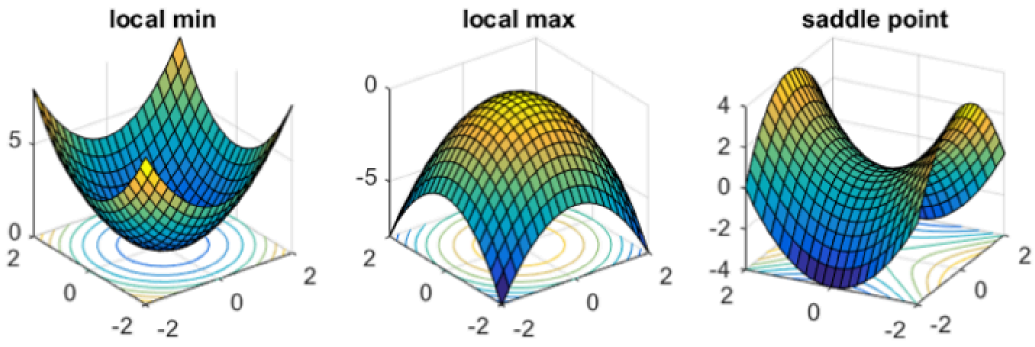
Local Minimum and Saddle Point

1. Critical Point（梯度为零的点）



在训练中，loss 不再下降，常见原因是：

- 梯度变得非常小（≈ 0）
- 参数更新停住（gradient descent 无法继续走）

此时我们可能来到所谓 **critical point**：

$$\nabla_\theta L(\theta) = 0$$

critical point 包含：

- **local minimum**
- **local maximum**
- **saddle point（鞍点）**

2. Taylor Series Approximation（泰勒展开）

loss 在某点 θ' 附近可近似为：

$$L(\theta) \approx L(\theta') + (\theta - \theta')^T g + \frac{1}{2}(\theta - \theta')^T H(\theta - \theta')$$

其中：

- 梯度（gradient）

$$g = \nabla_\theta L(\theta')$$

- Hessian（海森矩阵）

$$H_{ij} = \frac{\partial^2 L(\theta')}{\partial\theta_i\,\partial\theta_j}$$

在 critical point（$g = 0$）时：

$$L(\theta) \approx L(\theta') + \frac{1}{2}(\theta - \theta')^T H(\theta - \theta')$$

因此 **Hessian 决定点的性质**。

3. 用 Hessian 判断点的类型

设 $v \neq 0$，看二次型：

$$v^T H v$$

3.1 Local Minimum 的条件

$$v^T H v > 0, \quad \forall v$$

等价于：

$$H \text{ positive definite} \iff \text{所有 eigenvalues} > 0$$

3.2 Local Maximum 的条件

$$v^T H v < 0, \quad \forall v$$

等价于：

$$H \text{ negative definite} \iff \text{所有 eigenvalues} < 0$$

3.3 Saddle Point 的条件

$$v^T H v \text{ 有时 >0、有时 <0}$$

等价于：

$$H \text{ indefinite} \iff \text{eigenvalues 有正有负}$$

深度学习中最常出现的情况：saddle point，而不是 local minima。

4. 如何从 Saddle Point 逃离？

设 u 是 Hessian 的 eigenvector，对应 eigenvalue 为 λ。

$$Hu = \lambda u$$

代入二次项：

$$u^T H u = \lambda \|u\|^2$$

若 $\lambda < 0$，则：

$$L(\theta' + u) < L(\theta')$$

表示沿着 eigenvector 的方向 u 移动，可以降低 loss。

更新方式概念上为：

$$\theta \leftarrow \theta' + \alpha u$$

（实际中不会这么做，因为计算 Hessian 代价极大）

4. Deep Learning: 为什么 Saddle Point 比 Local Min 更常见?

- 神经网络参数通常 **百万到千万维**
- Error surface 是高维空间形状
- 高维中"看似无路"的低维凹点往往只是 saddle，而不是 local minima
- 在高维中，存在大量下降方向 → local minima 很少

实验亦支持：

- 训练过程中卡住的点几乎都有 **正负 eigenvalues 混合**
- 真正所有 eigenvalues > 0 的情形很罕见

因此：

> **训练停住的原因几乎总是 saddle point，而不是 local minimum。**