数学拾零与思考

Jerry Chaos

2025 年 11 月 13 日

Chapter 1

‡ name of this chapter

1.1 思考

数学中好多问题的解答，可能更希望是乘除运算而非加减运算，而一个简单的加减变乘除的方法就是：提取公因数。下面就是核心的用法：

$$ab - 13a = 0 \tag{1.1}$$

$$a(b - 13) = 0 \tag{1.2}$$

看似简单，但我们需要用的灵活，比如：

$$ab - 13a - 13b = 0 \tag{1.3}$$

$$a(b - 13) - 13b = 0 \tag{1.4}$$

$$a(b - 13) - 13(b - 13) = 169 \tag{1.5}$$

$$(a - 13)(b - 13) = 169 \tag{1.6}$$

如果说题目限定了a与b都是整数的话，那么就可以分解169为整数的乘积，从而分类讨论。

相加叫和，相减叫差，相乘叫积，相除叫商，商的整数部分叫模，余数部分叫余，指数结果叫幂，对数结果就叫对数。

欧拉公式 李永乐老师

$$e^{i\theta} = cos\theta + isin\theta \tag{1.7}$$

$$sin(a + b) = sina * cosb + cosa * sinb \tag{1.8}$$

$$cos(a + b) = cosa * cosb - sina * sinb \tag{1.9}$$

$$x = e^{lnx} \tag{1.10}$$

$$x = ln(e^x) \tag{1.11}$$

Lambert W function：$W(xe^x) = x$

我们可以看这个视频：Harvard Entrance Examination Question — Solve for ‘b’。一般在题目中突然冒出来一个魔法值（具体的数字），大概率意味着它有变化，要么拆成几个数字之和，要么拆成乘积，具体要看整个表达式的表现形式，比如这个click here

$$x^2 - x^3 = 12 \tag{1.12}$$
$$x^2 - 2^2 - x^3 - 2^3 = 0 \tag{1.13}$$
$$(x-2)(x+2) - (x^3 + 2^3) = 0 \tag{1.14}$$
$$\text{Extract the common factor:} \tag{1.15}$$
$$(x+2)[(x-2) - (x^2 - 2x + 2^2)] = 0 \tag{1.16}$$

在解决数学题期间，我们需要尽可能的“拼凑”整齐的表达式，这些表达式可能是既有公式的形式，也可能是便于计算的形式（W函数是Lambert W函数，与ln或者log计算一样，可以使用计算器求得，计算器中一般叫做lambertW或者productLog函数，故此答案中直接使用W是能接受的）。下面的计算中在两边同时乘以ln(2)，拼凑了完整的LambertW函数参数的形式：

$$(5-x)2^{(5-x)} = A \tag{1.17}$$
$$(5-x)(e^{\ln(2)})^{(5-x)} = A \tag{1.18}$$
$$\ln(2)(5-x)e^{\ln(2)(5-x)} = A\ln(2) \tag{1.19}$$
$$W\left[\ln(2)(5-x)e^{\ln(2)(5-x)}\right] = W(A\ln(2)) \tag{1.20}$$
$$\ln(2)(5-x) = W(A\ln(2)) \tag{1.21}$$
$$x = 5 - \frac{W(A\ln(2))}{\ln(2)} \tag{1.22}$$

1.2 verlet integration

here

Verlet integration与Euler intergration的本质区别是速度的计算，正因为Euler方法存在误差，是因为其使用了“瞬时速度”去计算“存在加速度下的一段时间”的距离，无论使用开始时刻还是结束时刻的瞬时速度，都是不准确的。而Verlet integration是利用此刻与前一刻的位置变化去计算速度，这种计算出来的速度是“平均速度”。

$$x_{n+1} = x_n + (v_n + a\Delta t)\Delta t \tag{1.23}$$
$$x_{n+1} = x_n + (\frac{x_n - x_{n-1}}{\Delta t} + a\Delta t)\Delta t \tag{1.24}$$
$$x_{n+1} = 2x_n - x_{n-1} + a\Delta t^2 \tag{1.25}$$

韦氏积分只需要保存前一次与本次的位置，即可计算未来的位置，不在需要保存速度信息。有人想到使用前后两者速度的平均值（见附录），这样一来会使verlet interation更加精确，当然计算量也会上升。这样的Verlet integration叫做速度V···I···

$$x_{n+1} = x_n + v_n\Delta t + \frac{a}{2}\Delta t^2 \tag{1.26}$$
$$\mapsto \tag{1.27}$$
$$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{a(t)}{2}\Delta t^2 \tag{1.28}$$

1.3 Taylor series

Dr Tom Crawford。该视频中，注重从Taylor系数入手，介绍其系数的由来。下面是f(x)在a处

展开后的表达式：

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

如果a=0，也就是在原点展开的话，Taylor Series也可以叫做Maclaurin Series，故：Maclaurin Series是Taylor Series的特殊形式：

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

为什么要有“展开点(expansion point)”？Taylor能用多项式的形式无限逼近真实的函数图像，但这种逼近是有代价的，这里的代价要从两方面理解：

- 计算量指数增长的代价：导数越高阶，幂级数越高，自然算量就需要越多；
- 逼近指数增长的代价：并非相同步长的多项式和就能“固定”增加一定程度的函数图像“样子”，而是越想相似就需要指数级增长的多项式项去拟合；

逼近图像是从两端延申的，展开点就是蔓延的中心，级数项越多圆圈的半径就越大。如果你就想知道在原点展开的泰勒级数在$f(6)$附近的样子，与其选择大量项去计算多形式和，还不如直接在6处展开。这里有个Desmos的示例：

1.4 综合除法

1.5 向量值函数（Vector-valued Function）

向量值函数（vector-valued function）是指输入一个或多个实数，输出一个向量的函数。简单来说，它是一个以标量变量t或多个变量u, v为自变量、以向量为值的函数。

1.6 Jacobian determinant

雅可比行列式用于空间变化后的系数计算，要理解其变换，我觉得一个重要的切入点是**偏导数**，当然这也是微积分的基础。回想一下一维函数的一阶导数，它表示因变量随着自变量的变化率，如果乘以dx，意味着：dx个单位的自变量引起多少个单位自变量的因变量变化，也就是$dy = f'dx$

如果在一维函数积分中，需要变换微分对象，那么就需要计算这样的变换系数，比如：

$$\int_0^2 (x)dx = \left.\frac{x^2}{2}\right|_0^2 = 2 \tag{1.29}$$

$$\text{if x=2t:} \tag{1.30}$$

$$dx = (2t)'dt = 2dt \tag{1.31}$$

$$\int_0^1 (2t)2dt = 2(t^2)\Big|_0^1 = 2 \tag{1.32}$$

那么在二维积分中该怎么算这个比例呢？这就用到了Jacobian Determinant. 回到前面说的，我们需要正确认知导数的含义，在二维积分中，自然就是偏导数了（partial derivative)。Let's first define what a partial derivative is: If a function is a multivariable function, we use the concept of partial differentiation (aka: partial derivative) to measure the effect of a change in one independent variable on the dependent variable, keeping the other independent variables constant. To apply the rules of calculus, at a time generally, we change only one independent variable and keep all other independent variables constant. In this way, we only look at the partial variation in the function instead of the total variation.

从微分推导的角度看待

$$z = f(x, y) \tag{1.33}$$

$$\rightarrow \tag{1.34}$$

$$dz = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \tag{1.35}$$

$$\text{def: x=g(u,v), y=h(u,v)} \tag{1.36}$$

$$\int dxdy \tag{1.37}$$

$$= \int \left(\frac{\partial g}{\partial u}du + \frac{\partial g}{\partial v}dv\right)\left(\frac{\partial h}{\partial u}du + \frac{\partial h}{\partial v}dv\right) \tag{1.38}$$

$$= \int \left(\frac{\partial g \partial h}{\partial u \partial u}du^2 + \frac{\partial g \partial h}{\partial u \partial v}dudv + \frac{\partial g \partial h}{\partial v \partial u}dvdu + \frac{\partial g \partial h}{\partial v \partial v}dv^2\right) \tag{1.39}$$

$$\approx \int \left(\frac{\partial g \partial h}{\partial u \partial v}dudv + \frac{\partial g \partial h}{\partial v \partial u}dvdu\right) \tag{1.40}$$

$$= \int \left(\frac{\partial g \partial h}{\partial u \partial v} - \frac{\partial g \partial h}{\partial v \partial u}\right)dudv \tag{1.41}$$

上面有2个重要的地方：

- 忽略了高阶无穷小du^2与dv^2项，因为它们更快趋向于0；
- 加号变减号，这个与微分几何学中叉乘的顺序有关，前者是$dudv$，后者是$dvdu$，变成一致的顺序后叉乘需要改变方向[1]。

从向量变化与叉积的角度看待

设想有这么一个积分运算及其参数的变换函数，如何求取$\mathbf{J}$?：

$$\text{def: } \begin{cases} x = g(u, v) = 2u + v \\ y = h(u, v) = u + 2v \end{cases} \tag{1.42}$$

$$\text{solve: } \iint dxdy = \iint \mathbf{J}dudv \tag{1.43}$$

$\frac{\partial x}{\partial u} = 2$意味着**u方向上增加du长度，x会在u方向上增加2du个长度**，同理v方向也是一样的，只不过增加相同的长度。我们取任意uv空间下的点$< a, b >$，向着u与v方向增加du与dv个极小的长度，那么x也会有一定的变化量，按照刚才提到的，x在u上的投影会增加$2du$个长度，而在v上的投影会增加$1dv$，同样我们也会知道y投影到UV空间的变化量$< du, 2dv >$，简单记作：

UV空间变化量	X投影到UV空间的变化量	Y投影到...
$< du, dv >$	$< 2du, dv >$	$< du, 2dv >$

[1] 具体细节需要进一步了解，查看微分几何学、函数行列式等内容。

从线性代数的角度看[2]，UV空间变化的微面元面积：

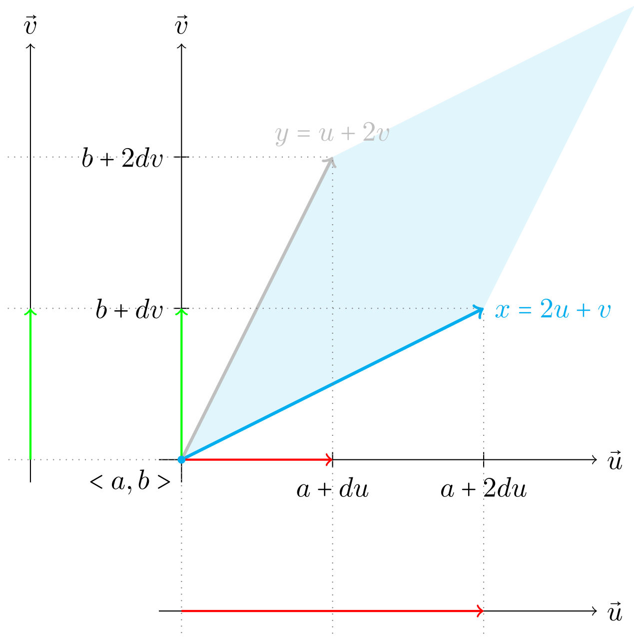
$$du \times dv = \begin{pmatrix}1\\0\end{pmatrix} \times \begin{pmatrix}0\\1\end{pmatrix} = 1$$

而XY空间它们各自的变化量投影到UV空间下的面积是：

$$dx \times dy = \begin{pmatrix}2\\1\end{pmatrix} \times \begin{pmatrix}1\\2\end{pmatrix} = 3$$

这里重点强调了**投影到**，这意味着上述计算的结果是在同一个空间，故此它们的数值具有可比性[3]。上述是在向量空间，在标量下的意义就是$dxdy = 3dudv$，因此计算的J应该是3，整个计算过程使用偏导数进行，组成的矩阵就是Jacobian matrix，其行列式就是Jacobian determinent.

$$\textbf{Jacobian matrix} = \frac{\partial x \partial y}{\partial u \partial v} - \frac{\partial y \partial x}{\partial u \partial v}$$



从图中我们知道新的空间$B ::= 0XY$，是定义在原来空间$A ::= 0UV$下的，因为只有存在了A，才能定义新的标架，以及它们的方向与大小。经过向量叉积我们知道x, y向量在$0UV$空间下的面积是：$\hat{x} \times \hat{y} = (2, 1) \times (1, 2) = 3$ 如果我们将视野搬到空间B下，面积会是多少呢？显然还是用叉积计算：向量$\vec{x}$在B空间下是其一个基向量，也就是$(1, 0)$，同理，$\vec{y}$会是$(0, 1)$，那么它们的面积是1，那么我们答案是$dxdy = 3dudv$，得出基本结论：

$$\textbf{if:} \begin{cases} x = g(u, v) \\ y = h(u, v) \end{cases} \tag{1.44}$$

$$\textbf{then:} \iint 1dxdy = \iint \begin{vmatrix} \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \\ \frac{\partial h}{\partial u} & \frac{\partial h}{\partial v} \end{vmatrix} dudv \tag{1.45}$$

[2]导数变化率在微观上具有线性近似性

[3]不同空间下的数值没有可比性，比如战国时期各个国家都有自己的度量衡，同一个物件赵国说3米，魏国说2米，虽然3大于2，但这种比较没有任何意义，必须是同一个度量衡，也就是这里说的投影到UV空间。

1.7 最小二乘法（least squares method）

最小二乘法，又称**最小平方法**，就是在多个样本中，找到一个表达式，使其能尽可能拟合所有给出的样本，在此之后，就可以用该表达式计算未知参数的结果。基本的思想就是利用偏导数等于0，计算参数的值。

$$(1,6)\Delta(2,5)\Delta(3,7)\Delta(4,10)$$

$$y = \beta_2 x + \beta_1$$

$$\beta_1 + 1\beta_2 = 6 \tag{1.46}$$

$$\beta_1 + 2\beta_2 = 5 \tag{1.47}$$

$$\beta_1 + 3\beta_2 = 7 \tag{1.48}$$

$$\beta_1 + 4\beta_2 = 10 \tag{1.49}$$

$$S(\beta_1, \beta_2) = [6 - (\beta_1 + 1\beta_2)]^2 + [5 - (\beta_1 + 2\beta_2)]^2 + [7 - (\beta_1 + 3\beta_2)]^2 + [10 - (\beta_1 + 4\beta_2)]^2 \tag{1.50}$$

最小值可以通过对$S(\beta_1, \beta_2)$分别求β_1和β_2的偏导数，然后使他们等于零得到。

$$\frac{\partial S}{\partial \beta_1} = 8\beta_1 + 20\beta_2 - 56 = 0 \tag{1.51}$$

$$\frac{\partial S}{\partial \beta_2} = 20\beta_2 + 60\beta_2 - 154 = 0 \tag{1.52}$$

$$\beta_1 = 3.5 \tag{1.53}$$

$$\beta_2 = 1.4 \tag{1.54}$$

$$y = 3.5 + 1.4x \tag{1.55}$$

1.8 emtpy

一个n维空间下的向量，怎么投影到另一个n维空间？投影之后各个分量是多少？

1.9 deep learning

1.9.1 损失函数 loss function

$$\ell^{(i)} = \frac{1}{2}[\hat{y}^{(i)} - y^{(i)}]^2$$

1.10 common equations

$$a^2 - b^2 = (a - b)(a + b) \tag{1.56}$$

$$a^3 + b^3 = (a + b)(a^2 - ab + b^2) \tag{1.57}$$

$$a^3 - b^3 = (a - b)(a^2 + ab + b^2) \tag{1.58}$$

二元一次方程的解：

$$ax^2 + bx + c = 0 \tag{1.59}$$

$$\Rightarrow \tag{1.60}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{1.61}$$

logarithm[4]:

$$\log_n(MN) = \log_n(M) + \log_n(N) \tag{1.62}$$

$$\log_n(M^p) = p\log_n(M) \tag{1.63}$$

$$\log_n(M) = \frac{\log_a(M)}{\log_a(n)} \tag{1.64}$$

[4] more equations to: https://en.wikipedia.org/wiki/Logarithm

Chapter 2

‡ chapter 2