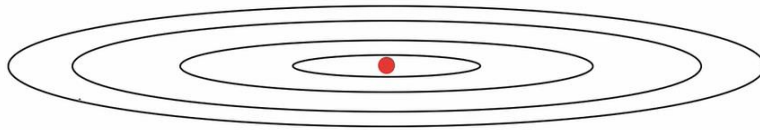
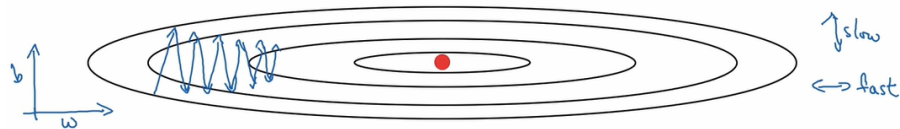


**RMSprop:** ( [https://www.youtube.com/watch?v=e-LFe\\_igno](https://www.youtube.com/watch?v=e-LFe_igno) )



假设这是我们需要优化的损失函数



它具有两个维度，假设为  $b$  和  $\omega$ ，显然，我们可以看到由于函数沿  $b$  方向的 gradient 更大 (larger  $db$ ) 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。为此，我们可以通过 RMSprop 来进行修正，其算法详情如下：

On interaction  $t$ , compute  $d\omega, db$  on current mini batch, the learning rate is  $\alpha$ :

$$S_{d\omega} = \beta S_{d\omega} + (1 - \beta)(d\omega)^2$$
$$S_{db} = \beta S_{db} + (1 - \beta)(db)^2$$

Here  $S$  means the exponentially weighted average (please refer to the [Exponentially Weighted Averages.pdf](#)), and the position will be updated using the above two equations:

$$\omega := \omega - \alpha \frac{d\omega}{\sqrt{S_{d\omega} + \epsilon}} \quad b := b - \alpha \frac{db}{\sqrt{S_{db} + \epsilon}}$$

$\epsilon$  is just to make sure that the denominator is not 0, so it's usually a very small number (default:  $10^{-8}$ )

简单而言，由于  $db > d\omega$  所以  $S_{d\omega} < S_{db}$ ，显然，直接作用到参数  $\omega$  和  $b$  上的结果就是，方向  $b$  上面的变化  $\frac{db}{\sqrt{S_{db} + \epsilon}}$  会被减弱，而方向  $\omega$  方向上的变化则会增强。最终结果如下图绿线所示。

