## **Gradient Descent with Momentum:**

图示

中度可信度描述已自动生成

这里我们会试图去优化上图所示的损失函数，红点代表着全局最小值，

卡通人物

中度可信度描述已自动生成

它具有两个维度，假设为和，显然，我们可以看到由于函数沿方向的gradient更大 这导致在梯度下降的过程中会出现上图蓝线样式的波动，使得我们需要更多的时间来达到全局最小值。我们可以通过**Gradient Descent with Momentum**来解决这个问题，算法细节如下：

On interaction , compute on current mini batch, the learning rate is :

Initial with

Herer means the exponentially weighted average (please refer to the [Exponentially Weighted Averages.pdf](https://github.com/GuoJiaqi-1020/Jacky-s-ML-notebook/blob/main/pdf%20note/Exponentially%20Weighted%20Averages.pdf)), and the position will be updated using the above two equations:

Usually, , which is like averaging the gradients of the last 10 iterations. Therefore, it smooths out the steps of gradient descent

这样一来，如下图所示，在b方向的gradient 会相互抵消，而在 方向的gradient 则不会出现这种情况，从而减小了波动，使得模型更快收敛

图示

低可信度描述已自动生成

**Note:** 这个算法收到了基础物理中的启发，我们可以把算法中的, 比作加速度(当前状态)，把, 看作速度