# 耳蜗：神经网络训练缓慢的根本原因

自从几年前在斯坦福上Stephen Boyd的“凸优化”课以来，我一直对数值优化的基本原理感兴趣。在这个领域发展起来的理论可以用来回答当今实践者面临的问题，特别是 - “为什么训练神经网络要花这么长时间？”？”

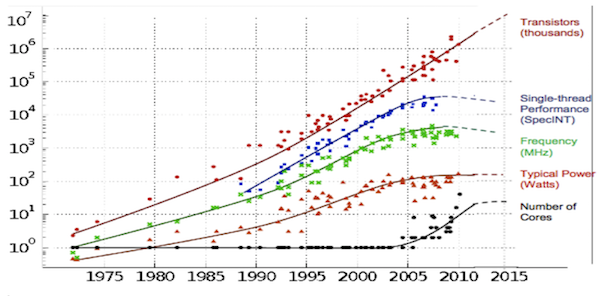
在80年代训练手写数字的Conv nets花了大约1周的时间来训练当时的硬件。今天的研究会议希望在ImageNet大小的数据集上演示您的方法，这也需要一周的硬件时间（一个高端GPU）。

拖延挂钟时间的一个因素是心理因素。对知识的渴求是不可抑制的，因此研究人员会不断增加计算量，直到它花费太长时间。谷歌的一些工程师使用10个PFlop TPU播客在MNIST上训练异国情调的模型。很明显，一个有创造力的人会找到一种方法来耗尽他们可用的任何计算量。

然而，也有一些基本因素使训练时间保持在高位。考虑以下代码来训练神经网络：

即使可用晶体管的数量一直呈指数级增长，但如果您必须按顺序执行300万次迭代，在某些情况下，额外的晶体管将不再起作用，您将受到单线程性能的限制。

单线程性能一直不太好。看看这个图表：



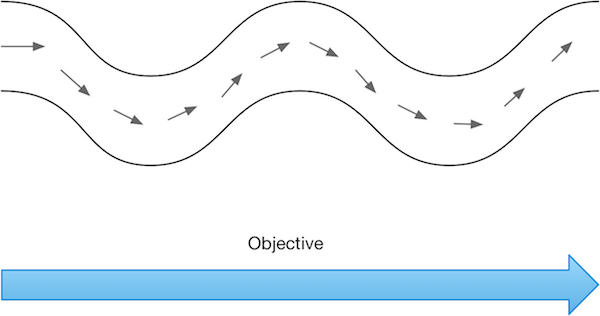
2007年前后，连续剧的表现达到顶峰，此后一直低迷。实际上，我发现在我3岁的笔记本电脑上训练一些Atari RL机型比在最新的Intel Xeon芯片上更快。

为了解决单线程性能差的问题，我们需要找出减少训练循环中迭代次数的方法。

由于以下3个OtodScIOS（减少非线性优化中的串行复杂度的障碍），迭代次数一直保持较高。

### OTODSCINO 1：非线性量

考虑非线性优化问题的以下情况：



因为梯度下降对局部信息起作用，它需要采取一定数量的步骤，直到它能“看到”最小值为止。

非线性可以通过目标函数组件之间的交互来描述，并且当您向网络添加更多层时，它增加了交互的可能性。

以一个玩具为例，将几个随机矩阵相乘，并通过调整任意条目a、b来最小化结果的范数：

即使对于这种纯线性的神经网络，优化问题也会随着层数的增加而变得非线性。

总结一下：

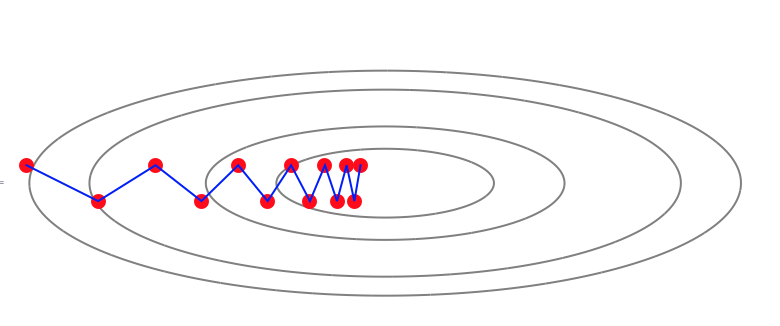


### OTODSCINO 2：本地条件号

考虑以下最小化问题：

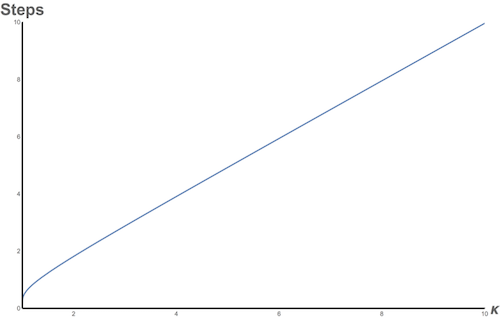


这是一个线性估计问题，所以OTODSCINO 1不适用。然而，梯度下降仍然很困难，因为梯度不指向最小值。这个问题的梯度下降将遵循一个特征的曲折路径：

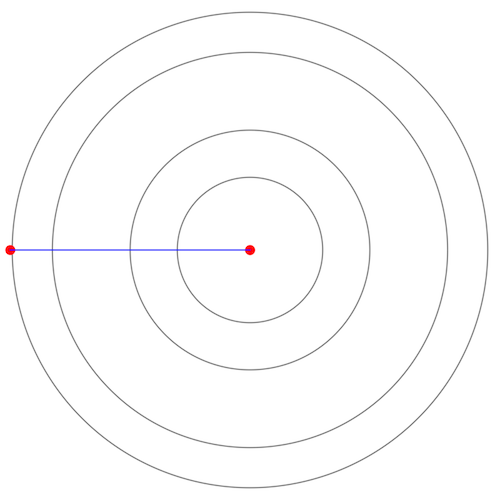


有一个量度梯度下降的条件数，它是最大直径和最小直径的比值，上面的问题是10。

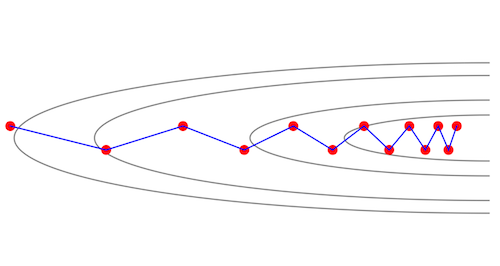
接近最小值所需的步骤数按O（条件数）增长：



步数随着O（k）的增加而增加。



条件1：最小化需要一个步骤。



条件10：最小化需要10个步骤。

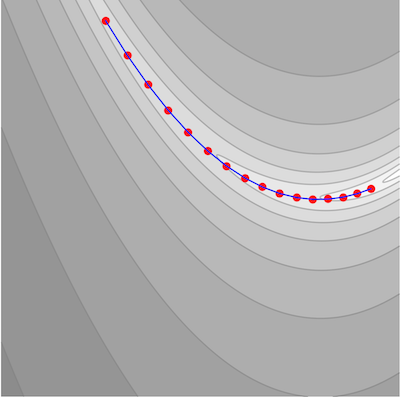
神经网络的优化曲面条件很差（中的经验评价），加上参数使条件更差。

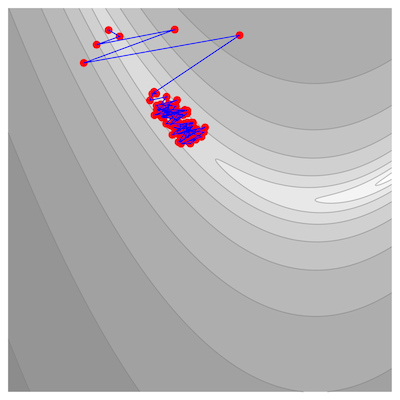
总结一下：



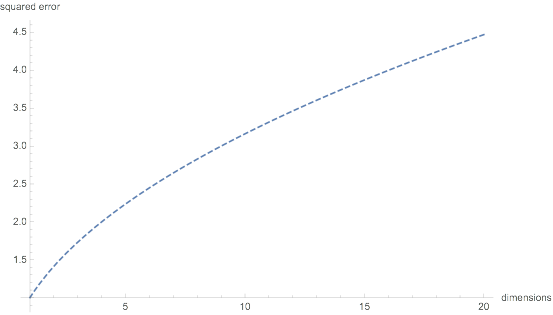
### OTODSCINO 3：梯度噪声量

神经网络优化使用随机梯度下降而不是梯度下降。我们对方向的错误估计被噪音进一步冲淡了。下面是一个在有噪声和无噪声的情况下最小化Rosenbrock函数的例子：





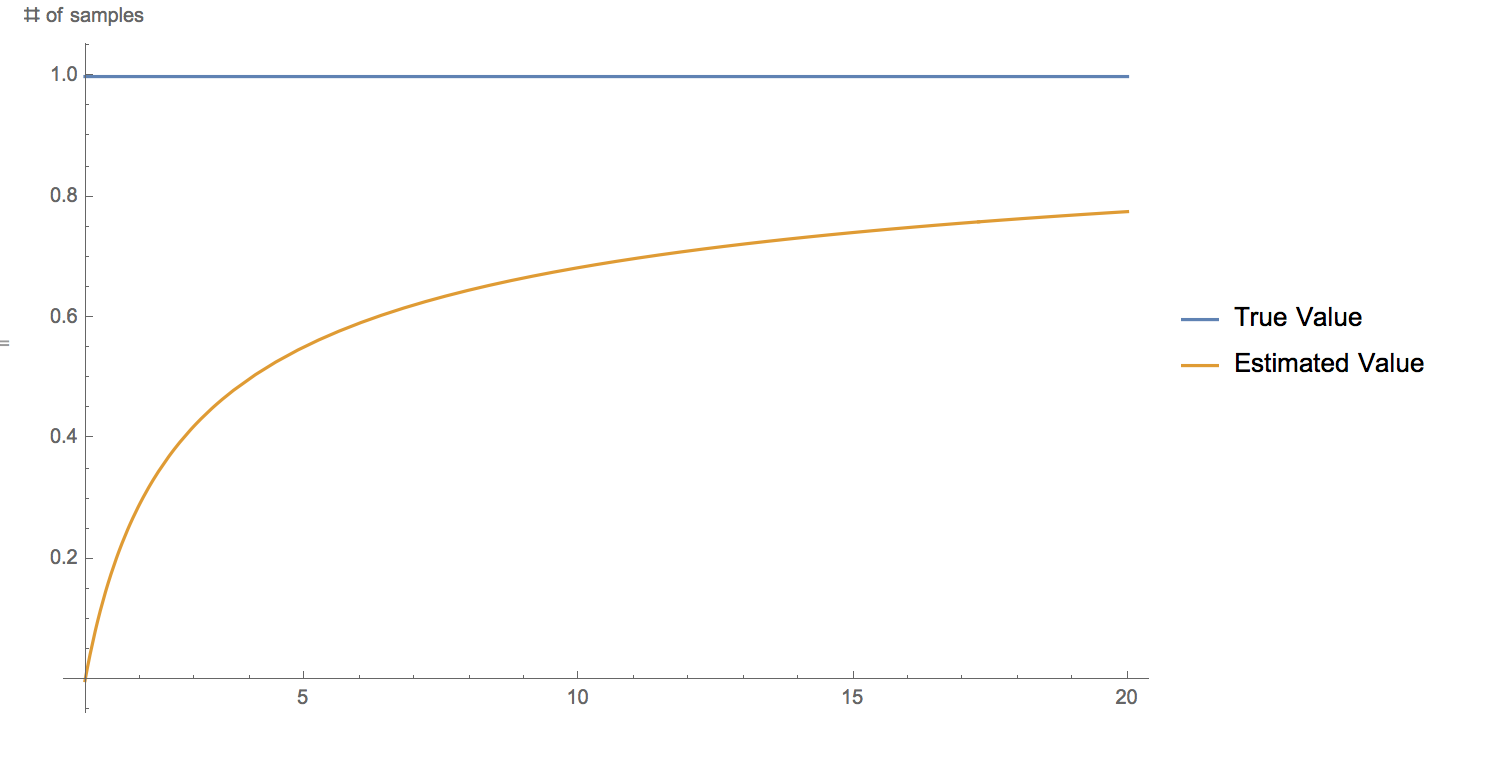
向参数空间添加尺寸时，噪波会增加。高斯正态噪声最容易看到。d维噪声引入的均方根误差近似增长为平方（d）



有关推导，请参见。

这最后一个OTODSCINO有一些希望，即使噪声随着附加维度的增加而增长，我们可以使用指数增长的晶体管池并行计算估计值并将它们平均起来。

从弱大数定律可知，误差收缩为sqrt（n），其中n是样本数。



在噪声和平均公式中都有平方根，这提供了一个容易记住的经验法则 - 为了避免额外的噪声使事情变得更糟，可以以与参数数量相同的速率增加批大小。

总结一下：



但是使用额外的并行计算，您可以否定这一点：



为了减少墙上的时钟时间，我们需要在三个耳蜗上芯片。第一个障碍可以通过使用“较少非线性”的神经网络参数化来解决。Resnet和ReLU激活是该领域进展的例子。第二个障碍可以通过采用从线性估计到神经网络的先进方法来缓解，比如KFAC。第三个障碍需要软件工程来简化多个设备的集成。最后，这三个障碍都可以通过下一代硬件来缓解，这将更有效地使用晶体管进行深度学习。最近在这个地方发现的一项调查。

To follow along with more [South Park Commons](https://www.southparkcommons.com/) members’ research and projects, sign up for the [SPC email newsletter](https://mailchi.mp/116e4aebefbc/southparkcommons).