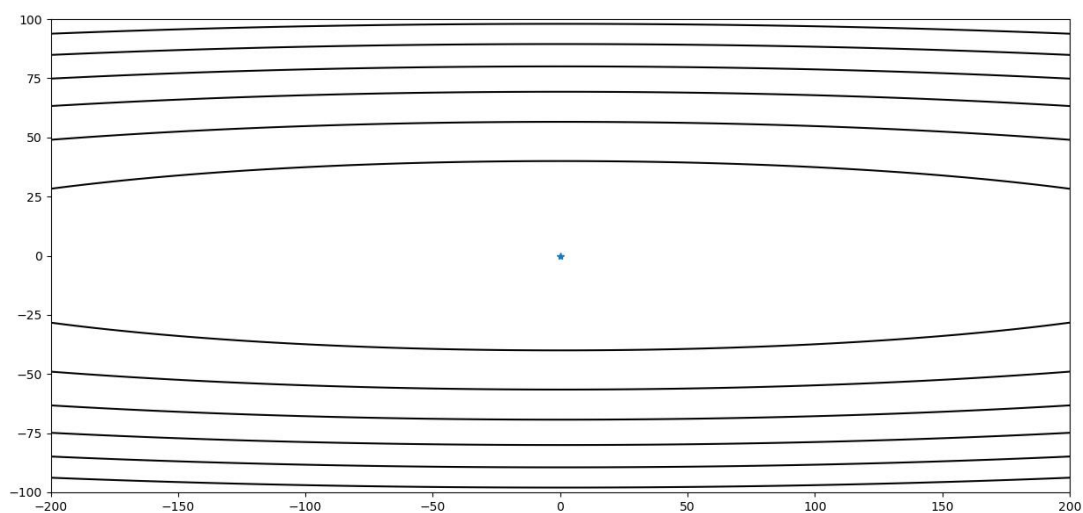


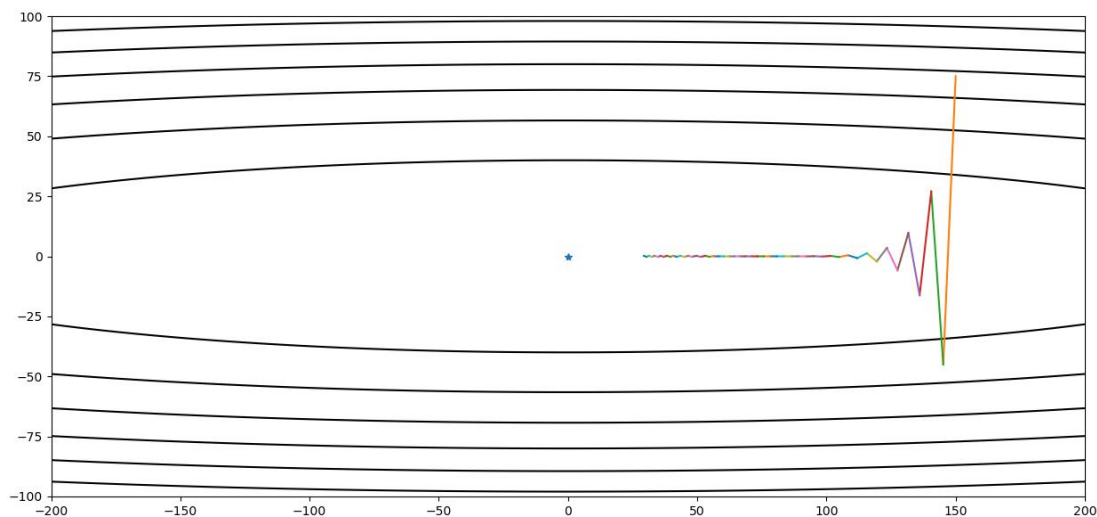
第八章 深度模型中的优化部分实验

一、验证 Hessian 矩阵存在病态时，sgd，momentum，nesterov 各自的表现

1、假设 loss function 为: $f(x,y) = x^2 + 50y^2$ ，则可以画出 loss 的等高线（蓝色星点为最优点），如下：

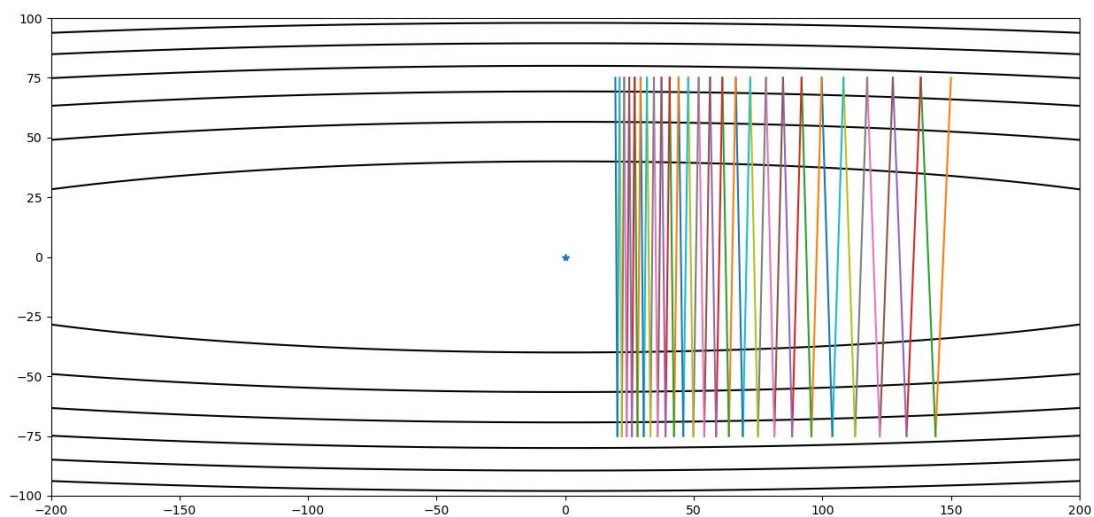


2、当使用较小学习率的 SGD 算法求解时：



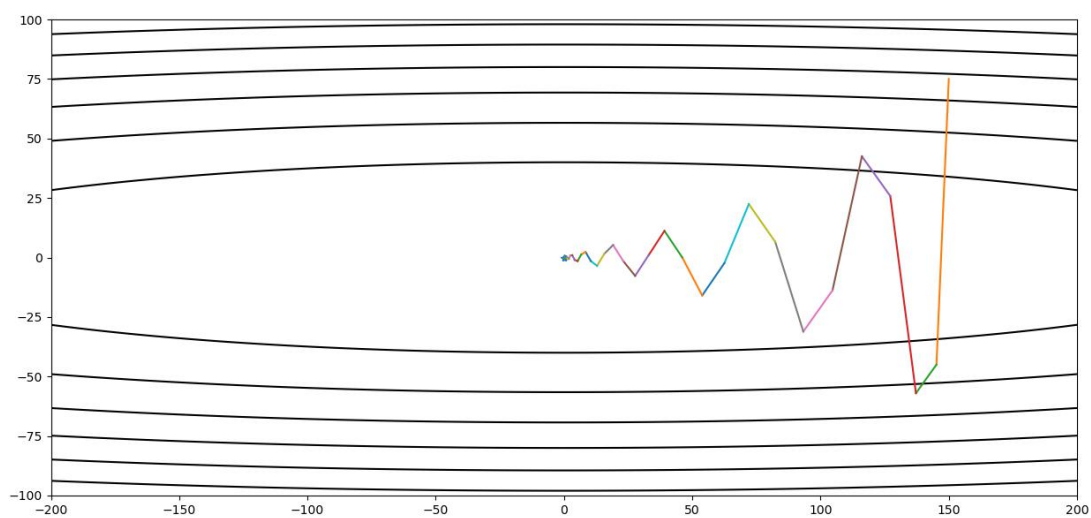
可以发现：因为 y 方向上的梯度是大于 x 方向上的梯度的，所以在 y 方向上会存在一些震荡，但是以为此时学习率较小，在 y 方向上还是能收敛的，只是在 x 方向上收敛很慢。

3、为了使得加速 x 方向上的收敛速度，我们加大学习率：



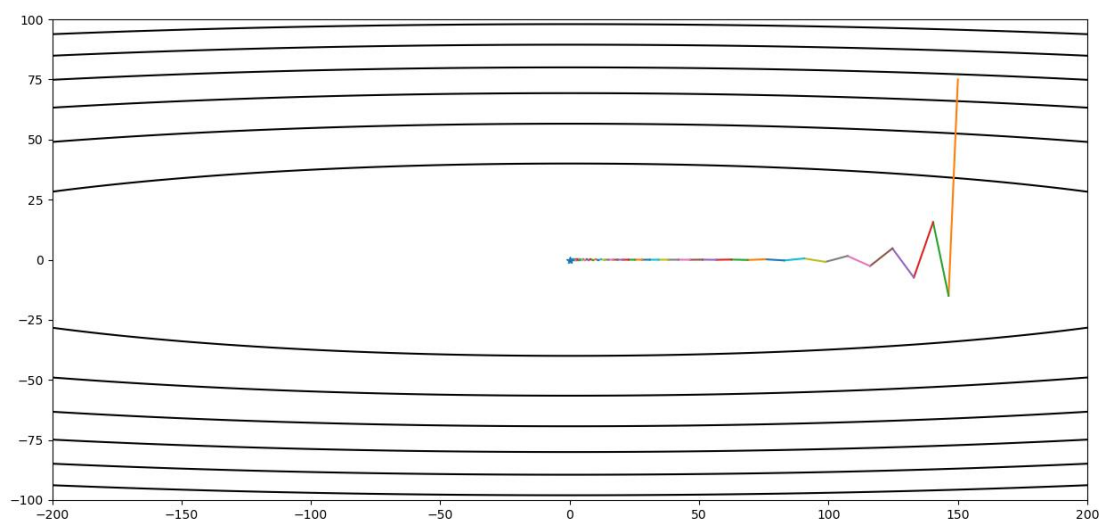
可以发现：当增大学习率之后由于 y 方向上梯度比较大的原因，会在 y 方向上不断震荡，没有收敛的趋势。

4、添加 momentum 后：



可以发现：添加 momentum 后，不仅能抵消一部分 y 方向上的震荡，还能加速 x 方向上的收敛速度。

5、使用 nesterov



比 momentum 有更小的震荡

二、验证权重初始化对每层输出的影响

网络设置为：9 个全连接层，每个全连接层均添加 ReLU 激活函数

1、权重随机初始化为均值为 0，标准差为 0.01

```
input mean -0.00055 and std 0.99942
layer 1 mean 0.11283 and std 0.16510
layer 2 mean 0.02177 and std 0.03197
layer 3 mean 0.00419 and std 0.00612
layer 4 mean 0.00073 and std 0.00106
layer 5 mean 0.00014 and std 0.00020
layer 6 mean 0.00002 and std 0.00003
layer 7 mean 0.00000 and std 0.00001
layer 8 mean 0.00000 and std 0.00000
layer 9 mean 0.00000 and std 0.00000
```

可以发现，随着网络的加深，输出分布的标准差趋近于 0

2、权重随机初始化为均值为 0，标准差为 1

```
layer 1 mean 11.29963 and std 16.55008
layer 2 mean 212.76242 and std 317.18530
layer 3 mean 3721.89941 and std 5660.79102
layer 4 mean 64864.29688 and std 96380.51562
layer 5 mean 1127515.00000 and std 1681990.62500
layer 6 mean 19708794.00000 and std 28712468.00000
layer 7 mean 307348512.00000 and std 446440992.00000
layer 8 mean 4226711808.00000 and std 6166444544.00000
layer 9 mean 57510981632.00000 and std 88393940992.00000
```

可以发现，随着网络的加深，输出分布的标准差变得很大

3、xavier 初始化

```
input mean -0.00061 and std 0.99968
layer 1 mean 0.39712 and std 0.58205
layer 2 mean 0.28575 and std 0.41919
layer 3 mean 0.19504 and std 0.29021
layer 4 mean 0.13776 and std 0.20061
layer 5 mean 0.08757 and std 0.13227
layer 6 mean 0.06961 and std 0.09486
layer 7 mean 0.05077 and std 0.07250
layer 8 mean 0.03598 and std 0.05434
layer 9 mean 0.02754 and std 0.03867
```

4、msra 初始化

```
input mean 0.00021 and std 0.99992
layer 1 mean 0.56466 and std 0.82630
layer 2 mean 0.53153 and std 0.80305
layer 3 mean 0.54207 and std 0.80047
layer 4 mean 0.57075 and std 0.82008
layer 5 mean 0.57368 and std 0.83010
layer 6 mean 0.56039 and std 0.81840
layer 7 mean 0.56670 and std 0.81028
layer 8 mean 0.58376 and std 0.81410
layer 9 mean 0.55047 and std 0.84539
```

5、随机初始化+bn :

```
input mean -0.00176 and std 0.99949
layer 1 mean 0.39657 and std 0.57974
layer 2 mean 0.39102 and std 0.57209
layer 3 mean 0.39029 and std 0.57162
layer 4 mean 0.38919 and std 0.57102
layer 5 mean 0.38820 and std 0.57024
layer 6 mean 0.38689 and std 0.56889
layer 7 mean 0.38538 and std 0.56731
layer 8 mean 0.38369 and std 0.56535
layer 9 mean 0.38161 and std 0.56324
```