



微积分基础入门 V2

Makiror Ouyang

引言

作为 Machine Learning 必要的数学基础，微积分对我以后想深挖的领域显得很重要，包括以后我写需要相关数学基础的文章都会直接引用本文内容作为知识前提，然后本文只需要初中数学常识即可。

对V2的说明

我承认我不想英翻一半的时候突然写V2，它打乱了我的节奏，但作为一个“选择性的”完美主义者，我在写新的关于微分方程的博客时自己看了一下这篇文章的内容以统一写法，此过程中对这篇文章的内容的结构和完整性感到很不满意，于是索性决定做一个比较彻底的更新。当然，还有一部分原因就是可能会成为我以后比较重要的一篇文章，会在之后经常被我自己的博客引用作为数学基础。

2023.11.11

部分符号的约定和消歧义

发布以后，所有引用本文的延伸文章均默认遵循该约定，我说的。

在本文约定中 \mathbb{N} 表示一个自然数集合，包含所有非负整数，但是在数学中，这个符号有时也会用来表示非零自然数，而在数列极限/函数极限的定义中，自然数集合通常不包括零，所以之后在没有下标注明的情况下，我们使用这个集合统一表示**不包括0**的自然数集合，即非零自然数。为了消歧义，若表示自然数集合（包括0）我就使用 \mathbb{N}_0 。

此处约定的依据参考Wiki：

Natural number: https://en.wikipedia.org/wiki/Natural_number

Limit of a sequence: https://en.wikipedia.org/wiki/Limit_of_a_sequence

极限的下标根据行内或独立的公式不同，显示的位置会有所不同，例如 $\lim_{x \rightarrow a}$ 和 $\lim_{x \rightarrow a}$ ，在这种区别下，下标的用法含义没有区别。

部分比较简单的地方可能会使用 $f(x)'$ 的标记来表示导数以提高内容可读性，但在部分地方为了内容的全面，会以形如 $\frac{dy}{dx}$ 的莱布尼兹记法来补充相关公式的写法。

在V1，微分算子 d 我为了省事直接写为斜体 d ，但为了更严谨，在V2都被改成了符合国际标准 ISO 80000-2 第12条 的直立体 d 。

极限与连续性

要说微积分，那肯定先从极限和连续性讲起。

数列极限

当我们谈论数列的极限时，我们关注的是数列中的数字随着项数的增加而趋于的特定值。换句话说，数列的极限就是数列逐渐接近的某个值。

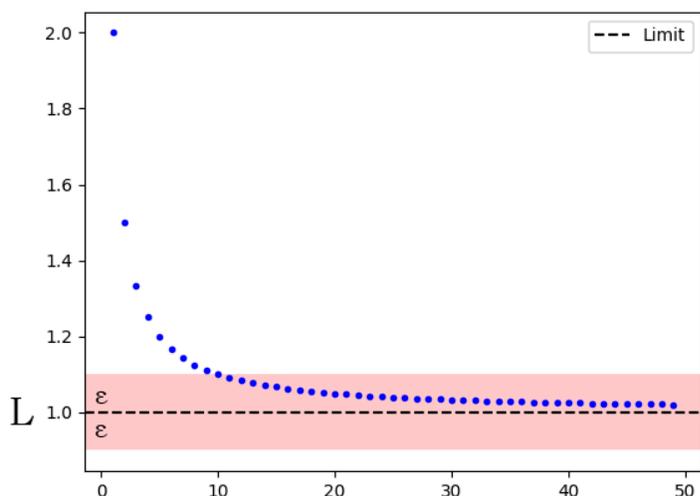
从定义上看，对于数列 $\{a_n\}$ ，若存在实数 L ，对于任意正实数 ϵ ，存在正整数 N ，使得当 $n > N$ 时， $|a_n - L| < \epsilon$ 成立，则称数列 $\{a_n\}$ 的极限为 L ，或是 $\{a_n\}$ 收敛于 L 。

尽管可能对于没有基础的读者而言定义上的内容没那么好理解，我们也能从中看出定义引入的几个关键概念：

- 实数 L ：数列逐渐接近的值，也就是极限的目标值
- 正实数 ϵ ：接近极限要求的精度，我们希望数列中的项与极限的距离都小于这个精度
- 正整数 N ：表示从这项开始数列中的每一项都满足与极限 L 的距离小于 ϵ ，也就是当项数大于 N 时，所有的后续项都在 ϵ 的范围内接近 L

我们用一阶逻辑可以将其表示为：

$$(\forall \epsilon > 0)(\exists N \in \mathbb{N})(\forall n \in \mathbb{N})(n > N \Rightarrow |a_n - L| < \epsilon)$$



如左图，数列 $\{a_n\}$ 的项根据公式 $a_n = 1 + \frac{1}{n}$ 给出，随着 n 的增大，数列的值逐渐接近极限 $L = 1$ 。粉红色的区域表示满足 $|a_n - L| < \epsilon$ 的范围，其中 $\epsilon = 1$ ，也就是我们希望数列项与极限值的差的绝对值小于0.1，数列的项必须在区域内才能满足接近极限值的条件。

所以，定义说明了数列的极限是指当数列的项数无限增加时，数列的值趋近于的一个确定的值。随着项数的增加，数列中的项与极限值

之间的差距不断减小，最终趋于无穷接近，也就是当 n 趋近于正无穷时，定义中的表示我们可以以以下两种形式表示

$$\lim_{n \rightarrow +\infty} a_n = L$$

$$a_n \rightarrow L (n \rightarrow +\infty)$$

值得注意的是， ∞ 前的+号代表了正无穷，若为-则是负无穷，在我们不区分正负地表示无穷大的数值时，就忽略这个加减号，也就是 $|n|$ 是正无穷大的。

我们要考虑这个极限 L 是否存在，若一个数列存在有限的极限值 L ，则它是收敛数列，若不存在有限的极限值，则我们称之为发散数列，例如一个没有上界或下界的数列、没有稳定趋势，无法用一个函数来描述其变化的数列。

函数极限

函数极限是函数在自变量趋近于某个值时，函数值的极限，老规矩我们先看定义。

当函数 $f(x)$ 的自变量 x 趋近于某个特定值 a (我听说 a 在定义中更常用) 时，若存在实数 L ，对于任意正实数 ϵ ，存在一个正数 δ ，使得当 $0 < |x - a| < \delta$ 时， $|f(x) - L| < \epsilon$ 成立，可以以这两种形式表示。

$$\lim_{x \rightarrow a} f(x) = L$$

$$f(x) \rightarrow L (x \rightarrow a)$$

例如，当自变量 x 趋近于正无穷时，函数 $f(x)$ 的值趋近于 L ，

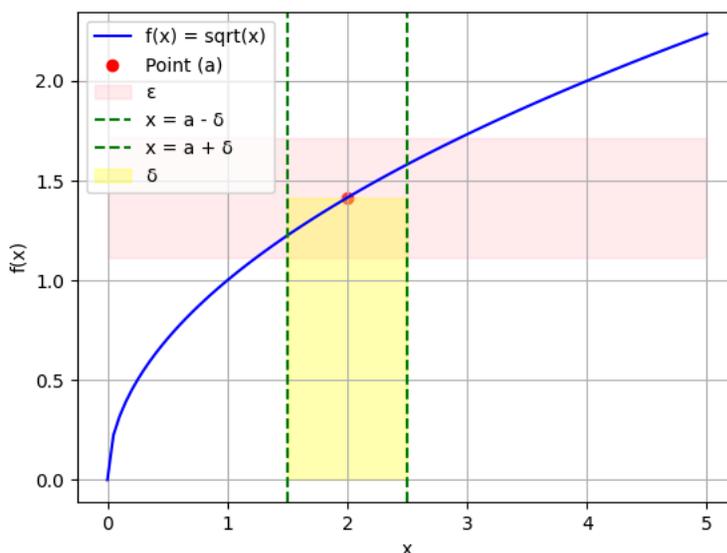
$$\lim_{x \rightarrow \infty} f(x) = L$$

$$f(x) \rightarrow L(x \rightarrow \infty)$$

这边要解释的是一个新的概念 $\epsilon - \delta$ 定义 (epsilon-delta definition)，我们会经常用它来定义极限，精确描述函数趋近某个点的行为方式，假设我们想知道函数在点 a 处的极限值，在这里的点 a 是通常被称为自变量趋近的极限点或收敛点，epsilon是一个正数，作为接近极限要求的精度（和之前解释的用法一样，只是这里是函数），delta则表示自变量与点 a 的距离，也就是自变量足够接近 a 的范围。另外，后面为了方便我就直接叫ed语言了。

看例图，函数解析式为

$f(x) = \sqrt{x}$ ，红色点是我们感兴趣的点 a ，对应函数图象上的 $(2, \sqrt{2})$ 。在粉色区中表示 ϵ ，值取0.3，表现为一个水平的带状区域，位于点 a 的上下方向（表示误差范围，也就是允许误差在 ± 0.3 ）。黄色区域表示 δ ，也就是自变量与点 a 的距离范围，在这个图像中我们的 δ 取0.5，所以，绿色虚线的范围也就是自变量可以变化的范围。



值得注意的是， ϵ 和 δ 是两个独立的变量，分别表示的是函数值和自变量与极限点的距离范围，它们的选择相互关联但没有直接依赖，所以我们可以独立选择 ϵ 和 δ 的值来满足极限定义的要求，但是一般在实际的证明或计算，我们应该要同时考量这两个值的取值。

我们需要进一步细分的是左极限和右极限，继续说我们感兴趣的点 a 。左极限表示 x 的值小于 a ，也就是 x 从左侧接近 a ，表示为

$$\lim_{x \rightarrow a^-} f(x) = L$$

同理，右极限则表示 x 的值大于 a ，也就是 x 从右侧接近 a ，表示为

$$\lim_{x \rightarrow a^+} f(x) = R$$

要注意的是，左极限或右极限存在并不意味着函数 $f(x)$ 在 a 处有极限，只有当左极限等于右极限且与 $f(a)$ 相等时，才能说函数 $f(x)$ 在 a 处有极限。

函数连续性

当我们谈论一个函数在某个点处连续时，可以想象这个函数的图像是一条平滑的曲线，没有任何断裂或跳跃。连续性通俗地说就是函数的平滑性和连贯性。

用 ϵ 语言来描述连续性的定义，设函数 $f(x)$ 在点 c 处有定义，对于任意正实数 ϵ ，存在一个正数 δ ，使得当 $0 < |x - c| < \delta$ 时， $|f(x) - f(c)| < \epsilon$ 成立，那我们就说函数 $f(x)$ 在点 c 处连续。

$f(c)$ 在这里表示函数在 c 处的函数值，也就是说，当 x 与 c 足够接近时， $f(x)$ 和 $f(c)$ 也足够接近，一样，我们可以将 ϵ 看作是我们期望的精度要求，而 δ 则是我们需要找到的对应『能容忍的范围』。这个要求也就是 c 处的极限存在且与函数值 $f(c)$ 相等，函数值点 c 处没有突变或跳跃，它是平滑且连续的。

基于上一节对函数极限的定义（当左极限等于右极限且与 $f(a)$ 相等时），函数在某点处连续，意味着该点的极限存在且与函数值相等，连续性是极限存在性的一种进一步表现。如果一个函数在某点处连续，那么该点的极限也必定存在。

如图，这是一个符号函数 (Signum function)，定义为：

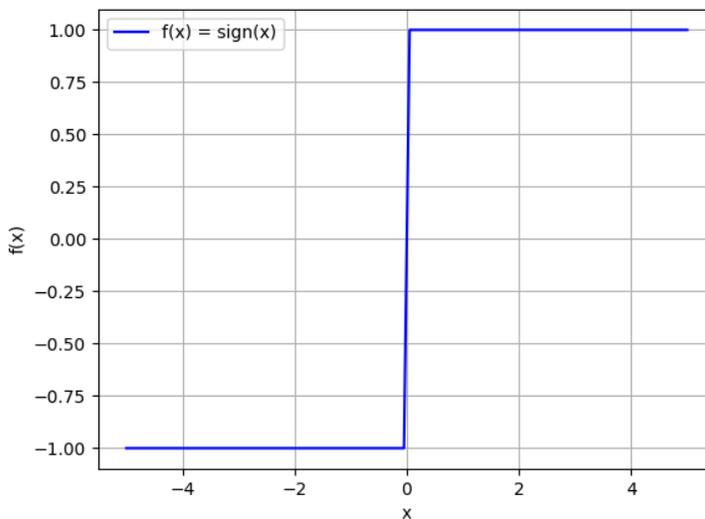
$$\text{sign}(x) = \begin{cases} -1 & (x < 0) \\ 0 & (x = 0) \\ 1 & (x > 0) \end{cases}$$

符号函数是一个用于判断实数正负号的逻辑函数，我们可以观察到，在 $x = 0$ 处，若 x 从左侧逼近0，即 $x < 0$ ，函数值为-1，所以 $\lim_{x \rightarrow 0^-} f(x) = -1$ ；同理，若 x 从右侧逼近0，即 $x > 0$ ，函数值为1，所以

$\lim_{x \rightarrow 0^+} f(x) = 1$ 。因为在 $x = 0$ 处的左右极限不相等，所以函数在 $x = 0$ 处不存在极限，无法满足连续性的要求，所以我们可以说符号函数是一个非连续函数。

这很好理解，这个没有连续性的点被叫做间断点，而间断点还要分几种类型，我们考虑函数自变量 x 趋近的值 a ：

- 可去间断点 (Removable Discontinuity)：若 $\lim_{x \rightarrow a} f(x)$ 存在，但它与 $f(a)$ 不相等，则这个间断点属于可去间断点



- 跳跃间断点 (Jump Discontinuity) : 若 $\lim_{x \rightarrow a} f(x)$ 不存在, 则这个间断点属于跳跃间断点
- 无穷间断点 (Infinite Discontinuity) : 若点 a 的左右极限至少有一个是无穷大, 或者其中一个极限不存在, 则这个间断点属于无穷间断点。

分类解释起来很简单, 先说可去间断点, 这种情况通常是由于函数在该点有一个缺失的点或者一个定义域的空隙, 可以通过修正或填充来使函数在该点连续。跳跃间断点也就是极限不存在而发生的一个跳跃, 也就是我们上面的符号函数那样, 符号函数因为在 $x = 0$ 处的左右极限不相等, 导致函数在这个地方不存在极限, 所以它是一个跳跃间断点。无穷间断点通常表现为函数值无法在这个点附近接近任何有限值, 例如函数 $f(x) = \frac{1}{x}$, 当 x 趋近于0时, 它的右极限是正无穷, 左极限则是负无穷, 这样就是一个无穷间断点。

简单极限求值

极限求值的方法很多, 各种各样的, 本文作为入门文章, 我会介绍一些基本的、常用的极限求值方法。

整体代入法

当我们计算极限时, 最基本的情况是函数在某一点处有定义, 并且可以直接代入该点得到函数值。这种情况下, 我们可以使用代入法来计算极限。

整体代入法, 就是我们把趋近的值代入到极限表达式即可。如果函数是连续函数, 我们就一定可以直接通过这种方式来求极限, 例如, 有极限表达式

$$\lim_{x \rightarrow 2} (x^2 - 4)$$

这种情况下我们直接将 2 代入 x 即可, 求:

$$2^2 - 4 = 4 - 4 = 0$$

得到结论

$$\lim_{x \rightarrow 2} (x^2 - 4) = 0$$

对于左右极限, 例如 $\lim_{x \rightarrow 2^+} (x^2 - 4)$ 和 $\lim_{x \rightarrow 2^-} (x^2 - 4)$ 的情况, 如果这个函数是连续函数, 我们也是直接将趋近的值代入即可。在数学的实际应用中我们研究的大部分都是连续函数, 连续函数在定义域内左右极限都是存在且相等的, 所以可以这样做, 但是你会发现我们在实际上很少需要先求左右极限。

如果是做题, 例如计算单侧极限, 求分段函数极限, 就得分别求单侧极限并判断极限是否存在。例如有分段函数:

$$f(x) = \begin{cases} 2x + 1 & (x < 1) \\ x^2 & (x \geq 0) \end{cases}$$

我们将讨论函数 $f(x)$ 在 $x = 1$ 处的左极限和右极限，当 x 从左侧接近1时，我们有极限表达式 $\lim_{x \rightarrow 1^-} 2x + 1$ ，我们可以直接带入得到 $\lim_{x \rightarrow 1^-} 2x + 1 = 2 \cdot 1 + 1 = 3$ 。同理，当 x 从右侧接近1时，我们可以得到极限表达式 $\lim_{x \rightarrow 1^+} x^2$ ，同样直接带入1得到 $\lim_{x \rightarrow 1^+} x^2 = 1^2 = 1$ ，因为左右极限不相等，所以我们可以说函数 $f(x)$ 在该点的极限不存在。

因子法和分式法

如果一个函数可以被分解为两个函数的乘积，我们则可以考虑使用因子法来计算它的极限。

考虑一个函数 $f(x) = g(x) \cdot h(x)$ ，我们关注的是 x 趋近于某一点 a 时的极限 $\lim_{x \rightarrow a} f(x)$ 。如果存在 $\lim_{x \rightarrow a} g(x) = L_1$ 和 $\lim_{x \rightarrow a} h(x) = L_2$ ，则 $\lim_{x \rightarrow a} f(x) = L_1 \cdot L_2$

例如有极限表达式：

$$\lim_{x \rightarrow 2} (x^2 - 4)$$

我们就能这样计算：

$$\begin{aligned} & \lim_{x \rightarrow 2} (x^2 - 4) \\ &= \lim_{x \rightarrow 2} (x + 2) \cdot \lim_{x \rightarrow 2} (x - 2) \\ &= 4 \cdot 0 \\ &= 0 \end{aligned}$$

如果一个函数涉及到分数形式，我们就可以将极限转化为一个分式然后简化运算。

考虑极限表达式 $\lim_{x \rightarrow a} \frac{f(x)}{g(x)}$ ，我们关注的是 x 趋近于某一点 a 时的极限，则可以直接应用分式法：

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

例如有极限表达式：

$$\lim_{x \rightarrow 1} \frac{(x + 1)(x - 1)}{(x - 1)}$$

则应用分式法：

$$\begin{aligned}
& \lim_{x \rightarrow 1} \frac{(x+1)(x-1)}{(x-1)} \\
&= \lim_{x \rightarrow 1} (x+1) \\
&= 1+1 \\
&= 2
\end{aligned}$$

复合函数法

对于复合函数的极限求值，我们可以应用复合函数法。

给定两个函数 $f(x), g(x)$ ，将 $g(x)$ 的输出作为 $f(x)$ 的输入的函数，表示为 $(f \circ g)(x) = f(g(x))$ ，我们称其为复合函数。我们关注的是 x 趋近于某一点 a 时的极限 $\lim_{x \rightarrow a} f(g(x))$ ，则复合函数的极限为：

$$\lim_{x \rightarrow a} f(g(x)) = \lim_{u \rightarrow L} f(u)$$

其中

$$L = \lim_{x \rightarrow a} g(x)$$

例如有函数 $f(x) = \sqrt{x}, g(x) = x^2$ ，关注极限表达式 $\lim_{x \rightarrow 2} f(g(x))$ ：

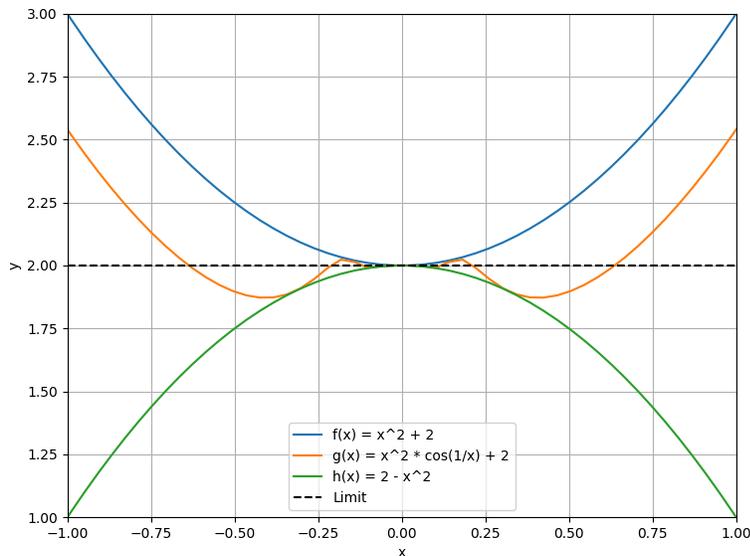
$$\begin{aligned}
\lim_{x \rightarrow 2} f(g(x)) &= \lim_{u \rightarrow \lim_{x \rightarrow 2} x^2} \sqrt{u} \\
&= \lim_{u \rightarrow 4} \sqrt{u} \\
&= 2
\end{aligned}$$

三明治定理

在写这一段时，我打 夹逼定理 被和谐了，咱就是说啊，心脏的东西看什么都脏（恼），没事，接下来我们继续将其叫做夹逼定理

夹逼定理是一种求极限的方法，当某些函数中某一点附近被夹在两个趋向相同的函数之间时，我们就可以利用夹逼定理来求极限。

若有函数 $f(x), g(x), h(x)$ 在某个点 c 的某个邻域内除了 c 外都有定义，且满足对于所有 x 在该邻域内， $f(x) \leq g(x) \leq h(x)$ ，且当 x 趋向 c 时， $f(x)$ 和 $h(x)$ 都趋向同一个极限 L ，则 $g(x)$ 也趋向极限 L 。



如图，三条曲线分别代表函数

- $f(x) = x^2 + 2$
- $g(x) = x^2 \cos\left(\frac{1}{x}\right) + 2$
- $h(x) = 2 - x^2$

Latex行内公式太乱吵到我眼睛了，这里分点写

其中 $g(x)$ 被 $f(x)$ 和 $h(x)$ 夹在中间，且当 x 趋近于0时，函数 $f(x)$ 和 $h(x)$ 的函数值都趋向于极限值 $L = 2$ ，根据夹逼定理， $g(x)$ 的函数值也趋向极限 L ，表示为 $\lim_{x \rightarrow 0} g(x) = 2$

夹逼定理可以在不求出确切的极限值的情况下证明一个极限的存在，我们只需要两个收敛于相同极限 L 的函数 $f(x)$ 和 $h(x)$ ，看函数是否被『夹在』它们之间，即可证明极限存在且等于 L 。这很利于我们解决不定式的极限问题，例如什么 $\frac{0}{0}$ ， $\frac{\infty}{\infty}$ 这样的不确定形式，我们可以将其转化为边界函数极限的比较。

因为还没写到导数，所以 L'Hôpital法则，Taylor展开等涉及到导数知识的求极限方法我们在后面的章节再了解。

变化率与导数

坡度

在数学中，坡度是描述曲线、曲面的陡峭程度的概念，所以也可以叫做斜率。

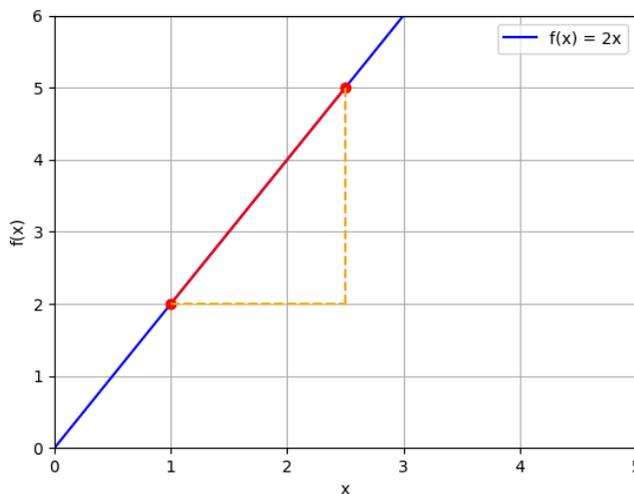
图中是一个函数 $f(x) = 2x$ ，很简单的一个直线函数，其中有两个点 $(1, 2)$ 和 $(\frac{5}{2}, 5)$ ，其中垂直方向的虚线 l_1 代表了它们两个纵轴的差，水平方向的虚线 l_2 代表横轴的差。若我们将点 $(1, 2)$ 表示为 (x_1, y_1) ；点 $(\frac{5}{2}, 5)$ 表示为 (x_2, y_2) ，它的坡度 m 可以表示为

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{l_1}{l_2}$$

进行一个简单的代入计算，也就是

$$m = \frac{5 - 2}{\frac{5}{2} - 1} = 2$$

进一步地，我们要求曲线在某一点的倾斜程度，也就是变化率。



导数与切线

本章写于 HongKong 国际机场的候机室，凌晨两点，明天早上的飞机，Pacific coffee的冰美式不错。

谈到导数的定义，我们可以从之前的极限开始看起，假设函数 $f(x)$ 在点 $x = a$ 处有定义，函数中该点处的导数就可以表示为

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

其中， h 代表 x 在 a 点的偏移量，当我们求函数 $f(x)$ 在点 $x = a$ 处的导数时，我们引入了偏移量 h 并考虑 $a + h$ ， h 趋近于0，所以我们代入得到的就是一个 $x = a$ 处偏移一个微小距离 h 后的函数值，然后我们再计算差分值 $f(a+h) - f(a)$ ，即可得到此处的变化量，最后我们将其除以偏移量 h ，即可得到函数 $f(x)$ 在点 $x = a$ 处的导数。 h 趋近于0，但不等于0，因为我们求导时希望得到的是函数在此处的瞬时变化率，而不是整个邻域的平均变化率。

当然，我们除了用 $f'(a)$ 表示函数 $f(x)$ 在点 $x = a$ 处可导，若因变量为 y ，我们还能用我们熟知的 $\frac{dy}{dx}$ 或者更一般的表达方式 $\frac{df(a)}{dx}$ 来表示，这是莱布尼茨记法，这么做的好处是明确自变量和因变量，但是我们要注意的是此处的记号 dx, dy 是一个整体。

切线的概念也不过如此，导数实际上就是切线的斜率，这条切线通过并与函数 $f(x)$ 在点 $x = a$ 处

相切，我们假设这个点为点 P ，根据前面那个偏移量 h 趋近于0我们可以得知，切线与曲线在相切点附近非常接近，距离在无限接近点 P 时逐渐趋近0（但是一样，不会等于0）我们可以使用点斜式方程表示切线：

$$y = f(a)(x - a) + f(a)$$

求导规则

所以我们现在要考虑如何求导数，为了简化求导过程，我们需要了解一些基本的求导规则和公式。

常数项规则

因为常数不会因自变量变化而变化，所以常数项的导数为0。

线性项规则

线性项的导数为其系数，例如 $f(x) = 2x$ 的导数 $f'(x) = 2$ 。

我们考虑函数 $f(x) = nx$ ，我们将其带入定义 $f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$ ，得到

$$f'(x) = \lim_{h \rightarrow 0} \frac{n(x+h) - nx}{h}$$

简化公式

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{nx + nh - nx}{h} \\ &= \lim_{h \rightarrow 0} \frac{nh}{h} \\ &= n \end{aligned}$$

所以线性项的导数与其系数相等。

幂规则

幂项的导数等于指数乘以原项的指数减一次幂，也就是函数 $f(x) = x^n$ 的导数为 $f'(x) = nx^{n-1}$

和差规则

若一个函数是多个函数加减组成，则可以对每个函数分别进行求导再进行加减。

例如，考虑函数 $f(x) = 3x^2 - 2x + 1$ ，分别对每个项求导，得到该函数的导数为 $f'(x) =$

$$6x - 2 + 0 = 6x - 2$$

若 $u = u(x), v = v(x)$ 并且它们都可导, 则和差规则的应用表示为

$$(u \pm v)' = u' \pm v'$$

乘积规则

若一个函数是两个函数的乘积, 则我们对其中一个函数求导, 再将其乘另外一个原函数的结果, 与另外一个函数的求导结果与这个函数的乘积相加。

若 $u = u(x), v = v(x)$ 并且它们都可导, 则乘积规则的应用表示为

$$(uv)' = u'v + uv'$$

商规则

若一个函数是一个函数除以另一个函数, 则我们可以称之为商函数, 其导数为分子的导数与分母的积, 减去分母的导数与分子的积, 除以分母。

若 $u = u(x), v = v(x) (v \neq 0)$ 并且它们都可导, 则商规则的应用表示为

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

常数倍规则

如果一个函数是另一个函数的常数倍, 则其导数也存在相同的比例关系, 等于常数与原函数导数的乘积。

若 $u = u(x)$ 且可导, N 为常数, 则常数倍规则的应用表示为

$$(Cu)' = Cu'$$

指数函数规则

对于指数函数 $f(x) = a^x$, 其中 a 是常数且 $a > 0, a \neq 1$, 则指数函数规则的应用表示为

$$f'(x) = a^x \cdot \ln(a)$$

特殊地, 对于自然对数 e , 导数为它自身 $f'(e^x) = e^x$ 。

链式法则

我们为复合函数求导的法则叫做链式法则, 复合函数的导数等于 外层函数对内层函数的导数 乘 内层函数的导数。

若函数 $f(x), g(x)$ 都可导, 则链式法则的应用可以表示为

$$(f(g))'(x) = f'(g(x)) * g'(x)$$

中值定理

拉格朗日中值定理

如果一个函数在闭区间上连续，在开区间上可导，那么在这个区间内至少存在一点，该点的导数等于函数在该区间两端点处的平均变化率。

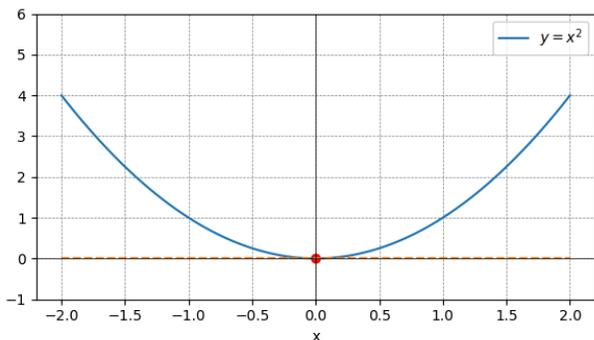
考虑函数 $f(x)$ 在区间 $[a, b]$ 上连续，且在 (a, b) 内可导，则至少存在一个 $c(a < c < b)$ 使得

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

这意味着如果一个函数在区间是连续且可微分的，那在这里面至少存在一个点使得函数在这个点上的切线斜率等于函数在整个区间的平均变化率。

罗尔定理

更特殊地，若在拉格朗日中值定理的基础上，两个端点的函数值相等，则必然存在一个点使导数为零。



考虑函数 $f(x)$ 在区间 $[a, b]$ 上连续，在 (a, b) 内可导且满足 $f(a) = f(b)$ ，则至少存在一个 $c(a < c < b)$ 使得 $f'(c) = 0$ 。

直观地，例图是一个很简单的抛物线，而这个 c 存在于开口向上抛物线的极小值，显而易见地，函数的极大值或极小值导数为0，这里的点 c 符合我们的定义。

柯西中值定理

考虑函数 $f(x), g(x)$ 在区间 $[a, b]$ 上连续，在 (a, b) 内可导且满足 $g'(x) \neq 0$ ，则存在 $c(a < c < b)$ 使得

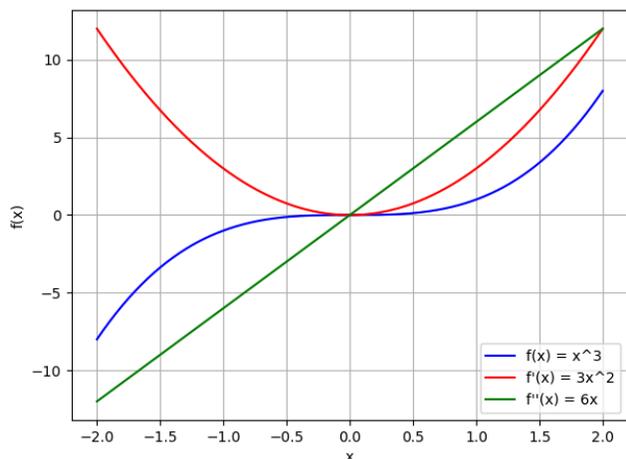
$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}$$

实质上这就是拉格朗日中值定理的推广，推广到两个函数在某一区间内的平均变化率与导数的关系。

二阶导数和凹凸性

我们可以使用二阶导数来描述函数的曲线的曲率和凹凸性，二阶导数就是对一个函数一阶导数的求导结果。对于函数 $f(x)$ 的二阶导数，我们可以将其记为 $f''(x)$ ，或者用莱布尼兹记法，因为 $\frac{d}{dx}$ 表示对 x 求导，所以若因变量为 y ，二阶导数我们可以记为 $\frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2}$ 。

在微积分中， d 表示微分或导数的运算符，代表一个无穷小的增量，所以我们在第二次求导后的 d^2y 并不是指 $d \cdot d \cdot y$ ，而是表示对函数进行二次微分。



图中，我们有函数 $f(x) = x^3$ ，我们可以运用前面所讲的内容对其进行求导，分别求一阶导数和二阶导数：

- $f'(x) = 3x^{3-1} = 3x^2$
- $f''(x) = 3 * 2x^{2-1} = 6x$

对应的求导结果也在图中了，图示中一阶导数描述了函数值每个点的变化率，在图像中它是一个递增的抛物线，开口朝上。这表示函数图像在每个点的斜率随着 x 的增加而增加。

二阶导数是一条斜率为正的直线，它反映了函数的凹凸性。对于函数 $f(x) = x^3$ ，在区间 $x > 0$ 内的曲线呈向下弯曲，开口向上的形状，所以在此区间的函数是凹性 (Concavity) 的。同时你可以观察二阶导数 $f''(x) = 6x$ ，在这个区间它是非负的，这反应了函数在该区间的凹凸性为凹性。同理，在这个函数的区间 $x < 0$ 内 函数呈向上弯曲，开口向下的形状，且 $f''(x) = 6x$ 是非正的，所以函数在该区间的凹凸性为凸性。

极值

在数学中，极值 (extremum) 是指函数在某个区间或定义域上取得的最大值或最小值，分别叫极大值 (maximum) 和极小值 (minimum)。根据这个区间，可能是邻域或是整个函数域，我们先看定义。

设函数 $f(x)$ 在点 $x = c$ 处的某个邻域内有定义：

- 若存在一个正数 ϵ 使得所有满足 $|x - c| < \epsilon$ 的 x 都满足 $f(c) \geq f(x)$ ，则这个点 c 对应的函数值 $f(c)$ 就是函数 $f(x)$ 的局部极大值。
- 若存在一个正数 ϵ 使得所有满足 $|x - c| < \epsilon$ 的 x 都满足 $f(c) \leq f(x)$ ，则这个点 c 对应的函数值 $f(c)$ 就是函数 $f(x)$ 的局部极小值。

这里的 ϵ 其实就是区间的定义，这种指定特定区间的就叫做局部极值 (Local Extrema)，这些极

值点只在该区间内成立，并不一定是整个函数的最大值或最小值。

- 若点 c 对于整个函数域的所有 x 都满足 $f(c) \geq f(x)$ ，则这个点 c 对应的函数值 $f(c)$ 就是函数 $f(x)$ 的全局极大值。
- 若点 c 对于整个函数域的所有 x 都满足 $f(c) \leq f(x)$ ，则这个点 c 对应的函数值 $f(c)$ 就是函数 $f(x)$ 的全局极小值。

与局部极值不同，全局极值考虑了函数在整个定义域内的取值范围，而不仅仅局限于某个特定区间。全局极大值/极小值是函数在整个定义域内的最高点/最低点，没有其他点的函数值能够超过/低过它。

二阶导数求极值

在某些情况，我们可以直接使用二阶导数来求函数的极值。我们通过求解方程 $f'(x) = 0$ 得到函数的驻点的横坐标，并计算该驻点的二阶导数：

- $f''(x) > 0$ ，该驻点为极小值点
- $f''(x) < 0$ ，该驻点为极大值点
- $f''(x) = 0$ ，无法使用二阶导数确定该点的极限

我们可以这么做是因为，若函数则意味着函数在该点的局部范围内从下向上走势逐渐增大在某点的导数为0，则在这个驻点处函数的斜率为0，也就是切线是水平的，而在极值点附近，函数在变化方向上达到了最大值或最小值，所以导数为0（所以驻点才有可能可能是极值点，但是它也有可能是函数的拐点或鞍点等，所以需要进行计算来判断）。

更高阶的导数同理，只需要连续地对函数进行求导即可。

偏导数与梯度

偏导数是多元函数对其中一个自变量的偏导数。考虑函数 $f(x_1, x_2, \dots, x_n)$ ，第 i 个自变量 x_i 的偏导数表示为 $\frac{\partial f}{\partial x_i}$ ，也就意味着是多元函数在某一方向上的变化率（这意味着它在其他自变量上保持不变），同理地，它也表示在多元函数上的某一点的切线斜率。

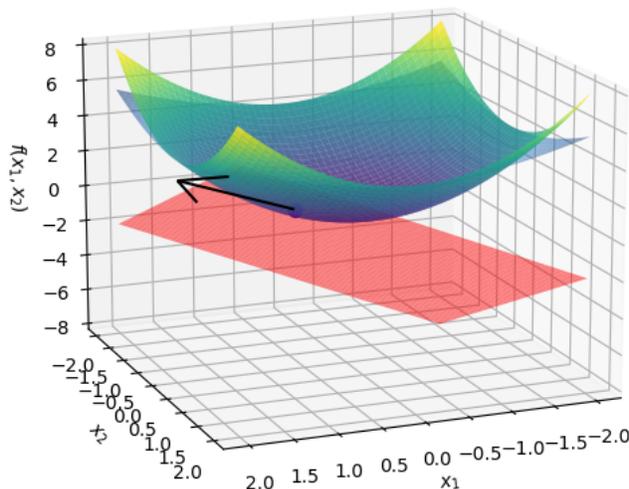
因为多元函数有多个自变量，所以就有一个叫梯度（Gradient）的概念，以向量的形式描述中多元函数上某一点的导数，其中每个分量就是函数在各个方向上的偏导数。对于一个 n 元函数 $f(x_1, x_2, \dots, x_n)$ ，梯度表示为：

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

例图的曲面是函数 $f(x_1, x_2) = x_1^2 + x_2^2$ 。黑色箭头是点 $(1, 1)$ 处的梯度矢量，红色平面为点 $(1, 1)$ 处通过偏导数构建出的切平面。淡蓝色曲面是函数在每个点处的梯度大小，其计算公式为：

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\right)^2}$$

$$= 2\sqrt{x_1^2 + x_2^2}$$



混合偏导数和施瓦茨定理

混合偏导数是多元函数的高阶偏导数，当一个多元函数关于不同自变量的偏导数被连续地取得时，我们可以考虑它们的次序，并可以得到混合偏导数。例如 $\frac{\partial f}{\partial x_1 \partial x_2}$ 表示先对 x_1 求偏导数，再对 x_2 求偏导数； $\frac{\partial^2 f}{\partial x_1^2}$ 表示对 x_1 进行两次偏导数，类似地也能推广到更高阶的混合偏导数，涉及对不同变量的多次求偏导数。

若一个函数 f 某一点的混合偏导数存在且连续，那么这些混合偏导数的次序可以交换，这就是 Schwarz 定理，例如一个二元函数 $f(x_1, x_2)$ ，则可以表示为

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1}$$

我们在求混合偏导数时可以使用这个定理有效地简化计算。

更多求极限法

讲完导数后，我们就能了解更多涉及到导数知识的求极限的方法。

L'Hôpital's 法则

偶尔听到饱受高数折磨的朋友说『（此处省略文明用语）洛就完了！』，对于求极限时遇到例如 $\frac{0}{0}$, $\frac{\infty}{\infty}$ 这样的不定形式，我们就能使用洛医院洛必达法则将其转化为一个导数的极限。

假设函数 $f(x)$ 和 $g(x)$ 在某一点 a 的附近可导，当 x 趋近于 a 时， $f(x)$ 和 $g(x)$ 的极限都为 0 或 $\pm\infty$ ，如果它们的导数 $f'(x)$ 和 $g'(x)$ 在该点存在且 $g'(x) \neq 0$ ，则原函数的极限可以被转化为导数的极限：

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

其实你可以这样理解，当函数的某个点 a 的行为趋近于不确定形式时，我们求极限可以关注函数的切线的斜率。如果 $f(x)$ 和 $g(x)$ 在点 a 附近的行为非常接近，那么它们的导数 $f'(x)$ 和 $g'(x)$ 在该点的行为也应该非常接近，所以我们可以通过计算导数的极限来近似原函数的极限。

这里你应该注意的是近似，洛必达法则并不难保证在所有情况下都能得到准确的极限值，我们在使用洛必达法则时应该注意它只适用于特定类型的不定形式（也就是上面定义的那些），且我们可以连续应用洛必达法则直到得到一个明确的结果，或者无法再应用。

Taylor expansion

我们可以使用泰勒展开（Taylor expansion）将函数表示为无穷级数，也就是按照一定规律相加得到的无穷和，来近似表示原始函数的行为，进一步地我们可以用其来计算函数的极限。

考虑光滑函数 $f(x)$ ，我们以点 $x = a$ 处为中心进行展开，则我们将函数表示为无穷级数的形式，函数在该处的值、一阶导数、二阶导数以此类推，每个阶次对应一个贡献程度。

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

因为级数是按照一定规律相加得到的无穷和，所以我们可以写成

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

在实际应用中，我们只需要将函数从 a 处展开到足够满足我们所需精度的项数（在定义中是无穷，实际按精度需求展开有限项数即可），再将展开式带入极限计算中，即可得到极限的近似值。

积分

定积分

定积分是一种积分，表示曲线下某一区间的面积，将曲线的函数表达式表示为一个区间的积分表达式。

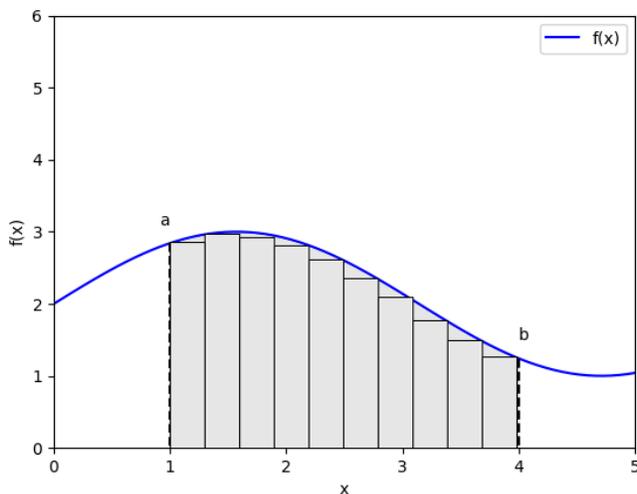
考虑函数 $f(x)$ 在区间 $[a, b]$ 内连续 (如右图), 则我们可以这样表示定积分:

$$\int_a^b f(x)dx$$

在定义中, $f(x)$ 是被积函数, dx 表示微元, 这意味着就是将区间 $[a, b]$ 划分为无数个小区间, 子区间长度的最大值为 $\max(x_i + 1 - x_i)$, 其中 $0 \leq i \leq n - 1$, 它们函数值与区间长度的乘积并求和就是面积。

我们在区间 $[a, b]$ 划分 n 个小区间后, 对于每个小区间 $[x_i, x_{i+1}]$ 选择一个代表点 ξ_i (这是希腊字母 ξ 的小写), 这可以是区间的左端点、右端点等, 总之在区间内即可。然后我们求得每个代表点 ξ_i 处的函数值 $f(\xi_i)$, 即可计算每个子区间的矩形面积, 按照定义我们可以得知每个子区间宽度为 $\Delta x_i = x_{i+1} - x_i$, 长度, 也就是 $\Delta x_i = f(\xi_i)$, 所以我们可以得到定积分的和式为:

$$\int_a^b f(x)dx = \sum_{i=0}^n f(\xi_i) \cdot \Delta x_i$$



右图的矩形是我懒得写matplot直接在ps拉的, 所以可能有点不齐, 自动忽略吧 (

值得注意的是, 这个代表点 ξ_i 作为代表积分区间的点, 意味着它是对被积函数进行取样的点, 对这个代表点的选择会影响到结果的近似程度, 例如上图的每个矩形与函数 $f(x)$ 接触到的点就是这个点, 因为比较粗糙 (区间分的少, 以及代表点只是单纯一个矩形往上拉到接触曲线的点) 所以这个图的误差是显而易见的。

Riemann 可积

一个函数被称为可积 (integrable) 通常指的是它在某个给定的区间上可积, 也就是说, 该函数的积分存在且有限, 这边我们应该先明确一下函数可积的条件, 之后我们的讨论默认为黎曼积分。

考虑函数 $f(x)$, 定义在闭区间 $[a, b]$ 上 (之后不再强调闭区间), 存在一个数 I , 使得对于任意给定的 $\epsilon > 0$, 都存在一个 $\delta > 0$, 使得当划分 $[a, b]$ 的任意一组标号为 x_0, x_1, \dots, x_n , 且每个子区间的长度都小于 δ 时, 对应的黎曼和 $R(f, x_0, x_1, \dots, x_n)$ 满足

- $\lim_{n \rightarrow \infty} R(f, x_0, x_1, \dots, x_n) = I$, 这个条件可以理解为, 当我们将闭区间 $[a, b]$ 划分为越来越多的小子区间, 并计算这些子区间上的黎曼和时, 这些黎曼和的极限值趋近于一个特定的数值 I , 这个 I 就是函数 $f(x)$ 在闭区间 $[a, b]$ 上的黎曼积分。
- 存在一个正整数 N , 对于所有的 $n > N$, 都有 $|R(f, x_0, x_1, \dots, x_n) - I| < \epsilon$, 简单的说就是按照之前对定积分的说法, 实际的积分中会存在一定误差, 而这个误差应该被

控制住一个范围内，这个 ϵ 就是对黎曼积分精度的要求。

必要条件

黎曼可积的必要条件是被积函数 $f(x)$ 在 $[a, b]$ 上有界。

充分条件

- 若被积函数 $f(x)$ 在 $[a, b]$ 上连续，则定积分存在，它是黎曼可积的
- 若被积函数 $f(x)$ 在 $[a, b]$ 上有界，且间断点有限，则定积分存在，它是黎曼可积的
- 若被积函数 $f(x)$ 在 $[a, b]$ 上单调，也就是函数在 $[a, b]$ 上单调增加/减少，则定积分存在，它是黎曼可积的

弧长与积分

弧长是曲线上的一段长度，我们可以使用定积分来计算弧长。对于连续且可导的平面曲线 $f(x)$ ($a \leq x \leq b$)，弧长 L 可以直接通过定积分计算：

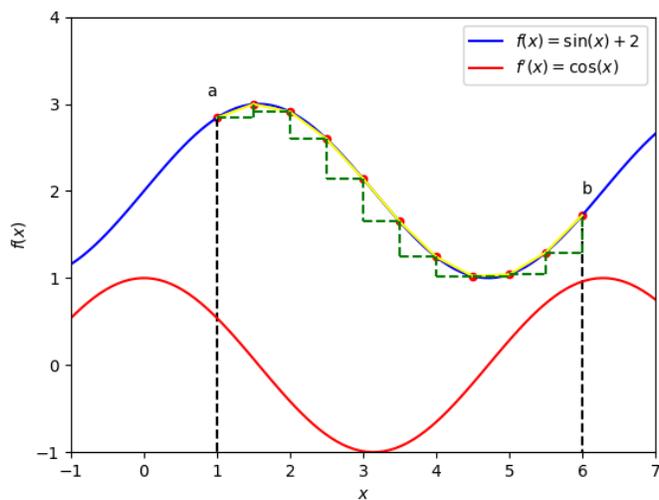
$$L = \int_a^b \sqrt{1 + [f'(x)]^2} dx$$

这个公式基于勾股定理，通过对弧微分进行积分，可以得到整段曲线的弧长。

例图展示的是使用定积分计算弧长的图像，公式就是这样计算近似某一段弧长的，当我们将我们计算的目标弧长分为无数个这样微小的一段然后相加，得到的就是在区间内的弧长。图中的函数是 $f(x) = \sin(x) + 2$ ，则它的一阶导数 $f'(x) = \cos(x)$ ，则公式应用到这里就变成了：

$$L = \int_a^b \sqrt{1 + \cos^2(x)} dx$$

当然，就算曲线是使用参数方程表示的，我们仍可以使用弧长公式来计算它的长度，我们考虑参数方程



$$\begin{cases} x = f(t) \\ y = g(t) \end{cases} \quad a \leq t \leq b$$

其弧长 L 的表达式如下，为了方便区分这里使用莱布尼兹记法。

$$L = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt$$

思想一样，但是在参数方程中我们要同时考虑到曲线在 $f(x)$ 和 $g(x)$ 两个方向上的弧长，我们用导数来考虑自变量 t 的变化对曲线在平面上的弧长的贡献即可。

不定积分

不定积分是另一种积分，不同于定积分，不定积分是不指定区间地对一个函数积分，它的结果是一个函数，称为原函数。

若 $f(x)$ 是被积函数，结果的原函数是 $F(x)$ ，则它满足 $F'(x) = f(x)$ ，（此处在下节微积分基本定理会补充）但是在定义中不定积分是没有设定区间的，对于一个函数 $f(x)$ 的不定积分而言，它存在无穷个在导数运算下是等价的原函数，所以不定积分的结果会与导函数相差常数 C ，我们应该将不定积分表示为：

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

你会发现，如果 $F(x)$ 是函数 $f(x)$ 的一个原函数， $F'(x) = f(x)$ 成立，它是导数运算的逆运算，所以不定积分也可称为反导函数 (Antiderivative)。

分部积分法

分部积分法是一种积分方法，用于计算两个函数的乘积的不定积分。

考虑两个函数 u, v 可导，同理， du, dv 分别是它们的导数，公式表示为：

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

这个公式来自乘积规则，这么一来，我们就能将一个乘积的不定积分转换为两个简单函数的乘积的形式，将 u 从不定积分中消掉，求出 v 的原函数，其实质就是对乘积规则的两边同时进行不定积分。

微积分基本定理

第一定理

考虑函数 $f(x)$ 在区间 $[a, x]$ 上连续, 其中 a 是常数, 其原函数 $F(x)$ 定义为:

$$F(x) = \int_a^x f(x)dt$$

实质上这就是说明, $F(x)$ 的导数等于 $f(x)$

$$F'(x) = f(x)$$

Newton-Leibniz 公式 (第二定理)

牛顿-莱布尼兹公式, 微积分第二基本定理, 描述了原函数 (不定积分) 和函数的定积分之间的联系。

考虑函数 $F(x)$ 为 $f(x)$ 的原函数, 也就是 $F'(x) = f(x)$ 且 $f(x)$ 在区间 $[a, b]$ 上连续, 则公式可以表示为:

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

也就是说, 函数 $f(x)$ 在闭区间 $[a, b]$ 上的定积分等于其原函数 $F(x)$ 在区间端点处的值的差, 这么一来, 我们可以通过找到函数的原函数来计算定积分。

Simpson's 法则

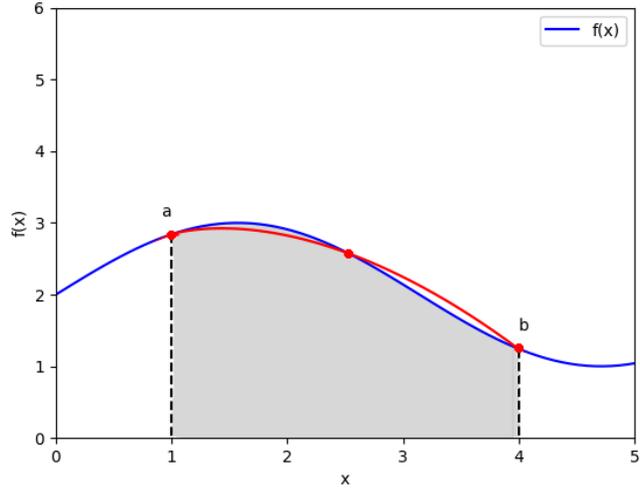
我们还能用辛普森法则近似计算函数的定积分值, 它也是数值积分的方法, 牛顿-柯特斯公式的其中一个公式 (另外一个为梯形法则) 类似地, 辛普森法则也是将积分分为若干个子区间, 并且每个小区间使用二次多项式逼近函数曲线。

将积分区间 $[a, b]$ 均分为 n 个子区间 $[a_i, a_{i+1}]$, n 为偶数, 每个子区间宽度为 $h = \frac{b-a}{n}$, 在每个子区间选三个等距的点, 分别是起点, 中点和终点, 分别为 $a_i, \frac{a_i+a_{i+1}}{2}, a_{i+1}$, 有些读者可能已经察觉到了, 辛普森公式的思想就是以抛物线逼近函数。

我们先考虑辛普森公式简单的形式，这是一种适用于被积函数曲线较为平滑，函数值区间变化较为均匀情况的特例。在这里我们并没有继续分子区间，而是直接取两个端点 a, b 以及区间中点 $4f(\frac{a+b}{2})$ 。

$$\int_a^b f(x)dx \approx \frac{1}{3} \frac{b-a}{2} \left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right]$$

$$= \frac{b-a}{6} \left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right]$$



右图是辛普森法则最简单用法的抛物线，虽然在这个函数图像很显然没有那么精准，但我们对每个子区间进行计算后求和是同理的。如果我们对每个子区间应用这样的近似方法，则表示为：

$$\int_{a_i}^{a_{i+1}} f(x)dx \approx \frac{a_{i+1} - a_i}{6} \left[f(a_i) + 4f\left(\frac{a_i + a_{i+1}}{2}\right) + f(a_{i+1}) \right]$$

所以我们可以得知求和后的公式，就是整个区间 $[a, b]$ 近似的积分值

$$\int_a^v f(x)dx \approx \sum_{i=0}^{n-1} \frac{a_{i+1} - a_i}{6} \left[f(a_i) + 4f\left(\frac{a_i + a_{i+1}}{2}\right) + f(a_{i+1}) \right]$$

蒙特卡洛积分

当我们想要计算某个函数在一个特定区域上的积分时，特别是对于高维或复杂的积分，有时候真的很难求解，所以我们要考虑另外一种方向，我们可以在特定的区域内进行随机抽样，然后通过对函数值求平均得到积分的近似值，这是一种叫做蒙特卡洛积分的数值方法。

考虑一个连续型随机变量 X 的样本空间 D ，定义 $Y = f(X)$ ，其概率密度分布函数是 $p(x)$ 满足 $p(x) \geq 0$ and $\int p(x)dx = 1$ ，然后我们考虑离散的值 x_i 和其概率 p_i ，则我们对随机变量和因变量的期望为：

$$E[X] = \int_D xp(x)dx$$

$$E[Y] = E[f(X)] = \int_D f(x)p(x)dx$$

上式中在对概率密度函数 $p(x)$ 的条件 $p(x) \geq 0$ and $\int p(x)dx = 1$ 下 x_i 在 $[a, b]$ 上均匀分布, 所以我们可以通过在 D 内随机生成样本点 $x_i, i = 1, 2, \dots, N$ 并计算对应的函数值 $f(x_i), i = 1, 2, \dots, N$, 于是我们可得在此范围积分的估计值:

$$\int_D p(x)dx \approx \frac{1}{N} \sum_{i=1}^N f(x_i)$$

符合大数定律, N 越大, 估值就越精确。

一些链接

特别推荐给初中或小学及以下数学基础的读者, 非常通俗的普林斯顿微积分读本:

<http://images.china-pub.com/ebook195001-200000/196991/D1-6Z.pdf>

Wiki: *Limit of a function*: https://en.wikipedia.org/wiki/Limit_of_a_function

Wiki: *Derivative*: <https://en.wikipedia.org/wiki/Derivative>

Wiki: *Integral*: <https://en.wikipedia.org/wiki/Integral>

积分的概念與反導函數: <http://www.ycvs.ntpc.edu.tw/ezfiles/0/1000/img/120/CH3-5.pdf> \

MAKIROR gzanan@gmail.com 14:24 20/07/2023

V2 Updated 14/11/2023 (Posted on 05/12/2023)