

病态问题一般被认为存在于神经网络训练过程中。病态体现在随机梯度下降会“卡”在某些情况，此时即使很小的更新步长也会增加代价函数。

回顾方程式(4.9)，代价函数的二阶泰勒级数展开预测梯度下降中的 $-\epsilon g$ 会增加

$$\frac{1}{2}\epsilon^2 g^\top H g - \epsilon g^\top g \quad (8.10)$$

到代价中。当 $\frac{1}{2}\epsilon^2 g^\top H g$ 超过 $\epsilon g^\top g$ 时，梯度的病态会成为问题。判断病态是否不利于神经网络训练任务，我们可以监测平方梯度范数 $g^\top g$ 和 $g^\top H g$ 。在很多情况中，梯度范数不会在训练过程中显著缩小，但是 $g^\top H g$ 的增长会超过一个数量级。其结果是尽管梯度很强，学习会变得非常缓慢，因为学习速率必须收缩以弥补更强的曲率。如图8.1所示，成功训练的神经网络中，梯度显著增加。

尽管病态存在于包括神经网络训练的其他情况中，有些适用于其他情况的解决病态的技术并不适用于神经网络。例如，牛顿法在解决带有病态条件数的Hessian矩阵的凸优化问题时，是一个非常优秀的工具，但是我们将会在下节中说明牛顿法运用在神经网络前需要很大的改动。



