就是，汉克尔错误地将 $\{\frac{1}{2^n}\}$ 这个集合当作了无处稠密集的一个代表，以至于认为所有无处稠密集都应该拥有与这个集合类似的性质。

总的来说，汉克尔混淆了“拓扑意义上可忽略”（即无处稠密）的集合与在测度论的观点下可忽略的集合。他循着狄利克雷指引的方向，尝试用拓扑的角度去刻画一个可积函数的 $S_\sigma$ 集合。与狄利克雷一样，汉克尔也是因为对无穷集（特别地，无处稠密集）的多样性的认知不足而误入歧途。直到人们发现了无处稠密集也可以拥有正的外测度，这些小集合在测度论的意义下的重要性才被发掘出来。在黎曼对于可积性的等价条件的定理 2.1.2中其实是隐含了一些测度论的思想的，但这些思想还没有被当时的人们发掘出来，主要是因为那个光鲜亮丽的看似等价的拓扑思想混淆了人们的视听。但是无论如何，通过定义函数在某点处的“跳跃”的概念，汉克尔还是将注意力集中到了“集合”（$S_\sigma$）的性质身上，这样人们迟早会认识到，集合在测度论意义下的性质在研究可积性的问题时是至关重要的。

黎曼通过认识到柯西在积分定义中对被积函数连续性的要求本质上是不必要的，对积分理论的发展做出了贡献。汉克尔虽然犯了一个历史性的错误，但在另一方面，他也证明了如下一个与黎曼的发现互补的结论：

**定理 2.1.4.** 一个黎曼可积函数的连续点必须是一个稠密集。

汉克尔的证明基于对任意的正实数 $\sigma$，集合 $S_\sigma$ 都必须无处稠密。由此，对于给定的区间 $I_0$ 与正实数 $\sigma$，都存在一个 $I_0$ 的子区间 $I_1$，使得 $I_1$ 与 $S_\sigma$ 不交；同样地，存在 $I_1$ 的子区间 $I_2$，使得 $I_2$ 与 $S_{\frac{\sigma}{2}}$ 不交；于是我们可以归纳地得到一系列 $I_n$，使得 $I_n$ 与 $S_{\frac{\sigma}{2^N}}$ 不交。由于 $\cap_{n=1}^{\infty} I_n \neq \varnothing$，区间 $I_0$ 就包含了一个连续点，于是便证明了这个定理。（此处的区间 $I_n$ 当然需要是闭区间，但是汉克尔的证明中遗漏了这一点）。当然，这种对于黎曼可积函数中的连续点的拓扑刻画，某种意义上也掩盖了它在测度论意义下的性质。

## 达布

法国数学家达布也在黎曼积分理论的发展过程中迈出了重要的一步。达布一直担任着教授的工作，1872 年，他回到了曾经学习的巴黎高等师范学院（*École normale*）。后来，他对法国数学的发展贡献了非常大的影响，并且在组织与促进科学活动中取得了非凡的成功。

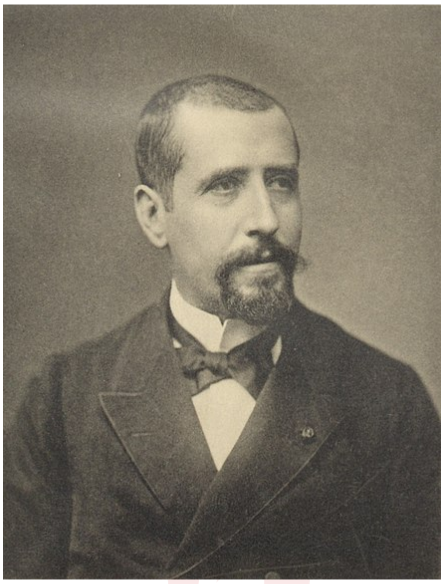


图 2.6: 达布

在他于 1872 至 1879 年发表的一系列文章中，他试图使法国数学界相信一个更严格的分析学研究方法的必要性。在这几篇文章中，达布为了说明相信直觉的潜在的（愈发上升的）危险性，给出了许多精巧的反例。在制造反例的时候达布基本上利用的是一致收敛的概念——这个概念十分重要，我们会在第 2.1.4 节中详细讨论。不过在这里让我们主要了解一下达布对积分理论做出的贡献。他通过对于黎曼的工作的翻译，以及对汉克尔的工作长期的高度评价，对黎曼积分理论在法国的转播产生的重要贡献。

达布首先明确，他考虑的仅限于有界函数 $f$（即那些满足 $A \leqslant f(x) \leqslant B,\ A, B, x \in \mathbb{R}$ 的函数）。对于这些函数，达布首先说明了对任意区间 $[x_0, x_1]$，存在三个实数 $m, M, \Delta \in \mathbb{R}$，$m$ 是 $f$ 在区间 $[x_0, x_1]$ 上的函数值的最大下界，$M$ 则是最小上界；$\Delta$ 则被定义为 $\Delta := M - m$，称作 $f$ 在对应区间上的振幅（注意到，振幅这个概念在前文中一直没有很好的定义）。达布明确指出，这些值不一定会被函数 $f$ 取到。这样，我们才真正区分开了最大/小值于上/下确界的关系。仅仅从此处，达布的叙述的严格性就可见一斑。

接着，对任意的有界函数 $f$，任意闭区间 $[a, b] \subset \mathbb{R}$ 以及其划分 $a = x_0 < \cdots < x_n = b$，定义 $\delta_i = x_i - x_{i-1}$，并且令 $m_i, M_i, \Delta_i$ 对应上文对各自区间的定义，那么我们

可以定义如下三个表达式：

$$M(n) := \sum_{i=1}^{n} M_i\delta_i \tag{2.1.25}$$

$$m(n) := \sum_{i=1}^{n} m_i\delta_i \tag{2.1.26}$$

$$\Delta(n) := \sum_{i=1}^{n} \Delta_i\delta_i \tag{2.1.27}$$

并且我们很容易得到，$\Delta(n) = M(n) - m(n)$，以及 $\Delta_i \leqslant B - A$。

接下来就是达布的伟大跨越：他证明了当 $n \to \infty, \delta_i \to 0$ 时，$M(n), m(n), \Delta(n)$ 都趋于唯一一个极限，并且该极限只与 $a, b$ 以及 $f$ 有关。由此，历史上第一次出现了上积分与下积分的概念：

$$\underline{\int_a^b} f(x)\mathrm{d}x := \lim_{n\to\infty} \sum_{i=1}^{n} m_i\delta_i \tag{2.1.28}$$

$$\overline{\int_a^b} f(x)\mathrm{d}x := \lim_{n\to\infty} \sum_{i=1}^{n} M_i\delta_i \tag{2.1.29}$$

现在再回头来看黎曼对于振幅的想法，达布自然便引出了一种函数的分类方式：即将那些满足

$$\lim_{n\to\infty} \Delta(n) = \overline{\int_a^b} f(x)\mathrm{d}x - \underline{\int_a^b} f(x)\mathrm{d}x = 0, \tag{2.1.30}$$

即

$$\overline{\int_a^b} f(x)\mathrm{d}x = \underline{\int_a^b} f(x)\mathrm{d}x \tag{2.1.31}$$

的函数与其他的函数区分开来。

紧接着，有了这些准备工作，达布准备好对黎曼积分做出一次改革了：对于一个划分 $a = x_0 < \cdots < x_n = b$，以及一系列实数 $\theta_i \in [0,1]$, 记 $\delta_i = x_i - x_{i-1}$，考虑如下的达布和：

$$S(n, \delta, \theta) = \sum_{i=1}^{n} \delta_i f(x_{i-1} + \theta_i\delta_i), \tag{2.1.32}$$

达布和只与 $n, \delta = (\delta_1, \delta_2, \cdots, \delta_n), \theta = (\theta_1, \theta_2, \cdots, \theta_n)$ 有关。由于

$$f(x_{i-1} + \theta_i\delta_i) \in [m_i, M_i] \tag{2.1.33}$$

恒成立，我们有：$m(n) \leqslant S(n, \delta, \theta) \leqslant M(n)$。于是我们有如下定理：

**定理 2.1.5.** *达布和有唯一的不依赖于 $\delta,\theta$ 的极限，当且仅当 $\Delta(n)$ 趋于 0，也就是 $\overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx$ 成立。*

最后，达布将达布和的唯一的极限称作 $\int_a^b f(x)\mathrm{d}x$。现在我们就可以由此推出一系列结论了，例如：

**命题 2.1.6.** *(1) 所有连续函数都是（达布）可积的。*
*(2) 映射 $x \mapsto \int_a^x f(y)dy$ 关于 $x$ 连续。*
*(3) 令 $F(x) = \int_a^x f(y)dy$。若 $f$ 在 $x_0$ 处连续，那么 $F$ 在 $x_0$ 处可导，并且 $F'(x_0) = f(x_0)$。*

同样值得一提的是达布得出的如下定理：

**定理 2.1.7.** *若在区间 $[a,b]$ 上定义的可导函数 $F$ 拥有一个有界并且可积的导数 $f = F'$，那么对任意 $x \in [a,b]$，都有 $F(x) - F(a) = \int_a^x f(y)dy$ 成立。*

这样，我们一直以来所学到的“黎曼积分”的记号便终于建立起来了。此外，最后这个定理对黎曼的概念进行了强力的支撑：它只用一个相当弱的可积性条件，便证明了柯西必须假设函数连续性才能证明的微积分基本定理。

但另一方面，这个叙述也带来了相应的问题：如果有这样一个 $f$，它是区间 $[a,b]$ 上的可导函数，并且它在区间 $[a,b]$ 上不是常函数，但它有一个有界的导数 $f'$，它的零点在 $(a,b)$ 上稠密，那么这个导数就会是黎曼不可积的。（否则，我们总是可以适当地选取黎曼和中分划区间里面的中间值，使得它的函数值是 0，从而 $f(x) - f(a) = \int_a^x f'(t)\mathrm{d}t = 0$，这推出了 $f$ 是常函数这一矛盾的结果。）

意大利数学家乌利塞 • 迪尼（Ulisse Dini）是第一个观察到这一关系的人。而在不久以后，人们在讨论所谓“病态”函数以及一些与直觉相悖的反例的过程中，找到了大量能佐证迪尼想法的例子。1881 年，意大利数学家维托 • 沃尔泰拉（Vito Volterra）就发现了非平凡函数 $f$，它拥有一个有界的导函数 $f'$，并且 $f'$ 的零点是一个稠密集。在 1896 年瑞典数学家托尔斯滕 • 布罗登（Torsten Brodén）发表了一个类似的例子，这个例子主要基于所谓的“奇点凝聚原理”，这是一个由黎曼提出，汉克尔做出完善的原理。

我们其实可以在其中看出一些讽刺性，因为黎曼发表“奇点凝聚原理”的目的是为了说明他的理论的一般性，但却在后来被布罗登用来证明黎曼积分的定义不够一般：

> 对一个函数 $f$ 求导的过程可以产生一个有界但黎曼不可积的函数 $f'$。所以黎曼的积分与求导并不是完全的逆过程。

这里我们第一次触及了黎曼积分理论的痛点，这也是后来促使法国数学家亨利 • 勒贝格（Henry Lebesgue）发展新的积分理论的原因之一。勒贝格在他的博士论文中发展了一个新的“勒贝格积分”，在这篇文章的序言中他这样写道：

> “我们知道，有些函数的导数是黎曼不可积的。如果我们接受黎曼的积分符号的话，这一理论就不允许我们完全地解决微积分的基本问题，即给定导数如何寻找原函数的问题。所以我们自然地希望找到一个在最大限度上使得积分成为求导的逆运算的积分定义。”

当然，这里需要指出的是，在勒贝格以前，我们所提到的数学家当中，没有人有利用上述结果来批判黎曼积分的意思——黎曼积分在当时还是最主流的积分符号。但无论怎样，我们回过头来看的时候，这些结果可以被解释成最初被发现的黎曼积分的弱点。后来，人们发现了越来越多黎曼积分的缺点，这些也共同推动了积分理论的进一步向前发展。

### 2.1.4 黎曼积分的问题：一致收敛性

黎曼积分的另一个症结来自于傅里叶曾经使用过的一个假设——一个收敛的函数项级数可以逐项求积分。假设函数 $u_i:[a,b]\to\mathbb{R}$ 有一个收敛的部分和 $s_n(x)=u_1(x)+\cdots+u_n(x)$，人们理所当然地认为

$$\int_a^b\sum_{i=1}^{\infty}u_i(x)\mathrm{d}x=\sum_{i=1}^{\infty}\int_a^b u_i(x)\mathrm{d}x \tag{2.1.34}$$

即对级数逐项求积分，在这里我们用到了如下求极限的过程与求积分的过程的可交换性：

$$\int_a^b\lim_{n\to\infty}s_n(x)\mathrm{d}x=\lim_{n\to\infty}\int_a^b s_n(x)\mathrm{d}x \tag{2.1.35}$$

在十九世纪的下半叶，人们研究的重心还是放在傅里叶级数的收敛性上，所以没有过多地考虑极限的交换性问题，所以包括柯西和高斯的许多数学家们在尝试证明所有连续函数都可以被其傅里叶级数表示这个（错误的）命题的时候，都毫不犹豫地使用了这种可交换性。

第一个对这种可交换性提出质疑的是挪威数学家尼尔斯・亨利克・阿贝尔（Niels Henrik Abel）。柯西在 1821 年所著的《分析教程》中曾列出了一条命题，说的是如果一个级数在点 $a$ 的一个邻域内收敛，并且级数中的每一项都在 $a$ 处连续，那么收敛到的函数也将在 $a$ 处连续。阿贝尔在其 1826 年发表的论文中用级数

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}\sin(nx)}{n} \tag{2.1.36}$$

举出了柯西上述命题的一个反例。不过他还证明了另一方面，有一类特殊的无穷级数是可以通过每一项的连续性推出和的连续性的；这里他所说的特殊性正是级数的一致收敛。但阿贝尔却并没有意识到一致收敛性的普遍意义，因此也没将这一类特殊的级数单独拿出来讨论。

图 2.7: 阿贝尔

或许是接收了阿贝尔的暗示，魏尔斯特拉斯在这个问题上又往前迈出了一步。至少早在 1841 年，魏尔斯特拉斯就注意到了逐点收敛与一致收敛有着理论意义上很重要的区别。在魏尔斯特拉斯于 1856 年入职位于柏林的一所大学后，他便在他的课堂上强调一致收敛这个概念的重要性，特别地，他证明了只要级数在积分区间上一致收敛，那么便可以进行逐项积分。但是显然，至少在 1870 年海涅提到它之前，魏尔斯特拉斯的观点并不为人所知。在此之前，德国数学家赛德尔（Philipp Seidel）与爱尔兰数学家乔治•斯托克斯（George Stokes）分别在 1850 与 1848 年先后独立地发现了，如果一列连续函数组成的级数没有在一点 $a$ 处收敛地“无穷慢”——即在 $a$ 的邻域上一致收敛——那么该级数的和将会在点 $a$ 处连续。但他们都没能将他们的结果与逐项求积分的操作联系起来，或许是因为他们在考虑问题的时候将一致收敛性限制在了特定点的邻域这个局部上，而没有考虑它的全局性质。事实上，斯托克斯还明确表示出同意柯西有关逐项求积分的命题。在柯西于 1853 年发表论文对于他之前所提出的级数的连续性的命题添加了一致收敛的条件作为修正之后，他仍然没意识到他关于逐项积分的命题在本质上是站不住脚的。真正第一个将一致收敛与逐项积分联系起来的是魏尔斯特拉斯，他通过对逐项求导提出一致收敛的前提条件，引起了数学界对于逐项积分一般意义上的成立性的严重怀疑。

图 2.8: 海涅

在魏尔斯特拉斯的学生圈子之外，直到 1870 年，魏尔斯特拉斯的一位朋友，德国数学家爱德华 • 海涅（Eduard Heine）在他有关三角级数的论文中提出呼吁，人们才意识到一致收敛性在逐项积分里的重要性，海涅也在论文中明确提及了魏尔斯特拉斯：

> “到目前为止我们一直认为，当对一个收敛的函数项级数求积分的时候，如果对每一项的积分都是有限的，那么最终结果就应该等于对逐项求积分以后的结果的和。但是是魏尔斯特拉斯先生首先观察到，这条定理的证明要求这个函数项级数不仅得是收敛的，而且得是一致收敛的。”

海涅是德国的哈勒-维腾贝格大学（the University of Halle）的一名数学教授，他或许是从德国数学家格奥尔格 • 康托尔（Georg Cantor）那里了解魏尔斯特拉斯的观点的，后者在 1867 年在柏林成为博士之后来到了德国的哈勒。

傅里叶在之前是将任意函数的三角级数表示的存在唯一性建立在逐项积分的过程上。海涅观察到，既然对于非一致收敛级数来说，逐项积分是有疑问的，那么问题就来了，即函数 $f$ 的三角级数展开

$$f(x)=\frac{1}{2}a_0+\sum_{n=1}^{\infty}(a_n\cos(nx)+b_n\sin(nx)) \tag{2.1.37}$$

中的系数是否一定是唯一的。

就像海涅所观察到的那样，一个满足迪利克雷条件（见第 2.1.1 节）的分段单调的连续函数的傅里叶展开在整体上大致是一致收敛的，不过在那些满足 $f(x-)\neq f(x+)$ 的点的邻域上级数是不一致收敛的。对一个在区间 $[a,b]$ 上定义的级数 $\sum\limits_{n=1}^{\infty} f_n(x)$，如果

存在有限个点 $x_1, x_2, \cdots, x_n \in [a,b]$，使得级数在任意不包含 $x_i$ 的 $[a,b]$ 的子区间上一致收敛，那么我们就称它“弱一致收敛”（converge uniformly in general）到 $f(x)$。于是一个满足迪利克雷条件的函数 $f$ 的傅里叶级数便“弱一致收敛”于 $f$。海涅希望证明一个函数 $f$ 至多能有一个弱一致收敛的三角级数表示。由定义我们知道，这种收敛允许级数在点 $x_1, x_2, \cdots, x_n$ 并不收敛于函数 $f$ 的函数值，海涅说，是康托尔激发他考虑这种可能性的。

通过考虑 $f$ 的两种形如式 (2.1.37) 的可能的表示的差，海涅得到了这个问题的如下转化：

*如果对所有 $x \notin P$（$P$ 为一个集合），*

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n\cos(nx) + b_n\sin(nx)) = 0, \tag{2.1.38}$$

*那么所有的系数 $a_n$ 与 $b_n$ 是否一定要等于 0 呢？*

海涅已经可以得到：如果式 (2.1.38) 所示的级数是关于有限集 $P$ 弱一致收敛的，那么答案是肯定的。海涅的证明思路主要是先证明 $a_n$ 与 $b_n$ 当 $n$ 趋于无穷的时候趋于 0，再利用一些性质去证明 $a_n$ 与 $b_n$ 只能都是 0。

康托尔受到海涅研究的启发，开始研究唯一性问题，他给自己设定了一个任务：去掉弱一致收敛的假设。他在 1870 年至 1871 年间发表了一系列论文，终于彻底去掉了弱一致收敛的条件，因此之前的 $P$ 也就没有必要是个有限集了。1872 年，康托尔开始着手研究 $P$ 是一个无限集时三角级数表示的唯一性问题。

图 2.9: 康托尔

在当时的数学界，尽管稠密集与无处稠密集已经渗透入了当时许多数学家的工作中，包括狄利克雷、黎曼、汉克尔，但还从来没有人谨慎地研究过无穷集的性质，康托尔是第一个。因此，康托尔从一些基础的定义开始，开创了集合论的新天地。首先如果一个点 $x$ 的每一个邻域都包含 $P$ 中的无穷多个点，那么就称 $x$ 为点集 $P$ 的**极限点**（又称，**聚点**），并且定义 $P$ 的所有极限点构成的集合称为 $P$ 的导集，记作 $P'$。如果 $P$ 的导集是无限集，那么我们可以进一步定义 $P$ 的导集的导集（或称，二阶导集），记作 $P''$。于是我们便可以归纳地定义 $P$ 的 $n$ 阶导集 $P^{(n)}=(P^{(n-1)})'$。最后，如果一个集合 $P$ 的 $n$ 阶导集是非空有限集，那么就称 $P$ 是第 $n$ 类集合。为了解释这些概念和记号，康托尔引入了一些例子。首先是 $[0,1]$ 区间上的所有有理数，由于它的导集就是 $[0,1]$ 区间本身，所以它不属于任何第 $n$ 类集合。另一个例子是所有形如 $\frac{1}{n}$ 的实数，此处 $n$ 为正整数，这种情况下它的导集就是单点集 0，从而它是第一类集合。这便是康托尔所给出的所有例子。某种程度上，这反映了当时的人们对于无穷集的例子还是知之甚少。

康托尔得出的主要结果是，如果式 (2.1.38) 对除了一个第 $n$ 类的集合（$n\in\mathbb{Z}_+$）$P$ 以外的点均成立，那么式 (2.1.38) 中的系数就必须全为 0。康托尔的证明基于海涅对 $n=0$ 的情况（即 $P$ 为有限集时）的完善，他发现证明可以对 $n$ 归纳从而推广到任意的正整数 $n$。康托尔同样也意识到，对任意的 $n$ 都一定存在第 $n$ 类集合。

在积分学的发展过程中海涅和康托尔的论文无疑是起到了关键性的作用：他们呼吁大家意识到了逐项积分的问题，同样开创了研究无限集的理论的先河，而后来这个理论也深刻地影响了积分理论的发展。

我们之前提到，数学家达布也非常关心这个问题，并且他也曾明确提及海涅的工作。在达布 1875 年的论文中，他充分地讨论了级数的一致收敛性问题。他的讨论开始于一个当时已经为人熟知的定义：

**定义 2.1.8.** 级数 $\sum\limits_{k=1}^{\infty}u_k(x)$ 在区间 $[a,b]$ 上一致收敛

$$\Leftrightarrow \forall\epsilon>0,\ \exists N\in\mathbb{N},\ s.t.\ \forall n>N,\ \forall x\in[a,b],\ |R_n(x)|:=\left|\sum_{k=n}^{\infty}u_k(x)\right|<\epsilon, \tag{2.1.39}$$

在口头上，达布将其解释为“误差项” $R_n$ 一致地变小。接着他指出了一致收敛级数是保持连续性与可积性的：*如果所有的 $u_n$ 都在区间上是连续的（可积的），并且级数一致地趋于函数 $u$，那么 $u$ 在这个区间上也是连续的（可积的）。*

达布在他的工作中举出了很多深刻而有价值的例子。他首先证明了，一致收敛并不

是保持连续性的必要条件：

$$f(x) := x^2e^{-x^2} = \sum_{n=1}^{\infty}\left[n^2x^2e^{-n^2x^2} - (n+1)^2x^2e^{-(n+1)^2x^2}\right]. \tag{2.1.40}$$

这个级数在区间 $(0,1)$ 上并不是一致收敛的。事实上，我们有 $R_n(x) = n^2x^2e^{-n^2x^2}$，所以特别地，我们有 $R_n(\frac{1}{n}) = \frac{1}{e}$，这代表误差项并不是一致地趋于 0。但级数求和得到的函数是连续的。

达布举出的另一个例子更加有趣，因为它说明了，存在极限与积分不能交换的情况：

$$f(x) := -2xe^{-x^2} = \sum_{n=1}^{\infty}\left[-2n^2xe^{-n^2x^2} + 2(n+1)^2xe^{-(n+1)^2x^2}\right]. \tag{2.1.41}$$

我们对这个函数逐项积分再求和会得到 $\sum\limits_{n=1}^{\infty}\left[e^{-n^2a^2} - e^{-(n+1)^2a^2}\right] = e^{-a^2}$，但对这个函数在 $[0,a]$ 上积分会得到 $e^{a^2}-1$。这个级数同样也不是一致收敛的。

*所以，事实上，黎曼积分与取极限的过程在一般意义上是不相容的。*

这时候，我们想问：当时的人们是怎么看待这个问题的呢？那么第一个对上述结果做出归纳总结的是迪尼的学生，意大利数学家阿尔泽拉（Cesare Arzelà）。阿尔泽拉证明了，如果一个由一致有界并且黎曼可积的函数构成的级数，收敛到一个黎曼可积函数，那么积分与极限的过程可交换。

在 1896 年，美国数学家威廉·福格·奥斯古德（William Fogg Osgood）在一篇论文中预见了勒贝格的结果。他考虑一列在区间 $[a,b]$ 上连续的函数 $s_n : [a,b] \to \mathbb{R}$（前述的部分和就是这样的函数的例子），并且这列函数逐点收敛到一个连续函数。奥斯古德将这个假设称为“条件（A）”。

奥斯古德给出了一个满足条件（A），但并不一致收敛的函数列：

$$s_n(x) = \frac{n^2x}{1+n^3x^2},\ f(x) = 0,\ x \in [0,1]. \tag{2.1.42}$$

当 $n$ 逐渐变大的时候，这个函数在 0 附近会出现一个不断变高变细的峰。奥斯古德接着定义了如下概念：若在点 $x_0$ 的邻域上，$s_n(x)$ 有着无限高的峰，那么就将点 $x_0$ 称为$X$-点（在式 (2.1.42) 中 0 就是一个 $X$-点）。

奥斯古德紧接着证明了如下定理：

**定理 2.1.9.** *“如果 $s_n(x)$ 是一个满足条件（A）的函数，$(\alpha,\beta)$ 是是一个不包含 $X$-点的区间，$x_0 < x_1$ 是区间中的两个点，那么*

$$\int_{x_0}^{x_1}\lim_{n\to\infty}s_n(x)dx = \lim_{n\to\infty}\int_{x_0}^{x_1}s_n(x)dx. \tag{2.1.43}$$

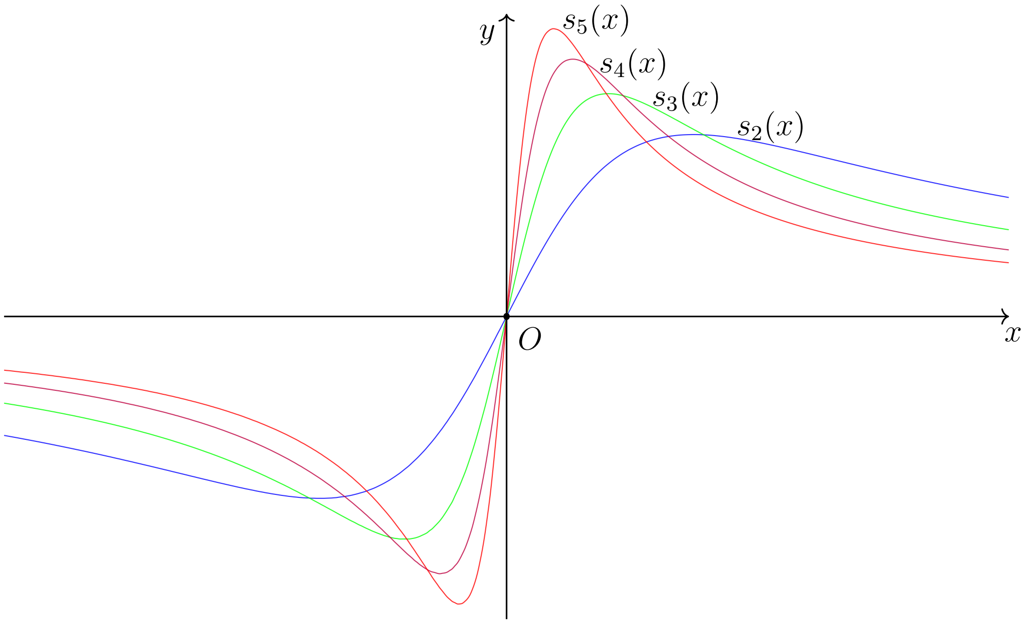



图 2.10: 奥斯古德的例子

如果 $s_n$ 是一个级数的部分和：

$$s_n(x) = u_1(x) + u_2(x) + \cdots + u_n(x), \tag{2.1.44}$$

那么这个级数就是可以逐项求积分的。”

换句话说，如果一列连续函数一致有界，并且逐点收敛到一个连续函数，那么极限与求积分就是可交换的。

我们注意，这里对于收敛到的函数可积的要求是必要的。首先数学家沃尔泰拉就曾经构造出一个沃尔泰拉函数 $g$，它处处可导，导函数 $g'$ 有界（$|g'| < M$），并且 $g'$ 是黎曼不可积的。那么我们可以构造一个一致有界的连续函数列：

$$f_s(x) = \frac{g(x+2^{-s}) - g(x)}{2^{-s}} \tag{2.1.45}$$

我们有：$f_s(x)$ 逐点收敛到 $g'(x)$，并且由拉格朗日中值定理，我们有：

$$\exists\ \xi_s \in (x, x+2^{-s}),\ |f_s(x)| = |g'(\xi_s)| < M, \tag{2.1.46}$$

从而 $f_s$ 一致有界，但 $g'$ 是黎曼不可积的。

但在勒贝格理论下，（如果只考虑有界集合的话）只需要勒贝格可积的函数列 $f_n$ 是一致有界就行了，我们可以推出极限函数是勒贝格可积的，从而积分与极限就可交换，这一美好的愿景得以实现。

与黎曼积分相反，勒贝格积分与取极限的过程在一般意义上是相容的。

勒贝格的结果也告诉我们，即使是“最一般”的黎曼积分记号也是可以进一步推广的——在积分理论重新用集合与测度的语言重新表述之前，这种洞察是难以想象的。

### 2.1.5 保罗·杜·布瓦-雷蒙（Paul du Bois–Reymond）

发展新积分理论的代表人物中最富热忱的便是德国数学家保罗·杜·布瓦-雷蒙（Paul du Bois–Reymond）。在十九世纪七十年代早期他熟悉了黎曼、汉克尔、海涅以及康托尔的工作。在 1873 年的秋天他完成了他的著作 *Versuch einer Classification der willkurlichen Functionen reeler Argumente nach ihren Aenderungen in den kleinsten Intervallen*，文章的主题是希望通过函数在最小的区间上的值的变化来区分一个函数。这部作品标志着他融入了分析学概念发展的进程，并指出了 1870 年代人们看待分析学的概念发展的观点。

图 2.11: 杜·布瓦-雷蒙

特别地，杜·布瓦-雷蒙接受了狄利克雷关于函数是一种对应关系的概念。这种一般意义上的函数被称为“无假设函数”，因为正如汉克尔已经强调的那样，它们不需要具有以前赋予的所有函数的性质。根据杜·布瓦-雷蒙的说法，函数最弱的限制条件是它的黎曼可积性；因此，可积函数被认为是无假设函数类中最大的子类。这种描述反映了黎曼的接班人对黎曼可积性条件的看法。从勒贝格积分理论的角度来看，黎曼的积分理论确实显得不够普遍。但是在 1870 年代——甚至很大程度上，在 1880 年代——可积

函数类似乎包含了一个比迄今为止所设想的要大得多的函数域。进一步地，将定积分定义为柯西和的唯一极限，这一为黎曼理论奠基的定义，被毫无疑问地接受为定义积分的自然方式。因此，黎曼的扩展被认为是终极的。就好像杜・布瓦-雷蒙所说的那样，黎曼已经对可积函数类进行了最大程度的扩展。

黎曼所给出的可积函数的例子（见式 (2.1.17)）表明，狄利克雷的可积性条件（不连续点必须形成一个无处稠密的集合）对于黎曼可积性并不是必要的。这个条件同样也不是充分的，我们将在第 2.2.1 节中介绍具有正测度的无处稠密集 $S$ 的发现，于是我们很容易构造出在一个无处稠密集上不连续但不可积的函数。汉克尔以为这样的集合 $S$ 并不存在，杜・布瓦-雷蒙或许也是这样认为的，因为他认为狄利克雷的可积性条件比黎曼的条件要强，也就是一个可积性的充分条件。

造成杜・布瓦-雷蒙的错误的原因是相对清晰的。一切要源于康托尔之前所定义的一个概念，在第 2.1.4 节中我们提到康托尔研究了在关于集合 $P$ 的弱一致收敛的条件下，一个函数的三角级数的系数的唯一性问题，他证明了只要 $P$ 在有限次求导过后是有限集，那么唯一性就是成立的。事实上，康托尔为这样的集合专门起了一个名字，叫做 *first species sets*，在本文中我们类比可解群的定义将其暂且称做"可解集"。我们容易知道，可解集一定是无处稠密集，并且比汉克尔在图 2.5中所能想象到的任何集合都要复杂。但我们并不难得到，可解集一定是可以被总长无限小的区间覆盖住的。因此，如果无处稠密集被错误地认为一定可解的话，那么汉克尔的错误命题（命题 2.1.3）就将是合理的，而狄利克雷的可积条件将是黎曼可积的一种特殊情况。

在讨论杜・布瓦-雷蒙的想法之前，我们对前述有关可积性条件的文字进行一个小小的总结。在康托尔定义了可解集之后，摆在数学家面前的是三种无限集合：零测集、无处稠密集以及可解集。最初的数学家们寄希望于利用可解集的概念来区分零测集与无处稠密集，但是很不幸的是，可解集既是零测集，又是无处稠密集，所以它几乎对考虑这个问题没有任何帮助，甚至可能帮了倒忙。当时数学家对无限集贫乏的认知反而导致了对这三种概念灾难性的混淆，直到十九世纪七十年代，人们才普遍意识到这三种概念是互不相同的。杜・布瓦-雷蒙便在这个概念中掉进了坑里。

#### 杜・布瓦-雷蒙的想法

杜・布瓦-雷蒙大概是以为可解集与无处稠密集是同一个概念，在下面杜・布瓦-雷蒙在 1875 年发表的文章中便可见一斑：

"无穷个不形成任何线段的点可以以两种方式分布在区间上。第一种方式是，

在每个任意小的区间中都存在这样的点，就像是有理数的分布一样。第二种方式是，在区间的每个任意小部分中，可以找到一个有限的线段，使得这些点都不位于该线段上。

我们希望更仔细地考虑第二种情形：这些点要么数量有限，要么只有在接近单个点时才变得稠密。因为如果它们在数量上是无限的，那么它们之间的所有距离不可能都是不趋于 0 的。但是，它们在任意小区间上的所有距离也不能消失，否则将化为上述第一种情况。因此，它们的距离只能逐点趋于 0，或者更准确地说，在无限小的区间上趋于零。

例如，方程 $\sin(\frac{1}{x})=0$ 的解集在 $x=0$ 附近稠密，而方程 $\sin\left[\frac{1}{\sin(\frac{1}{x})}\right]=0$ 的解集则在 $\sin(\frac{1}{x})$ 的根附近稠密，以此类推，于是下述方程

$$0=\sin\cfrac{1}{\sin\cfrac{1}{\sin\cfrac{\ddots}{\cfrac{1}{\sin\cfrac{1}{x}}}}} \tag{2.1.47}$$

的解集便给出了第二种情形的分布情况的例子。”

于是，方程 $\sin(\frac{1}{x})=0,\sin\left[\frac{1}{\sin(\frac{1}{x})}\right]=0,\cdots$ 的解集分别是第一、二、…类集合。显然，杜・布瓦-雷蒙认为，如果一个无限的点集只在点上而不是在整个区间上变得“无限稠密”的话，那么这些点就可以可解集的层次结构特征进行分类。在随后的一篇论文中，他暗示道，如果一个集合不是可解集，那么它必须在某个区间内是稠密的。这种对无处稠密集的结构的解释代表了利普西茨的论文中所指出的趋势的延续。

因此，杜・布瓦-雷蒙混淆了无处稠密集的概念与可解集的概念；因此，他便无法认识到这两个概念与零测集这个概念之间有何区别。在当时人们缺乏对汉克尔的命题（命题 2.1.3）的反例，甚至没有人提出具体的批评，这强烈表明，杜・布瓦-雷蒙所表现出的那种困惑是普遍存在的。

### 2.1.6 黎曼积分的问题：高维积分

最后一个关于积分理论发展的重要问题，来源于我们对于如何处理像 $f:\mathbb{R}^n\to\mathbb{R}$ 这样的多元函数的疑问；最简单的，式子

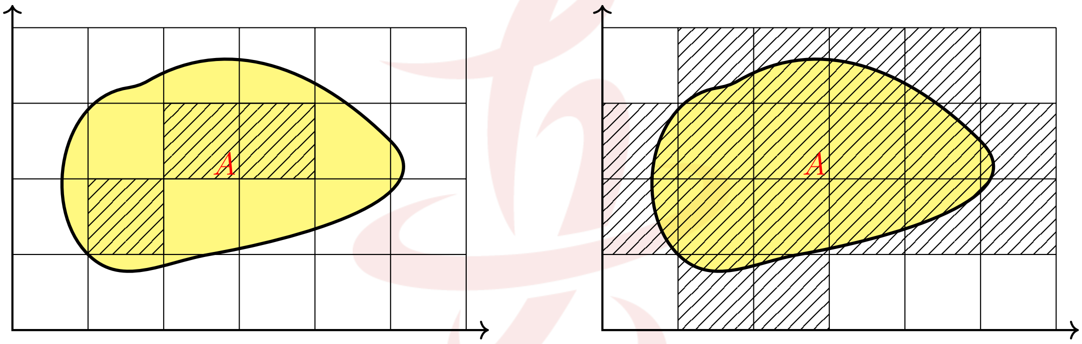
$$\int_A f(x,y)\mathrm{d}a \tag{2.1.47}$$

当 $A$ 是一个在平面 $\mathbb{R}^2$ 中的定义域时代表了什么意义呢？

一个直接的想法是类比区间划分为小区间，将平面划分为小矩形 $R_{ij}$，每个小矩形的面积是 $a(R_{ij})=\Delta(x_i)\Delta(y_j)$，然后考虑如下形式的柯西-黎曼和：

$$\sum f(x_i,y_j)a(R_{ij}),\quad (x_i,y_j)\in R_{ij}. \tag{2.1.48}$$

但是有一个问题是，我们到底应该规定哪些小矩形的面积为这个和式做贡献呢？是那些完全被 $A$ 包含的矩形？还是只要包含 $A$ 中的点的矩形？图 2.12给出了这两种贡献方式的区别。




(a) 完全被 $A$ 包含　　(b) 包含 $A$ 中的点

图 2.12: 两种贡献方式

我们一般会认为，这两种贡献方式得到的值的差异，也就是那些与 $A$ 的边界有接触的矩形的面积和，在分划越来越细的过程中会趋于 0。从而图 2.12所示的两种贡献方式应该得到同样的结果。但由于意大利数学家朱塞佩 • 皮亚诺（Guiseppe Peano）发现了一条“空间填充曲线”——一条经过所有点 $(x,y)\in[0,1]^2$ 的连续曲线，上述断言的正确性就得到了质疑。

> 我们需要对 $A$ 的边界所满足的条件做何种程度的加强，才能给出一个唯一且充分的积分定义呢？

在皮亚诺的曲线发表两年以后，法国数学家卡米尔 • 若尔当（Camille Jordan）给出了这个问题的部分回答，也因此为积分的发展引出一条新的路线。虽然若尔当的思维

与旧的积分理论有着紧密的联系，但是他也是第一个厘清测度论与积分理论之间的联系的人。

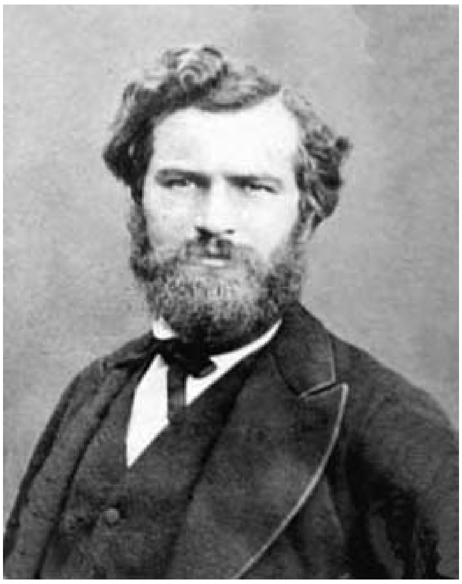


图 2.13: 若尔当

在讨论若尔当的贡献之前，让我们先对我们之前所进行的关于黎曼积分的讨论做一些总结。最开始，黎曼、汉克尔以及特别是达布的贡献，使我们不断完善了黎曼概念的表述；但同时，勒贝格等人的结果也越来越清晰地暴露了黎曼积分——这一原本被认为是拥有最一般的性质的概念——的缺点：

(1) 存在函数，它的导数有界，但是黎曼不可积。求导与黎曼积分在一般意义上并不是互为逆过程。

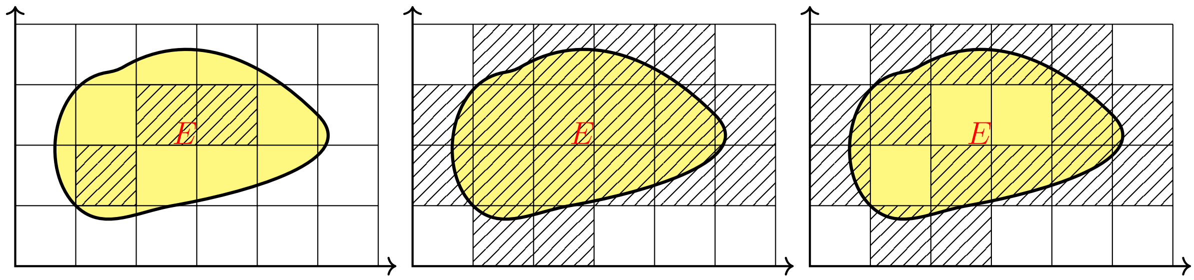
(2) 黎曼积分与求极限的过程并不能随意交换。由可积函数的序列 $f_n$ 逐点收敛得到的函数 $f$ 并不一定是黎曼可积的。即使 $f$ 是可积的，序列 $\int_a^b f_n(x)$ 也有可能收敛到一个与 $\int_a^b f(x)$ 不同的值。

(3) 黎曼的概念对于高维积分的问题尚不能给出令人满意的答案。

黎曼与达布的工作使得对一个函数 $f$ 在给定区间上的定积分有了清晰的定义，但是这种定义并不能涵盖对面积甚至更高维积分的情况。就像若尔当所说：“这种积分区域在之前并没有得到同样谨慎的对待”。

若尔当的出发点是对面积积分的集合论分析。为此，他首先发展了一个计算 $\mathbb{R}^n$ 的子集的大小（测度）的理论。他的论述集中在 $n=2$ 的情况上——也就是平面 $\mathbb{R}^2$ 的子集——虽然这一理论对于一般的 $n$ 都适用。首先，我们假设 $E \subset \mathbb{R}^2$ 是平面上一块有界的区域，并且用两条边平行于坐标轴的边长为 $r$ 的正方形去分割它。

在此种情形下，若尔当的策略似乎与我们之前所提到的处理高维积分的直觉想法很类似：对边长为 $r$ 的分划 $P$，若尔当一方面考虑所有那些严格处于 $E$ 中的（也就是与 $E$ 的补集 $E^c$ 不相交的）正方形的并集 $S$；另一方面考虑那些与 $E$ 至少有一个公共点的正方形的并集 $S+S'$。那么这两个集合的差集 $S'$ 就由那些包含的 $E$ 的边界 $\partial(E)$ 的正方形构成。（这里 $E$ 的边界被定义为所有 $E$ 中可被 $E^c$ 中的点逼近的点与 $E^c$ 中可被 $E$ 中的点逼近的点的并，从而一个边界点可以被描述为任意邻域都同时包含 $E$ 与 $E^c$ 中的点的点。）




(a) $S$ (b) $S+S'$ (c) $S'$

图 2.14: 若尔当的分划

所有这些集合的面积都是良定义的，因为它们都是 $r^2$ 的整数倍。由此，若尔当证明了如下定理：

**定理 2.1.10.** *如果对平面的分划越来越细，使得 $r$ 趋于 0，那么 $S$ 与 $S+S'$ 的面积将各自趋于一个定值 $A$ 与 $a$。*

这个定理的证明不仅是（几何）直观的，并且它清晰地重演了那个我们之前很熟悉的过程：即对函数的定义域（积分区域）进行分划并且不断细分。

现在，如果 $A$ 与 $a$ 相等的话，覆盖 $E$ 的边界的集合 $S'$ 的面积就会趋于 0。这引发了如下的定义：

**定义 2.1.11** (若尔当). 令 $E \subset \mathbb{R}^2$ 代表一个有界的集合。那么称 $\underline{\mathcal{J}}(E) := A$ 为集合 $E$ 的**若尔当内容度**，$\bar{\mathcal{J}}(E) := a$ 为集合 $E$ 的**若尔当外容度**。集合 $E$ 称为**若尔当可测的**，如果 $\underline{\mathcal{J}}(E) = \bar{\mathcal{J}}(E)$。在此情况下，$\mathcal{J}(E) := A = a$ 被称作 $E$ 的**若尔当容度**。

上述定义非常清晰地定义了 $E$ 在若尔当意义下的可测性，因此只要 $E$ 是若尔当可测的，我们就可以直接由柯西-黎曼和推广得到积分 $\int_E f(x,y)\mathrm{d}e$ 的定义。具体地来说，假设函数 $f:\mathbb{R}^2 \to \mathbb{R}$ 是在若尔当可测集 $E \subset \mathbb{R}^2$ 内部的一个有界函数，那么我们可以依照下面几步来定义（若尔当意义下的）$f$ 在 $E$ 上的积分 $\int_E f(x,y)\mathrm{d}e$：

若尔当的第一步是考虑将 $E$ 分划为有限个不交的若尔当可测集的并 $E = \sum_{k=1}^{m} E_k$，记

$$M_k := \sup\{f(x) : x \in E_k\}, \quad 1 \leqslant k \leqslant m, \tag{2.1.49}$$

$$m_k := \inf\{f(x) : x \in E_k\}, \quad 1 \leqslant k \leqslant m, \tag{2.1.50}$$

接着若尔当对达布的上下和做了直接的推广：

$$s(P) := \sum_{k=1}^{m} m_k \mathfrak{J}(E_k), \quad S(P) := \sum_{k=1}^{m} M_k \mathfrak{J}(E_k). \tag{2.1.51}$$

这一过程明显是出于对达布的模仿。两者之间唯一的区别就是区间与区间长度被分别替换成了若尔当可测集与它的若尔当容度。若尔当这样写道：

> "达布先生已经证明过：如果分划不断变细，直到其中的区间的直径趋于 *0*，那么 $S(P)$ 与 $s(P)$ 就会趋于一个定值。"

若尔当利用自己的术语给了上述命题一个新的证明，在他的证明中，他记 $\lim S = T, \lim s = t$。接着他将这两个极限分别定义为上积分与下积分（这里对更高维情况的推广是直接的）：

**定义 2.1.12.** $t$ 被称为 **"下积分"**，$T$ 我被称为 **"上积分"**。如果 $f$ 的上积分与下积分相等，就称 $f$ 是可积的。在此情况下，$\int_E f(x,y)\mathrm{d}e = T = t$ 就被称作 $f$ 在 $E$ 上的**积分**。

现在如果我们将其限制在 $n = 1$ 的情况下，这样得到的一维的积分的定义首先在连续性方面就沿袭了达布的积分的性质，进一步地，它还对分划的概念进行了拓宽，所以它又是达布的定义的推广。当达布规定将区间 $[a, b]$ 分划成有限多小区间的时候，若尔当则允许将区间分划成若干有限个若尔当可测集（包括区间）。所以很清晰的一点是，所有黎曼可积（也就是在达布的意义下可积）的函数必定也是在若尔当的意义下可积的函数。

此外，这种考虑方式还引出了如下问题：我们是否可以通过扩大可测集的概念来拓宽分划的概念，以此来定义更一般的积分？

为了回答这个问题，我们将在第 2.2 节中详细讨论测度论的发展，我们将见证若尔当与博雷尔有关可测集的想法被写进勒贝格集大成的博士论文中。

事实上，若尔当几乎一字不差地将他的结果带到了他的著名教科书《巴黎综合理工学院分析教程》（*Cours d'Analyse de l'École Polytechnique*）的第二版中。他的理论于是开始变得广为人知（特别是在法国）。这本书被广泛传播，埃米尔・博雷尔（Emile Borel）和亨利・勒贝格（Henri Lebesgue）等人阅读了它。用 1894 年至 1897 年间在巴黎高等师范学院学习的 Lebesgue 的话来说：

> “通过将部分集合论插入他的《巴黎综合理工学院分析教程》中，……若尔当在某种意义上恢复了这一理论——他已经证实它是数学的一个分支。他不仅证实了这一点，而且还在证明中引入对面积和集合的测度以及积分的研究，为某些工作，尤其是我自己后来的工作铺平了道路。”

若尔当——1900 年左右的法国著名数学家之一——验明了分析学研究中集合论的角度的可行性。我们将在接下来的一节中进一步了解它所带来的结果。

## 2.2 测度论，勒贝格积分

### 2.2.1 测度论的发展

在第 2.1.3 节中我们已经看到，一些集合表现出来的状态非常混乱。汉克尔在此之前已经尝试过去刻画某种特殊的集合——函数的不连续点集——的“大小”，也就是在这样研究的过程中，他做出了无处稠密等价于可以被总长任意小的区间覆盖的假设。

为了简化我们接下来的语言，我们定义：

**定义 2.2.1** (零测集). 若一个集合 $A \subset \mathbb{R}$ 可以被可数个（有限个）总长趋于 0 的区间覆盖，即对任意 $\epsilon > 0$，存在一个可数的（有限的）下标集 $M \subset \mathbb{N}$，使得存在一系列小区间 $I_m \in M$，满足

$$A \subset \bigcup_{m \in M} I_m, \quad \sum_{m \in M} |I_m| < \epsilon, \tag{2.2.1}$$

我们就称它为**勒贝格（若尔当）零测集**。

于是汉克尔的错误可以被翻译为：拓扑意义下的小集合（无处稠密集）同样是在测度论意义下的小集合（零测集）。

事实上，如果勒贝格（若尔当）零测集与无处稠密集被认为是同一种东西的话，那么一个基于零测集语言的测度理论就显得有些多余。但是，测度理论的必要性很快变得清晰起来，因为人们发现了一个无处稠密但并非零测的集合。

**具有正若尔当容度的无处稠密集的发现以及对勒贝格（外）测度理论的尝试**

英国数学家、牛津大学教授斯蒂芬·史密斯（Henry John Stephen Smith）在 1875 年已经发表了一个具有正若尔当容度的无处稠密集的例子，但起初并没有得到认可。史密斯曾在欧洲大陆学习，多亏他的外语知识，使他能够跟上数学界的最新发展。他主要研究可积性问题，尤其是对黎曼准则的优化。似乎他的读者，甚至他自己都没有意识到他在这里所举出的例子与测度理论的相关性。

图 2.15: 史密斯

大概是受到汉克尔的启发，史密斯考虑了函数的跳跃间断点，并研究对于 $\delta > 0$ 有无限多个跳跃大于 $\delta$ 的跳跃间断点的情况。为了对可能出现的不同情况进行分类，他（明确提出他）引入了汉克尔对稠密集和无处稠密集的区分，并将这两种情况分别称为“紧密”和“松散”。但与汉克尔相反，他意识到了一个无处稠密集可以展现出的复杂结构。

在史密斯构造的不能被无限小总长的集合覆盖的集合例子中，包含着一个非常基础性的想法，就是从区间 $[0,1]$ 中删去一些总长小于 1 的区间，这一想法在后来的许多有关的例子中都有所体现，所以我们在这里跟随史密斯的脚步，具体叙述这个例子的构造方式：

- 首先取一个正整数 $m > 2$。

- 第一步，将 $[0,1]$ 区间划分为等长的 $m$ 个小区间，并将最后一个区间称为 $I_1$（在这里为了补充史密斯的想法我们假设 $I_1$ 是开区间）。并将 $I_1$ 这个长度为 $\frac{1}{m}$ 的区间删去。

– 第二步，将剩余的 $m-1$ 个小区间中的每一个都划分为等长的 $m^2$ 个长度为 $\frac{1}{m}\cdot\frac{1}{m^2}$ 的小区间，继续将每个小区间中的最后一个（假设为开的）小区间删去，称这些被删去的区间为 $I_2, I_3, \cdots I_m$。

– 剩余的 $(m-1)(m^2-1)$ 个小区间继续被等分为 $m^3$ 个部分，再将最后一个删去……

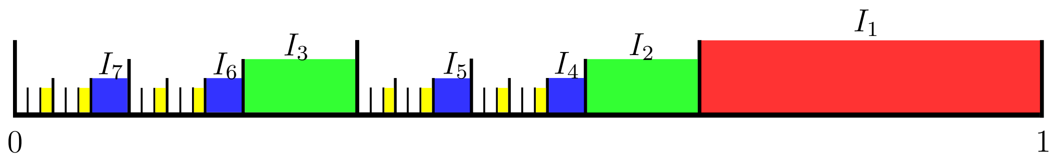



图 2.16: $m=3$ 的情况

在 $k$ 步之后，我们将会删去

$$N(k) = 1 + (m-1) + (m-1)(m^2-1) + \cdots + (m-1)(m^2-1)\cdots(m^{k-1}-1) \quad (2.2.2)$$

个区间 $I_1, I_2, \cdots, I_{N(k)}$，它们的总长为

$$\sum_{n=1}^{N(k)} |I_n| = 1 - \left(1-\frac{1}{m}\right)\left(1-\frac{1}{m^2}\right)\cdots\left(1-\frac{1}{m^k}\right). \quad (2.2.3)$$

当 $k$ 趋于无穷的时候，我们最终会得到一个集合 $X=[0,1]\setminus\sum_{n=1}^{\infty} I_n$，注意到，对于任意 $[0,1]$ 的子区间，总会有一个区间 $I_n$ 包含于它，从而在构造 $X$ 的过程中会将 $I_n$ 删去，于是 $X$ 在 $[0,1]$ 中无处稠密。但由于删去的区间的总长趋于 $1-\prod_{k=1}^{\infty}\left(1-\frac{1}{m^k}\right)<1$，从而我们不可能用总长任意小的区间覆盖 $X$。

事实上，史密斯自己是用一个稍微不一样的方式来得到这个结论的。他考虑了所有这些区间的端点构成的集合 $Q$，这是无处稠密集 $X$ 的子集，从而它本身也是一个 $[0,1]$ 上的无处稠密集。史密斯证明了：一个在 $Q$ 上有着有界的跳跃间断点的函数 $f$ 是不可积的。

史密斯的这个例子的重要性在于，它与汉克尔关于不连续函数的理论是矛盾的。史密斯提到，他先前在汉克尔的论文中读到（在第 2.1.3 节中提到过的）其对稠密集与无处稠密集的区分，他从而去研究了汉克尔有关所有无处稠密集都可以被总长趋于 0 的区间覆盖的“证明”。史密斯认为有限的情况是容易理解的，但他觉得汉克尔的构造“一旦点数变得无穷大，脑海里就开始无法产生直观清晰的印象了”——事实上，他的构造证明了汉克尔的构造是错误的——并且史密斯还认为，“无论我们是否承认汉克尔博士

的命题是正确的，他用此命题来证明的不连续点是无处稠密集的函数的可积性这个结论本身，看起来都像是错误的”。

### 来自德国的进展

史密斯自己，包括与他同时代的数学家们（在一开始）可能并没有完全意识到这个例子将会带来的影响，因此他对这个理论的构想有点谨慎和保守。但在遥远的德国——一个几乎没人知道史密斯及另一个数学家沃尔泰拉给出的例子的地方——在十九世纪八十年代初期也有一些数学家们独立地取得了一些进展，当时德国数学家们很重视这些发现，以至于后来创立了一个新的理论，也就是测度论。

第一个呼吁大家注意具有正的若尔当外测度的无处稠密集的德国数学家是保罗 • 杜 • 布瓦-雷蒙，在第 2.1.5 节中我们提到，因为混淆了无处稠密集与可解集保罗 • 杜 • 布瓦-雷蒙将狄利克雷的可积性条件（即不连续点必须形成一个无处稠密的集合）视为黎曼可积的一种特殊情况。但是后来他发现，无处稠密集可以有所谓的无穷阶聚点（见第 2.1.4 节中康托尔的定义），构造步骤如下：

1. 令 $p$ 为一个给定的实数，取一个严格单调递增且收敛于 $p$ 的数列 $(a_n)_{n\in\mathbb{Z}_+}$；
2. 令 $I_k=[a_{2k-1},a_{2k}],k\in\mathbb{Z}_+$，则这是一列互相不交的闭区间；
3. 在每个区间 $I_n$ 上定义一个第 $n$ 类集合 $P_n$。（式 (2.1.47) 给出了一种可行的定义方式）；
4. 定义 $P=\bigcup_{n=1}^{\infty}P_n$，那么由于 $p\in P^{(n)},\forall n\in\mathbb{Z}_+$，此时点 $p$ 就是一个无穷阶聚点。于是 $P$ 并不是一个可解集，但是它很明显是无处稠密的。

不同于构造可解集时我们只去考虑一些点，上述集合 $P$ 的构造背后所隐含的思想大概是去分配区间。杜 • 布瓦-雷蒙从这里意识到，构造无处稠密集可以通过先在一个固定的区间例如 $[-\pi,\pi]$ 上分配无限个不交且稠密的区间，那么剩下那些不在这些区间中的点就会构成一个无处稠密集 $Q$。进一步地，如果我们可以让这些删去的区间的总长度小于 $2\pi$ 的话，那么 $Q$ 就不能被总长趋于 0 的区间覆盖住了。于是，就像杜 • 布瓦-雷蒙所指出的那样，$Q$ 上的特征函数（即在 $Q$ 中的点为 1，其余点为 0 的函数），尽管满足狄利克雷的可积性条件，但并不黎曼可积。

杜 • 布瓦-雷蒙在 1882 年的著作《函数理论》（*Die allgemeine Functionenthëorie*）中，将上述思路叙述得更加明确。这些思路与史密斯和沃尔泰拉的方法基本一致，而在当时他们的论文在德国仍然不为人所知。

除了杜·布瓦-雷蒙以外，还有一位德国数学家，阿克塞尔·哈纳克（Axel Harnack），在测度论的发展上做出了贡献。

图 2.17: 阿克塞尔·哈纳克

最开始，哈纳克关心的问题是任意函数的三角级数的可表示性。在他 1880 年的工作中他开始考虑级数的一致收敛性（魏尔斯特拉斯的工作使他认识到一致收敛的重要意义），紧接着，他就碰到了一种至关重要的集合，即零测集。但其实，一开始哈纳克也对这些集合的定义有些混淆。

但无论如何，哈纳克似乎很快就克服了这个困难。在他 1881 年出版的课本《关于微积分的要素》（“*On the elements of calculus*”）中“关于定积分作为和的极限的一般定理”（“*general theorems on the definite integral as a limit of a sum*”）这一章节中，他给出了非常准确的记号。在定义黎曼积分以后，如汉克尔一样，他分析了函数的不连续点构成的集合。对于无穷集的情况，他将其分成了两类：一类是若尔当容度（哈纳克显然想的是用有限个区间去覆盖）为 0 的集合，他称为“离散的”；其余的情况，则统称为“线性的”。

为了将这两类集合应用到积分理论中，他区分了所谓“逐点不连续”与“线性不连续”函数（与汉克尔非常类似，因为“离散集”也无处稠密）：如果对于所有 $\sigma > 0$，跳跃大于 $\sigma$ 的点构成的集合 $S_\sigma(f)$ 是离散的（若尔当容度为 0），就称 $f$ 是逐点不连续的；相反，如果这个集合是线性的，就称 $f$ 是线性不连续的。在这种定义下，哈纳克现在通过将黎曼的定理用自己的语言叙述，纠正了汉克尔曾犯下的错误：

**定理 2.2.2.** 逐点不连续函数是可积的，$\cdots$，线性不连续函数是不可积的。

若尔当容度的理论被康托尔进一步地发展。在他 1883 年的论文《关于无限线性点流形》（“*On infinite linear point manifolds*”）的第四部分里写道：

> “在杜·布瓦-雷蒙和哈纳克对积分定理推广的研究中，使用了一类点集，即可以被有限个总长趋于 0 的区间覆盖的集合。”

集合“在任意小的区间上都不稠密”这个条件被康托尔描述为若尔当容度为 0 的必要不充分条件。紧接着在文章的第六部分，康托尔再次考虑了若尔当测度的问题，并定义了今天我们所称的“外容度”（“outer contents”）。

在这一小节的最后，值得一提的是，在 1885 年哈纳克的文章《有关点集的容度》（“*On the content of point sets*”）中，他第一次明确指出，集散集——即容度为 0 的集合——“以防误解”，是需要覆盖区间为有限个。他补充道，某种意义上任何可数集都可以被可数个总长趋于 0 的区间覆盖。以 $[0,1]$ 区间上的有理数为例，这是一个稠密的集合，将这个可数集的元素写成一列 $a_1, a_2, \cdots$，则对任意的 $\delta > 0$ 我们总可以逐步地选取不交的区间 $I_1, I_2, \cdots$，使其长度依次为 $\epsilon_1, \epsilon_2, \cdots$，并且 $\sum_{i=1}^{\infty} \epsilon_i = \delta$。因此，如果我们不把覆盖区间的个数限制为有限个，就会出现一个奇怪的现象：拓扑意义上很大（稠密）的集合，测度论的意义上很小（零测）。

我们将在下一节中看到，在法国数学家埃米尔·博雷尔（Émile Borel）的手中，这些重要的想法是怎样被用于推动测度论的进一步发展的。

### 2.2.2 博雷尔测度

埃米尔·博雷尔出生于 1871 年 1 月 7 日，是一个乡村牧师的儿子。1889 年，他进入巴黎高等师范学院，在那里学习到 1893 年，并于 1894 年毕业。1909 年，他成为索邦大学（Sorbonne）的复分析教授（他已经在那里完成了博士论文），并于 1920 年被授予概率论和数学物理方向的主席。

除了他的政治活动（他自 1924 年以来一直是法国议会议员，1925 年甚至担任海军部长）以外，他还是一位杰出的科学工作的推动者和组织者。他撰写了一系列有影响力的复分析专著，后来又开始了关于概率微积分及其应用的大量工作。在他的倡议下，亨利·庞加莱研究所于 1928 年在巴黎成立，他自己是第一任所长。

我们还应该提到，除了他对复分析的贡献和他关于测度论的思想（见下文）之外，博雷尔早在 1905 年就指出了测度论对概率论的适用性。后来，这个想法影响了关于概

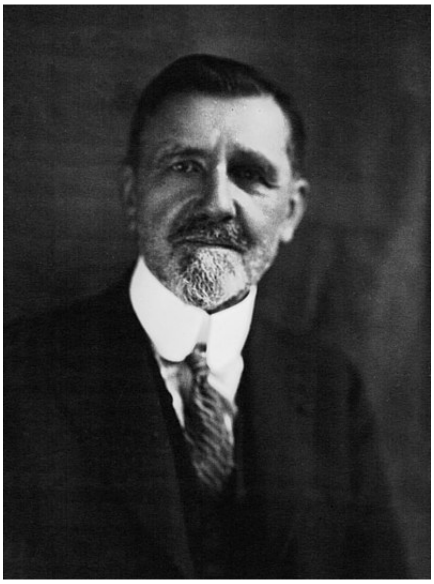


图 2.18: 博雷尔

率论中概念的重要的公理化定义的讨论，直到 1933 年俄罗斯数学家安德雷·柯尔莫哥洛夫（Andrei N. Kolmogorov）的著作最终确立了这一公理化定义。

博雷尔于 1956 年 2 月 3 日在巴黎去世。毫无疑问，他是二十世纪初最有影响力的数学家之一。

### 博雷尔的想法

在他 1894 年的博士论文中，博雷尔为复函数理论提出了一个富有成效的想法：他得以运用一个来自康托尔的想法，跨越奇点稠密的障碍，来对如下表达式：

$$f(z) = \sum_{n=1}^{\infty} \frac{A_n}{z - a_n}, \quad z, A_n, a_n \in \mathbb{C}, \quad \sum_{n=1}^{\infty} |A_n| < \infty, \tag{2.2.4}$$

进行解析延拓：假设 $\{a_n : n \in \mathbb{N}\}$ 是一条简单闭凸曲线 $C \in \mathbb{C}$ 的稠密子集，那么存在不可数无穷个 $C$ 中的点，使得上述级数收敛。这就是该表达式可以被延拓到 $C$ 之外的原因。

现在下一个问题是“所有收敛点的集合是什么样子的？”，博雷尔在 1896 至 1897 学年在他以前硕果频出的地方——巴黎高等师范学院的一系列讲座中详细考虑了这个问题。他的演讲备受好评，因此在 1898 年，《函数理论课程》（*Leçons sur la théory des fonctions*）一书得以问世。这本书首次考虑了我们今天称之为“博雷尔集”（Borel sets）的集合。在这书的第一部分，博雷尔发展了他在第二部分中处理他的原始问题所需的集合论语言。

在这里，为了说明可测集及其测度所发挥的关键作用，我们将研究最特殊的情况，即用刚才提到的级数 $f(z)$ 的实值情况进行类比，因此我们将问题限制在单位区间 $[0,1]\subset\mathbb{R}$ 上，考虑级数：

$$\sum_{n=1}^{\infty}\frac{A_n}{|x-a_n|},\quad x,a_n\in(0,1),\quad A_n\in\mathbb{R}^{+},\quad n\in\mathbb{N},\quad A:=\sum_{n=1}^{\infty}\sqrt{A_n}<\infty,\tag{2.2.5}$$

这里，假设集合 $\{a_n:n\in\mathbb{N}\}$ 在 $[0,1]$ 上稠密。

跟着博雷尔的思路，我们想要找到这个级数的收敛点。利用表达式 $v_n(k):=\frac{\sqrt{A_n}}{2k}$，我们定义一列区间：

$$I_n(k):=(a_n-v_n(k),a_n+v_n(k)),\quad n\in\mathbb{N}.\tag{2.2.6}$$

进一步，令 $B_k:=\bigcup_{n=1}^{\infty}I_n(k)$，那么下列命题成立：

**命题 2.2.3.** *若 $x\in[0,1]\setminus B_k$，我们有：*

$$|x-a_n|\geqslant v_n(k)\Leftrightarrow\frac{A_n}{|x-a_n|}\leqslant\frac{A_n}{v_n(k)}=2k\sqrt{A_n}\quad\forall n\in\mathbb{N}.\tag{2.2.7}$$

*从而级数 $\sum\limits_{n=1}^{\infty}\frac{A_n}{|x-a_n|}$ 在 $[0,1]\setminus B_k$ 上一致收敛。*

现在令 $D$ 代表所有该级数的不收敛点构成的集合。那么我们有：$D\subset\bigcap_{k=1}^{\infty}B_k$，即 $D\in B_k,\quad\forall k\in\mathbb{N}$。

由于 $B_k$ 是由总长不超过

$$\sum_{n=1}^{\infty}2v_n(k)=\sum_{n=1}^{\infty}\frac{\sqrt{A_n}}{k}=\frac{A}{k},\tag{2.2.8}$$

的区间构成，那么集合 $D$ 就可以被可数个总长趋于 0 的区间 $\{I_n(k):n\in\mathbb{N}\}$ 覆盖住。

于是用我们之前提到的记号来说，$D$ 是一个勒贝格零测集。这里着重强调的原因是，仅用若尔当的容度理论是不能充分说明问题的：如果只用有限多个区间的话，是不足以区分不收敛点集 $D$ 与收敛点集 $[0,1]\setminus D$ 的。因为这两个集合都是若尔当不可测的（因为 $\{a_n:n\in\mathbb{N}\subset D\}$，于是这两个集合的若尔当内容度均为 0，而若尔当外容度则均为 1）。

或许与他的数学哲学有关（他后来明确地提出了这种哲学）——这要求他只使用“可构造的定义”，即那些只用进行至多可数个操作就足够了的定义——这使他得出了以下定义：

**定义 2.2.4.**

*“如果一个集合由可数多个互不相交且总长度为 $s$ 的区间构成，那么就称这*

个集合的**测度**为 $s$。如果两个不交的集合的测度分别为 $s, s'$，那么他们的并集的测度就为 $s + s'$。"

更一般地，如果有可数多个不交的集合，它们的测度分别为 $s_1, s_2, \cdots$，那么它们的并集的测度就为 $s_1 + s_2 + \cdots$。

"如果一个测度为 $s$ 的集合 $E$ 包含一个测度为 $s'$ 的集合 $E'$，那么集合 $E \backslash E'$ 的测度就为 $s - s'$…"

最后：

"对那些可以通过上述定义分配一个测度的集合，我们就称其为**可测集**。"

博雷尔自己并没有触及可测集系统有多广泛的问题，取而代之的是，他强调了从上述定义中所能得出的四个重要性质的关键性：

- "可数多个（不相交）集合之和的测度对应于它们的测度之和"，即用现代的话来讲，测度是可数可加的；
- "两集合之差的测度（假设其中一个覆盖另一个）对应于它们测度的差值"；
- "测度始终是非负的"；
- "任何具有正测度的集合都是不可数的。"

在此，我们从根源上了解了博雷尔集合及其测度——尽管是以一种相当泛泛的形式，对于具有现代分析知识的读者来说，这或许看起来会比较熟悉，但对于跟博雷尔同时代的数学家们就并非如此了。事实上，一开始人们并不理解博雷尔的想法。例如，阿瑟·舍恩弗利斯（Arthur Schoenflies）准备了一篇关于"点流形理论的发展"的长篇报告，其中总结了这个理论的研究在世纪之交得出的知识，但其中博雷尔的想法并没有得到公正的对待。

但至少有一位年轻的数学家，在巴黎高等师范学院聆听博雷尔的课堂，并且在后来得以将博雷尔的思想推向繁荣发展，他就是亨利·勒贝格。

### 2.2.3 勒贝格的新方向

亨利·勒贝格（Henri Lebesgue）只比博雷尔小四岁。他于 1875 年 5 月 28 日出生，父亲是法国博韦一名印刷工人，母亲是一所私立小学的教师。通过社区的奖学金，他得

图 2.19: 勒贝格

以完成中学学业。在他得到数学教育的时期，主要使用的是若尔当的《分析教程》，并且那时集合论方法在分析处理方面也正处于突破时期。1894 年至 1897 年间，他在巴黎高等师范学院学习——因此他很有可能参加了博雷尔的讲座。

随后，他在南茜的一所女子中学教数学；但与此同时，他完成了博士论文《积分、长度、面积》(*Intégrale, Longueur, Aire*)，该论文于 1902 年答辩并出版。我们在第 2.1.3 节中已经提到了勒贝格研究的出发点：即在黎曼积分概念的框架中一些函数的导数的原函数并不存在。

毕业后，Lebesgue 开始了学术生涯。首先，他在雷恩和普瓦捷工作，直到 1910 年成为索邦大学的讲师，并于 1919 年成为教授。从 1921 年起，他在法兰西学院任教，并在那里度过了余生（他于 1941 年 7 月 26 日去世）。在此期间，他主要考虑的是教学问题和数学史。他的积分概念起初只是被人们带有犹豫地接受，最终获得了成功。它在应用方面的重要性，特别是对自然界中不连续和统计现象的分析，很快就不容小觑了。除了其他应用外，勒贝格积分还成为现代概率理论中的重要工具。

所以，勒贝格积分的关键点是什么？

勒贝格吸取了若尔当有关集合论的观点，像若尔当一样从集合的可测性入手。在一切的开始，他就要求了集合的测度 $m$ 具有可列可加性：即对至多可数个不交的集合 $E_i, i \in \mathfrak{I}$，我们有

$$m\left(\sum_{i\in\mathfrak{I}} E_i\right) = \sum_{i\in\mathfrak{I}} m(E_i). \tag{2.2.9}$$

现在让我们通过一个特殊情形来解释勒贝格有关测度定义的想法：即对区间 $[0,1]$ 的子集怎么定义其测度。

为此，如果我们选取一个任意区间的长度（例如 $[0,1]$）为单位长度，那么我们对于任意区间 $I$，就可以将它的长度 $L(I)$ 与它的测度 $m(I)$ 联系起来。于是对于可数多个不交的区间，我们所希望的可列可加性就可以成立：

$$m\left(\sum_{n=1}^{\infty} I_n\right) = L\left(\sum_{n=1}^{\infty} I_n\right) = \sum_{n=1}^{\infty} L(I_n) = \sum_{n=1}^{\infty} m(I_n). \tag{2.2.10}$$

我们很容易在这些式子中看到博雷尔的影子，勒贝格的想法实际上是对博雷尔的补充。事实上，如果我们想要对一个任意的集合 $E$ 赋予测度的话，那么如果用可数多个区间 $\bigcup_{n\in\mathfrak{J}} I_n$ 将 $E$ 覆盖，那么我们必须有：

$$m(E) \leqslant m\left(\bigcup_{n\in\mathfrak{J}} I_n\right) \leqslant \sum_{n\in\mathfrak{J}} L(I_n). \tag{2.2.11}$$

于是，覆盖 $E$ 的区间的长度之和 $\sum\limits_{n\in\mathfrak{J}} L(I_n)$ 的下确界就给出了 $E$ 的测度的一个自然的上界。

**定义 2.2.5** (勒贝格 1902). 对一个有界集 $E \subset \mathbb{R}$，令

$$m_e(E) := \inf\left\{\sum_{n\in\mathfrak{J}} L(I_n) : \mathfrak{J} \in \mathbb{N}, E \subset \bigcup_{n\in\mathfrak{J}} I_n\right\} \tag{2.2.12}$$

表示 $E$ 的**外测度**。

接着对于集合 $E \subset [0,1]$，勒贝格定义了它的补集 $C(E) = [0,1] \setminus E$。于是（如果测度 $m$ 有定义的话）我们有 $m(C(E)) \leqslant m_e(C(E))$。勒贝格得出的结论是: 无论 $m(E)$ 能否被定义，我们都有

$$m(E) \geqslant m([0,1]) - m_e(C(E)). \tag{2.2.13}$$

这一 $m(E)$ 的下界就被定义为 $E$ 的内测度。

**定义 2.2.6** (勒贝格 1902). 对于 $E \in [0,1]$ 如下表达式

$$m_i(E) := m([0,1]) - m_e(C(E)). \tag{2.2.14}$$

称为 $E$ 的**内测度**。

观察到我们总是有 $m_e(E) + m_e(C(E)) \geqslant m([0,1])$，于是

$$m_i(E) \leqslant m(E) \leqslant m_e(E). \tag{2.2.15}$$

现在通过这些方式，测度的问题就可以以一种自然的方法解决：

**定义 2.2.7** (勒贝格 1902). 对一个有界集 $E \subset \mathbb{R}$，如果有 $m_e(E) = m_i(E)$，那么就称其为**可测集**。并且令 $m(E) := m_e(E) = m_i(E)$ 表示 $E$ 的**测度**。

可测集还有以下性质：可数多个可测集的并集也是可测集；如果它们不交，那么测度还是可列可加的。

### 三种测度之间的区别

现在一个很自然的问题跃然纸上：勒贝格测度、博雷尔测度、若尔当测度这三种测度之间有什么联系呢？

首先，若尔当测度（容度）与勒贝格测度之间的关系是最明显的。勒贝格的外测度 $m_e(E)$ 的定义中允许至多可数个区间来覆盖 $E$，而若尔当的外容度 $\bar{\mathfrak{I}}(E)$ 则只允许使用有限个区间来覆盖 $E$。由此我们可以直接得到 $m_e(E) \leqslant \bar{\mathfrak{I}}(E)$。进一步地，我们还能观察到二者在内容（测）度的定义上的相似性：内容度定义为 $\underline{\mathfrak{I}}(E) = 1 - \bar{\mathfrak{I}}([0,1] \setminus E)$，而内测度定义为 $m_i(E) := m([0,1]) - m_e(C(E))$。由此我们可以直接得到：$\underline{\mathfrak{I}}(E) \leqslant m_i(E)$。

利用 $\underline{\mathfrak{I}}(E) \leqslant m_i(E) \leqslant m_e(E) \leqslant \bar{\mathfrak{I}}(E)$，我们便可得到：任意若尔当可测集都是勒贝格可测的，并且此时该集合的若尔当容度与勒贝格测度相等。

其次，由勒贝格测度的构造方式我们可以得到，博雷尔集也是勒贝格可测的（当我们把单位长度取为 $[0,1]$ 区间的长度时）。于是勒贝格可测性是博雷尔集的一种推广：博雷尔集的定义是构造性的，它不能保证零测的博雷尔集也是博雷尔可测的，但同样的结论在勒贝格的理论下就是成立的：勒贝格零测集的任何子集都是勒贝格零测的。

现在我们看到，勒贝格的想法是博雷尔的一种自然推广，但与博雷尔不同，勒贝格将这些想法运用到了积分理论上，这也是它的博士论文的第二部分的主要内容。

### 勒贝格积分

多年以后，在 1926 年 5 月位于哥本哈根举行的一场会议上，勒贝格主持了一场有关积分概念的发展史的讲座，并讲述了勒贝格积分这一想法的来源。

他解释道：通常意义下我们考虑积分（像黎曼与达布那样），遵从一种非常自然的逻辑：当我们将区间分划地越来越细，对于那些连续的函数 $f$，如下的量

$$\underline{f_i} := \inf\{f(x) : x \in (x_i, x_{i+1})\} \tag{2.2.16}$$

与

$$\overline{f_i} := \sup\{f(x) : x \in (x_i, x_{i+1})\} \tag{2.2.17}$$

之间的差也会越来越小，从而积分定义式的下界

$$\underline{S} := \sum \underline{f_i}(x_{i+1} - x_i) \tag{2.2.18}$$

与上界

$$\overline{S} := \sum \overline{f_i}(x_{i+1} - x_i) \tag{2.2.19}$$

就会越来越接近，最终趋于该积分的值。但是勒贝格认为，对于几乎处处不连续的函数，我们没有理由希望会有同样好的情况发生。我们不能仅仅通过选取足够细的分划，保证对应的量 $\underline{f_i}$ 与 $\overline{f_i}$ 之差会变得足够小。

为了解决这种情况，我们必须从另一个方向入手——而这便是勒贝格的重要突破：

> “现在很清楚的是，我们必须分划的不是区间 $(a, b)$，而是区间 $(\underline{f}, \overline{f})$ 这个由 $f$ 在区间 $(a, b)$ 上的函数值的上下界构成的区间。”

勒贝格确实找到了一个“新方向”：他考虑对有界函数的值域进行分划，进而考虑对应的原像构成的集合。从而所谓的“可测函数”的定义域就被划分为了若干个可测集，最后利用它们得到我们想要的结果。

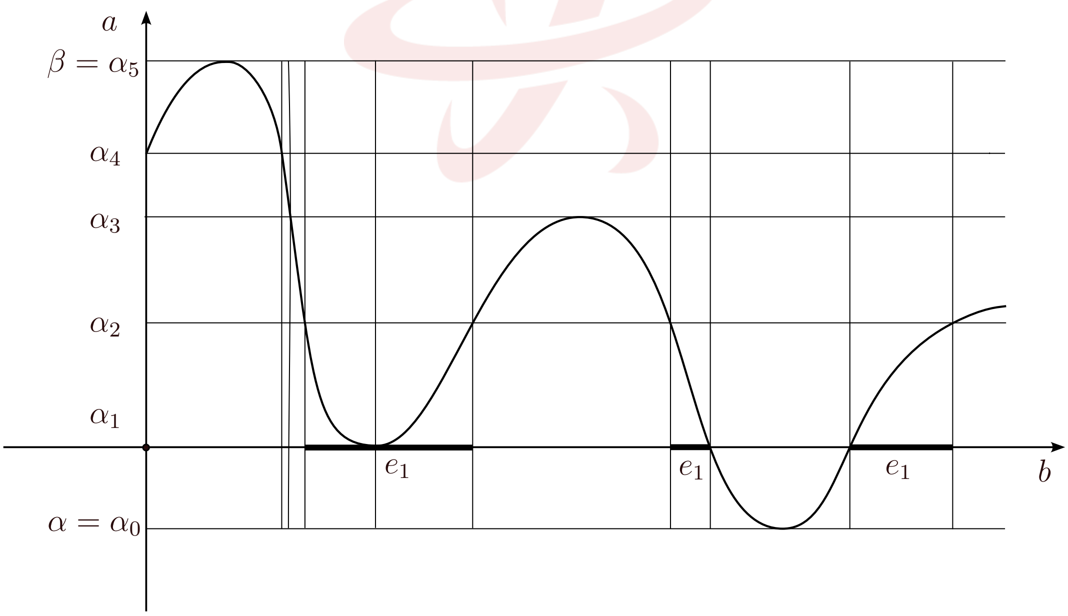



图 2.20: 勒贝格积分

详细地来说，我们首先对有界函数 $f$ 的值域进行分划：

$$\inf\{f(x) : x \in [a, b]\} := \alpha_0 < \alpha_1 < \cdots < \alpha_n := \sup\{f(x) : x \in [a, b]\} \tag{2.2.20}$$

然后定义对应的原像集：

$$e_i := \{x \in [a,b] : \alpha_i \leqslant f(x) < \alpha_{i+1}\}, \quad i = 0, \cdots, n-1, \tag{2.2.21}$$

（勒贝格在论文中的表述实际上分别考虑的开区间及其端点 $\alpha_i$ 的原像，在这里我们稍作修改）。如果这些集合 $e_0, e_1, \cdots, e_{n-1}$ 都是可测集，那么自然地我们可以去考虑

$$\sigma := \sum_{i=0}^{n-1} \alpha_i m(e_i), \tag{2.2.22}$$

以及对应的

$$\Sigma := \sum_{i=0}^{n-1} \alpha_{i+1} m(e_i), \tag{2.2.23}$$

于是它们便同之前一样，分别对应了积分的一个下界和上界。

接着令 $\|P\| := \max\{\alpha_{i+1} - \alpha_i : i = 0, 1, \cdots, n-1\}$，那么我们有：

$$\Sigma - \sigma = \sum_{i=0}^{n-1} (\alpha_{i+1} - \alpha_i) m(e_i) \leqslant \|P\| \sum_{i=0}^{n-1} m(e_i) = \|P\|(\beta - \alpha). \tag{2.2.24}$$

于是当 $\|P\| \to 0$ 时，$\Sigma$ 与 $\sigma$ 会收敛到同一个量，这便可以得到勒贝格积分的定义：

**定义 2.2.8** (勒贝格积分). 若一个函数 $f : \mathbb{R} \to \mathbb{R}$ 满足，任意形如

$$\{x \in \mathbb{R} : c \leqslant f(x) < d\}, \quad c, d \in \mathbb{R}, c < d, \tag{2.2.25}$$

的集合都是勒贝格可测集，就称 $f$ 为**可测函数**。若函数 $f$ 在区间 $[a,b] \in \mathbb{R}$ 上有界，可测，那么 $f$ 的**勒贝格积分** $\int\limits_a^b f(x)\mathrm{d}x$ 就被定义为上述 $\sigma$ 和 $\Sigma$ 的共同的极限。

在勒贝格的论文中他提到，他的定义有部分动机是希望定积分的概念可以拥有一些较好的性质，他希望新定义能够

- 将黎曼积分作为一种特殊情况涵盖在内；
- 同时适应单变量与多变量函数；
- 保证导数的变上限积分是一个原函数，从而解决微积分的基本问题。

事实上，勒贝格积分是黎曼积分的推广：任意黎曼可积函数都是勒贝格可积的。同时存在勒贝格可积的函数，它是黎曼不可积的，例如我们在第 2.1.1 节中介绍的狄利克雷函数式 (2.1.12)。对 $f = 1_{\mathbb{R}\setminus\mathbb{Q}}(x)$ 我们有 $\int\limits_a^b 1_{\mathbb{R}\setminus\mathbb{Q}}(x)\mathrm{d}x = b - a$，因为有理数是一个零

测集。进一步，勒贝格积分的定义也很容易推广到多变量的情况，因为我们只需要考虑 $\mathbb{R}^n$ 中的测度即可。

但如果这些就是勒贝格积分的全部的话，那么它相对于若尔当的积分还并不称得上是多大的跨越，事实上它还有很多真正将其与之前的积分区分开来的性质，我们这里介绍其中的两个：

**定理 2.2.9** (勒贝格 1902). *若 $[a, b]$ 上的函数 $f$ 有一个有界的导数 $f'$，那么 $f'$ 勒贝格可积，并且 $\int\limits_a^b f'(x)dx = f(b) - f(a)$.*

从而黎曼积分的一大弱点，同时我们在第 2.1.4 节的最后部分提到也是勒贝格的研究的一个出发点，在勒贝格的定义下被很好地解决了：

*只要导数是有界的，积分与求导的操作就互为逆过程。*

在第 2.1.4 节中提到的另一个黎曼积分的弱点便是积分与极限的可交换性问题，这个问题在勒贝格积分的框架下也很好地解决了，只要积分区域的测度有限，并且函数列是一致有界的，那么极限与积分就是可交换的：

**定理 2.2.10** (勒贝格 1902). *“如果一列有着一致的上下界的函数 $f_n$ 收敛到函数 $f$，那么 $f$ 的积分就等于 $f_n$ 的积分的极限。”*

所以我们观察到，勒贝格积分这个新概念并不是为了引入而引入；相反，它是消除黎曼积分弱点的重要工具。然而，勒贝格的理论还是花了一些时间才建立起来的：特别是在法国，他一开始就遇到了反对意见，特别是来自老一辈数学家的反对。因此，他的论文的首次发表是在意大利，并且在 1919 年之前都没有获得教授职位也就不足为奇了。

然而，人们可以得出结论，亨利・勒贝格的工作证实了傅里叶那天真的愿景，即“任意”函数确实可以用数学分析的方法处理——尽管人们寻寻觅觅了大概八十年的时间……