偏导数与全微分

引言

在多元微积分里，**偏导数**和**全微分**是一对孪生概念：前者告诉你当**某一条坐标轴**上微移时函数的瞬时变化率，后者把所有坐标方向的信息**整合成一条"总账"**，给出最精确的一阶线性近似。为了让概念扎根，我们用 **Why → What → How → How Good** 的四步框架来"拆"它们。

---

1 Why —— 为什么需要它们?

典型情境	为什么少不了偏导数 / 全微分
工程调参：发动机效率 $E(p,T)$ 随压力 p、温度 T 变化	想知道"单独提高 1 K 温度能带来多少效率提升？"→ 用 $\partial E/\partial T$；想估计"两者各变一点"后的整体变化 → 用全微分 $\mathrm{d}E$。
数据敏感度：模型输出 $f(\mathbf{x})$ 对输入误差的放大	需要量化"哪一维更敏感"与"合成误差的首阶估计"。
最优化：在 $\mathbb{R}^n$ 中找极值	梯度 ∇f（偏导数打包）= "爬坡指南针"；全微分=切平面，用来验证可微性与泰勒展开。
坐标变换：极坐标、球坐标、机器人的关节坐标	链式法则写成 $\mathrm{d}f=\nabla f\cdot\mathrm{d}\mathbf{r}$ 才能快速"换账"。

---

2 What —— 概念到底是什么?

2.1 偏导数（Partial Derivative）

对二元函数举例：

$$\frac{\partial f}{\partial x}(x_0,y_0)=\lim_{h\to 0}\frac{f(x_0+h,y_0)-f(x_0,y_0)}{h}.$$

- **几何意涵**：沿 x 轴方向作水平切片，看到的是一条单变量曲线；偏导就是那条曲线在点 (x_0,y_0) 处的斜率。
- **向量视角**：把全部偏导排成列就得到了 **梯度**

$$\nabla f(x_0,y_0)=\Big(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\Big).$$

2.2 全微分（Total Differential）

若 f 在 $\mathbf{x}_0$ 处可微，则存在线性映射 $Df(\mathbf{x}_0):\mathbb{R}^n\to\mathbb{R}$ 使

$$f(\mathbf{x}_0+\mathbf{h})=f(\mathbf{x}_0)+Df(\mathbf{x}_0)[\mathbf{h}]+o(\|\mathbf{h}\|).$$

写成"坐标 + 形象"版本：

$$\boxed{\mathrm{d}f=\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\,\mathrm{d}x_i=\nabla f\cdot\mathrm{d}\mathbf{x}}$$

它就是 **最优一次线性近似** 的"公式化身"；$\mathrm{d}x_i$ 是"你愿意怎样小动"的自由度，全微分把每一维的影响加权并求和。

3 How —— 如何计算与运用?

3.1 一步步算

1. **先算所有偏导。**
2. **组装梯度** ∇f。
3. **替换** $\mathrm{d}x_i$ 得到全微分。

> **示例** $f(x,y,z)=x^2yz+\sin(xz)$

$$> \begin{array}{c} > f_x = 2xyz + z\cos(xz), \quad > f_y = x^2 z, \quad > f_z = x^2 y + x\cos(xz) \\ > \Rightarrow \ \mathrm{d}f = (2xyz + z\cos xz)\,\mathrm{d}x + x^2 z\,\mathrm{d}y + (x^2 y + x\cos xz)\,\mathrm{d}z. > \end{array} >$$

3.2 常用操作

任务	做法	说明
误差传播	$\sigma_f^2 \approx \sum_i \left(\frac{\partial f}{\partial x_i}\sigma_{x_i}\right)^2$	把全微分看作随机增量的线性组合；前提是误差小且独立。
链式法则（复合函数）	若 $u=g(\mathbf{x})$，$y=f(u)$	先算 $\mathrm{d}u$，再 $\mathrm{d}y=f'(u)\,\mathrm{d}u$。
隐函数微分	方程 $F(x,y)=0$	取全微分 $F_x\,\mathrm{d}x+F_y\,\mathrm{d}y=0 \to \dfrac{\mathrm{d}y}{\mathrm{d}x}=-\dfrac{F_x}{F_y}$。
切平面/线性化	$L(\mathbf{x})=f(\mathbf{x}_0)+\nabla f(\mathbf{x}_0)\cdot(\mathbf{x}-\mathbf{x}_0)$	把全微分写在 $\mathbf{h}=\mathbf{x}-\mathbf{x}_0$ 上即可。

---

4 How Good —— 线性近似到底有多好?

1. **误差阶数**

$$f(\mathbf{x}_0+\mathbf{h}) = f(\mathbf{x}_0) + \nabla f\cdot\mathbf{h} + \underbrace{\tfrac{1}{2}\mathbf{h}^\top H_f(\mathbf{x}_0)\mathbf{h}}_{\text{二阶项}} + o(\|\mathbf{h}\|^2),$$

其中 H_f 是 Hessian。

- **结论**：只要 $\|\mathbf{h}\|$ 足够小，遗漏的部分是 **二阶无穷小。**

2. **误差上界**（均匀 Lipschitz）
   若梯度连续且有界 $\|\nabla f\|\le M$，则

$$|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - \nabla f\cdot\mathbf{h}| \le \frac{M}{2}\|\mathbf{h}\|^2.$$

3. **偏导数存在≠可微**
   - 经典反例：$f(x,y)=\frac{|xy|}{x^2+y^2}$ 在 $(0,0)$ 处虽有全部偏导，却不可微。
   - 检验标准：偏导连续 ⇒ 一定可微；只给存在不够。

---

5 联系 vs. 区别（一句话版）

> **偏导数** = 按坐标轴“局部单项考试”；
> **全微分** = 把所有考试成绩按权重汇总，变成一份“总评估 + 线性预测”。

---

6 常见误区

1. **把 $\mathrm{d}f$ 当作“具体数值”**
   $\mathrm{d}f$ 是线性算子，只有给定一组 $(\mathrm{d}x_1, \ldots, \mathrm{d}x_n)$ 后才会产出数字。
2. **忽视“高阶小量”**
   $\Delta f \approx \mathrm{d}f$ 不是等号；差别正是一阶近似的误差。
3. **梯度方向**
   ∇f 指向最速**增**方向，不是“切线方向”。切向量在梯度的正交子空间里。

---

7 总结·回味

- **Why**：在多变量世界里量化“单向敏感度”与“整体线性变动”，离不开这两个概念。
- **What**：偏导数 = 单轴极限斜率；全微分 = 梯度·微位移 = 最佳一次线性映射。
- **How**：先求偏导→组梯度→套公式；以误差传播、链式法则、隐函数、优化为典型舞台。
- **How Good**：误差为 $O(\|\mathbf{h}\|^2)$，梯度连续即可确保“好”；若仅有偏导则仍需警惕不可微陷阱。

掌握了“单项考试 + 总评估”的思路，你就能从任何角度“拨动”变量，立即评估函数的首阶响应，并确信自己的估算**有多靠谱**。