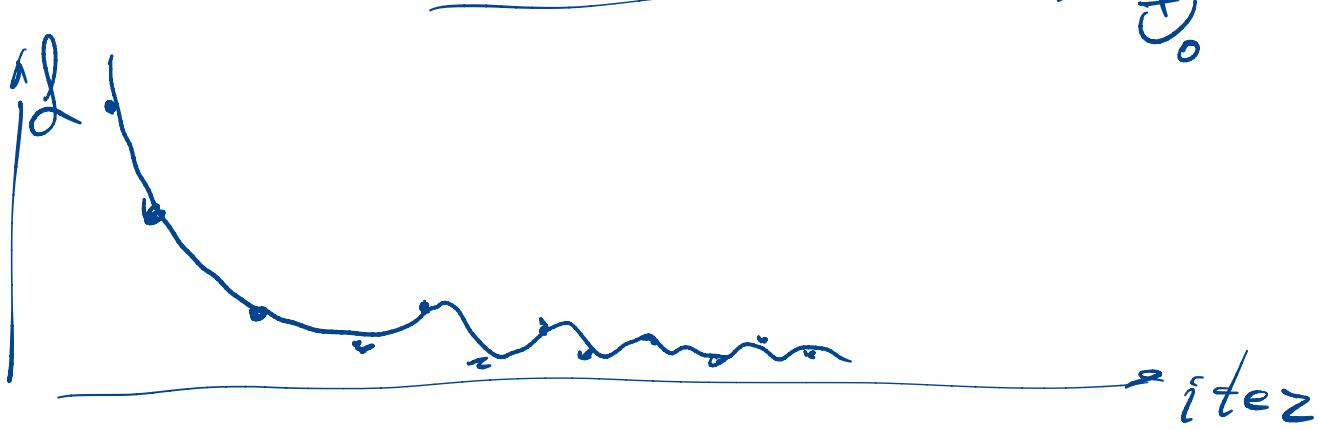
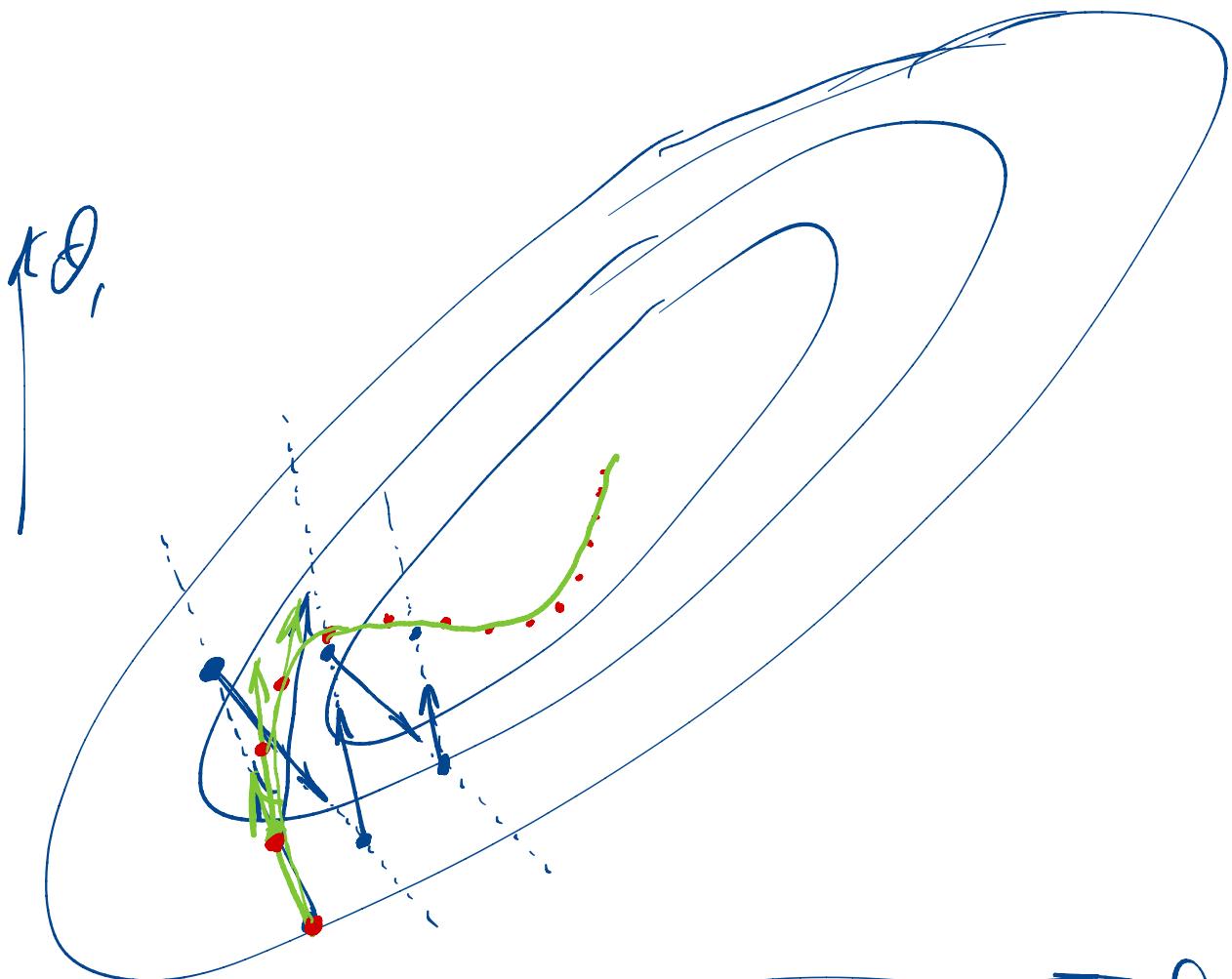


SGD

$$\underline{\theta}^{(t+1)} = \underline{\theta}^{(t)} - \eta \nabla_{\underline{\theta}} L(\underline{J}, \underline{\theta}^{(t)}) \\ = \vec{g}$$



# Итерационный метод (Momentum)

$$\vec{g}^{(t)} = \nabla_{\theta} L(\beta, \theta^{(t)})$$

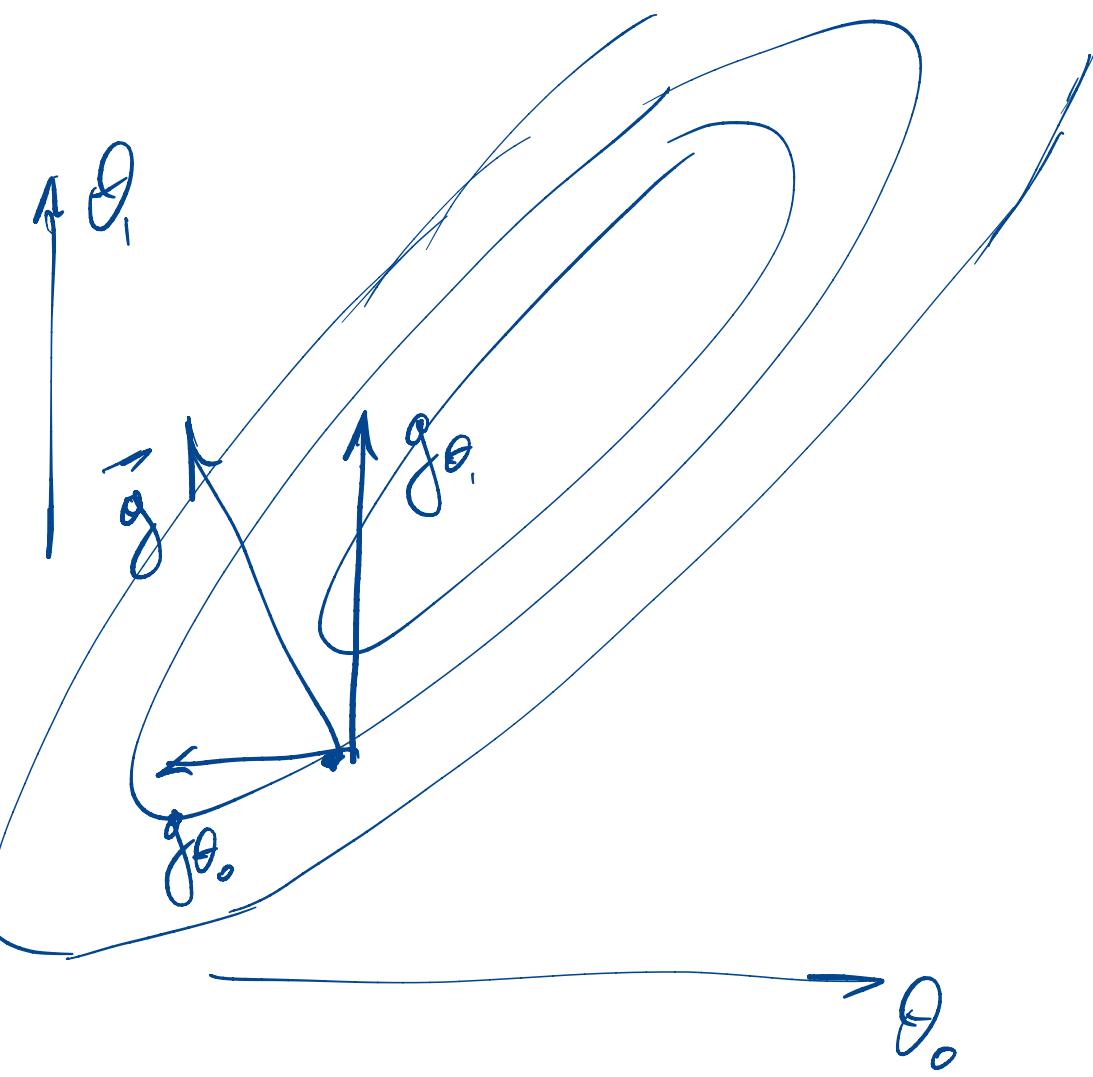
$$\vec{m}^{(t+1)} = \beta \vec{m}^{(t)} + (1 - \beta) \vec{g}^{(t)}$$

$$\beta \sim 0,9 - 0,99$$

$$\theta^{(t+1)} = \theta^{(t)} - \gamma \vec{m}^{(t)}$$

Фактор:

- ① Сложенный оптимизанд
- ② Погенулятивно более быстрый оптимизанд
- ③ Погенулятивно изобличенное обобщенное неподходящее.



Може наблюдава съвпадащо  
направление  $\hat{g}$  по направление  $\theta_i$ .

$$g = \nabla_{\theta} L \quad g = \frac{\hat{g}}{\|\hat{g}\|}$$

$$S^{(t+1)} = \beta_2 S^{(t)} + (1 - \beta_2) g^2$$

$$\beta_2 \sim 0.999$$

$$g' = \frac{g}{\sqrt{S + \epsilon}}$$

$$\epsilon \sim 10^{-8}$$

AdaGrad

# SGD, AdaGrad & Momentum

Adam

$$g^{(t)} = \nabla_{\theta} \mathcal{L}(B, \theta^{(t)})$$

$$m^{(t)} = m^{(t-1)} \beta_1 + (1 - \beta_1) g^{(t)}$$

$$s^{(t)} = s^{(t-1)} \beta_2 + (1 - \beta_2) (g^{(t)})^2$$

$$\theta^{(t+1)} = \theta^{(t)} - \gamma \frac{m^{(t)}}{\sqrt{s^{(t)}} + \epsilon}$$