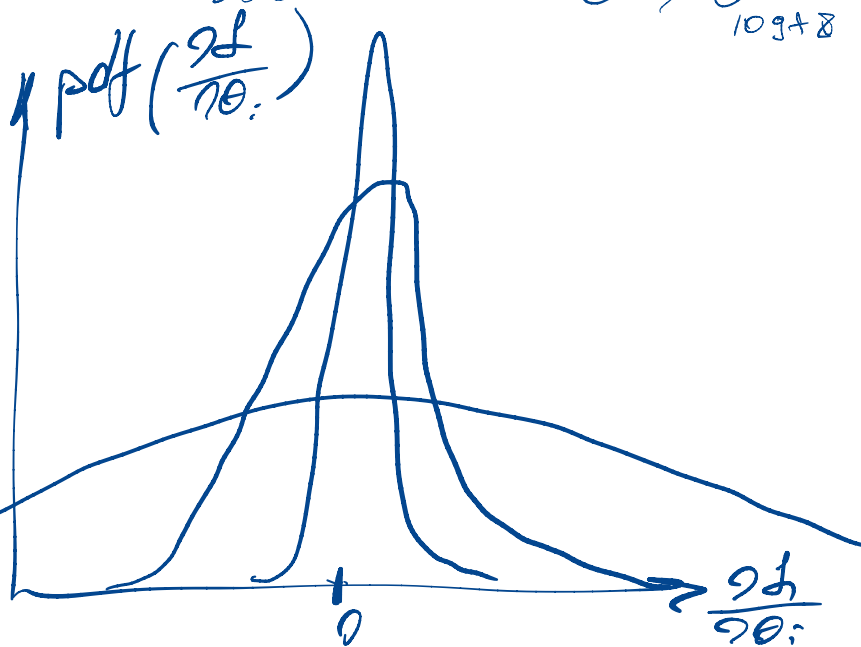


SGD:

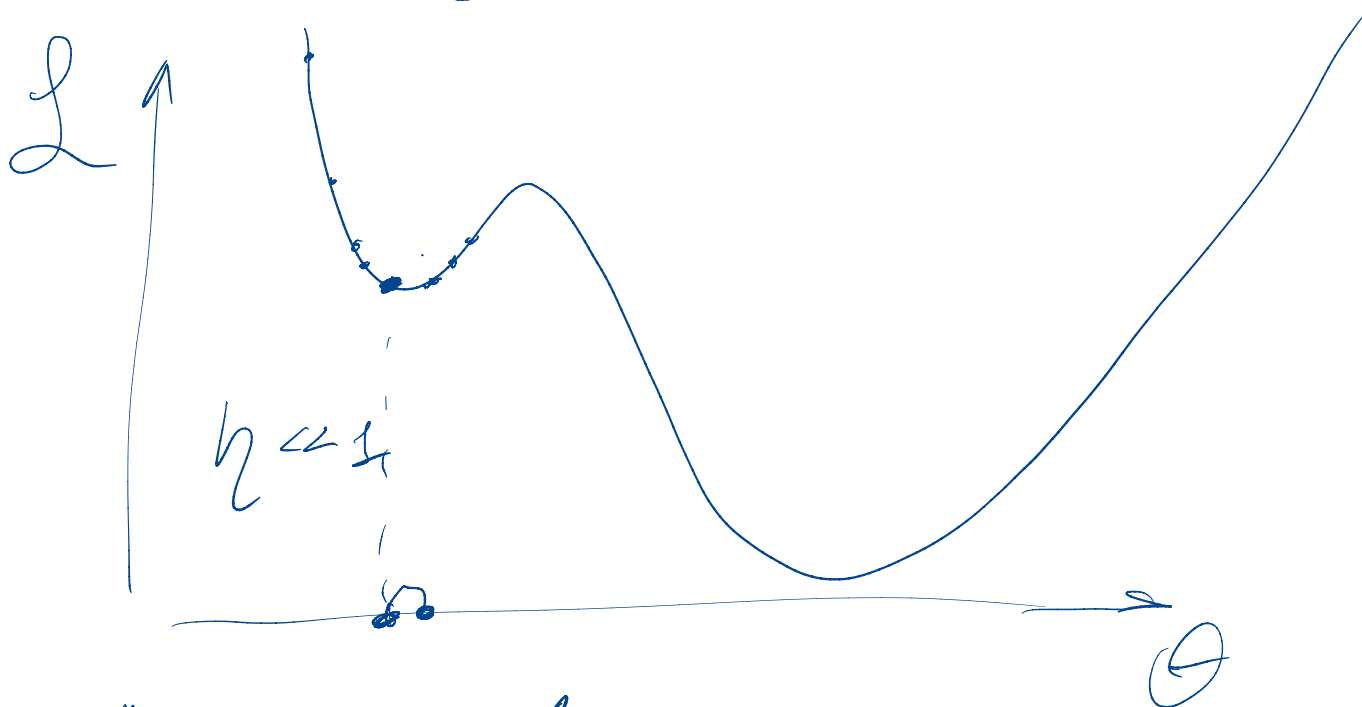
$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_{\theta} L$$

$$\theta_{10978}^{(t+1)} = \theta_{10978}^{(t)} - \eta \frac{\partial L}{\partial \theta_{10978}}$$



Приёмы в градиентной оптимизации:

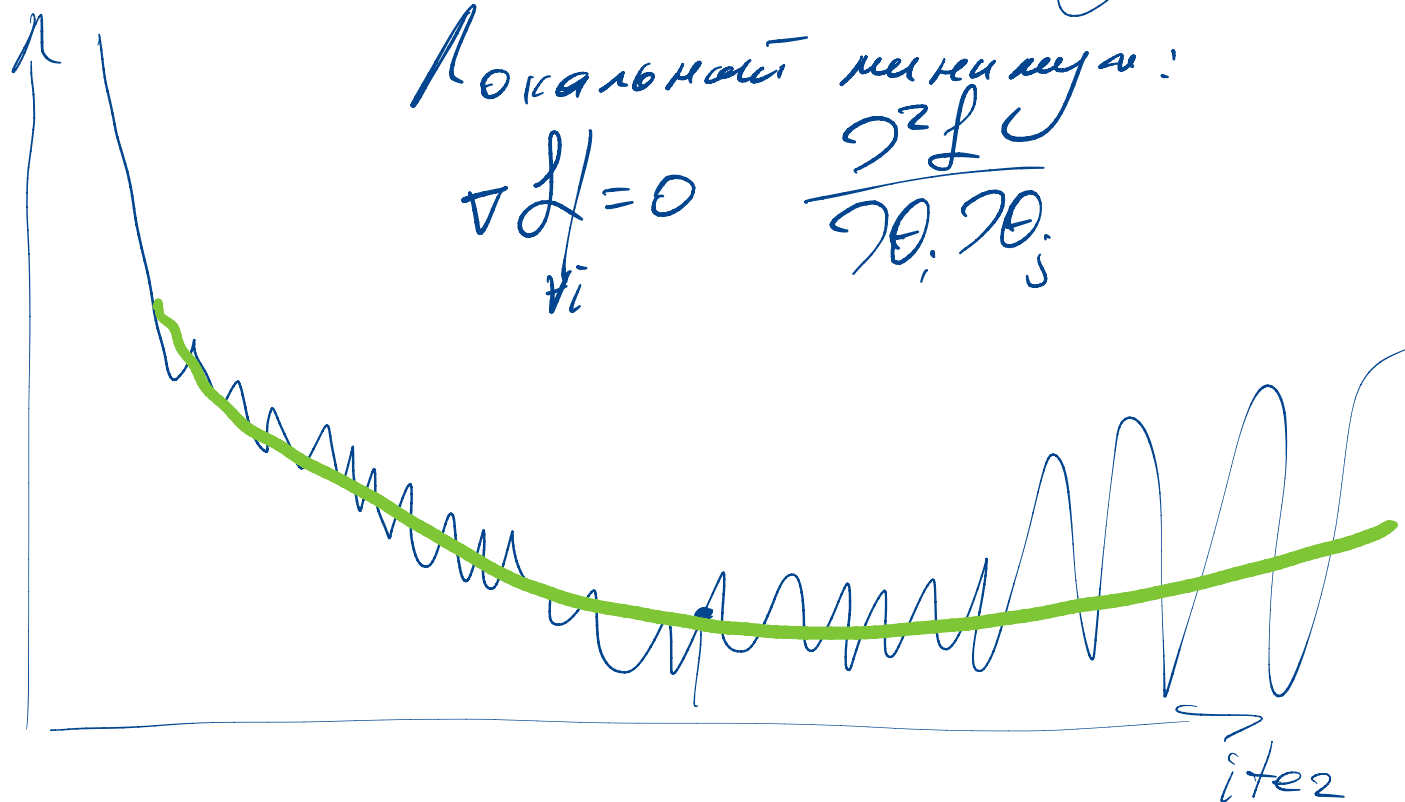
① Имитация обжига

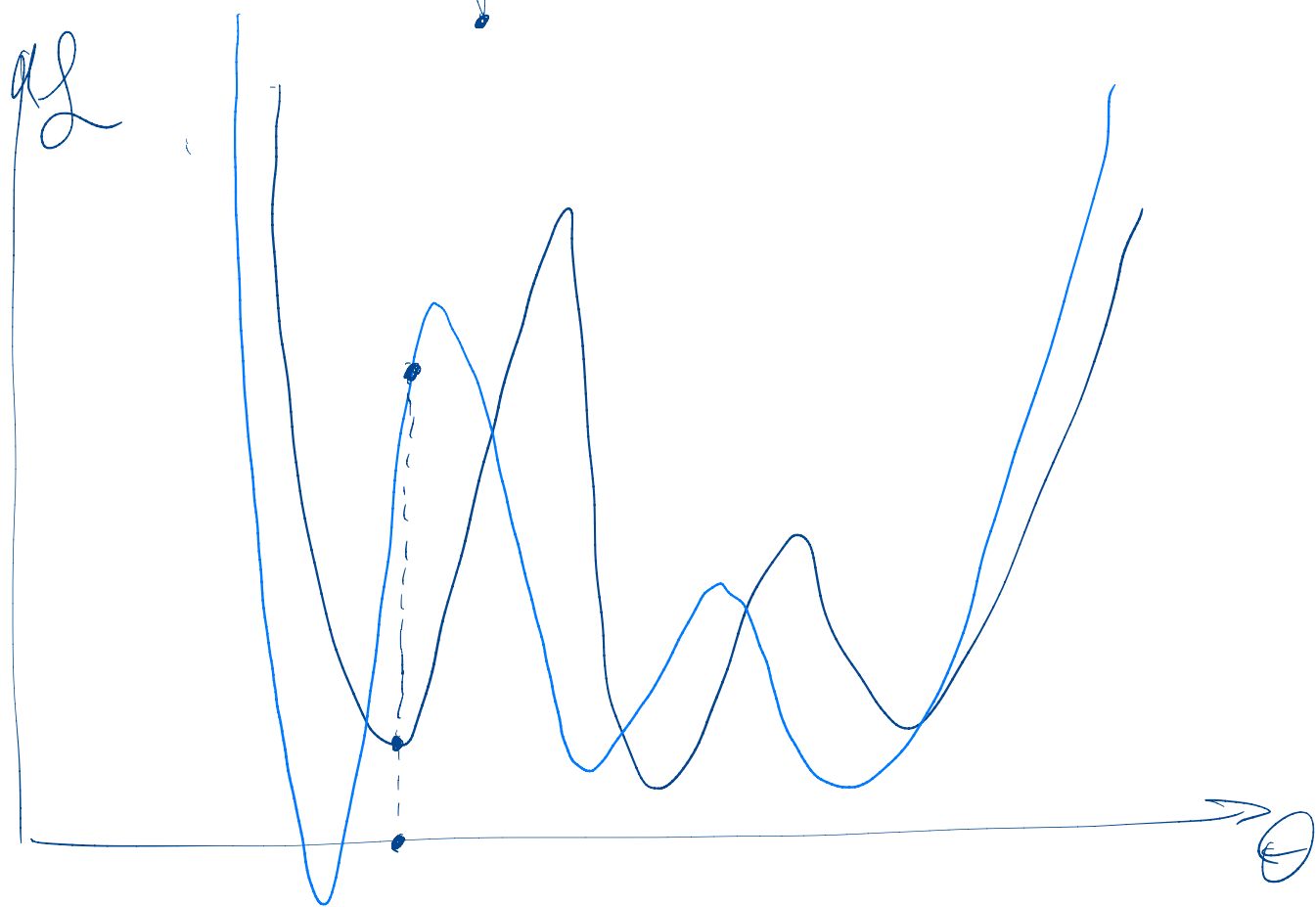
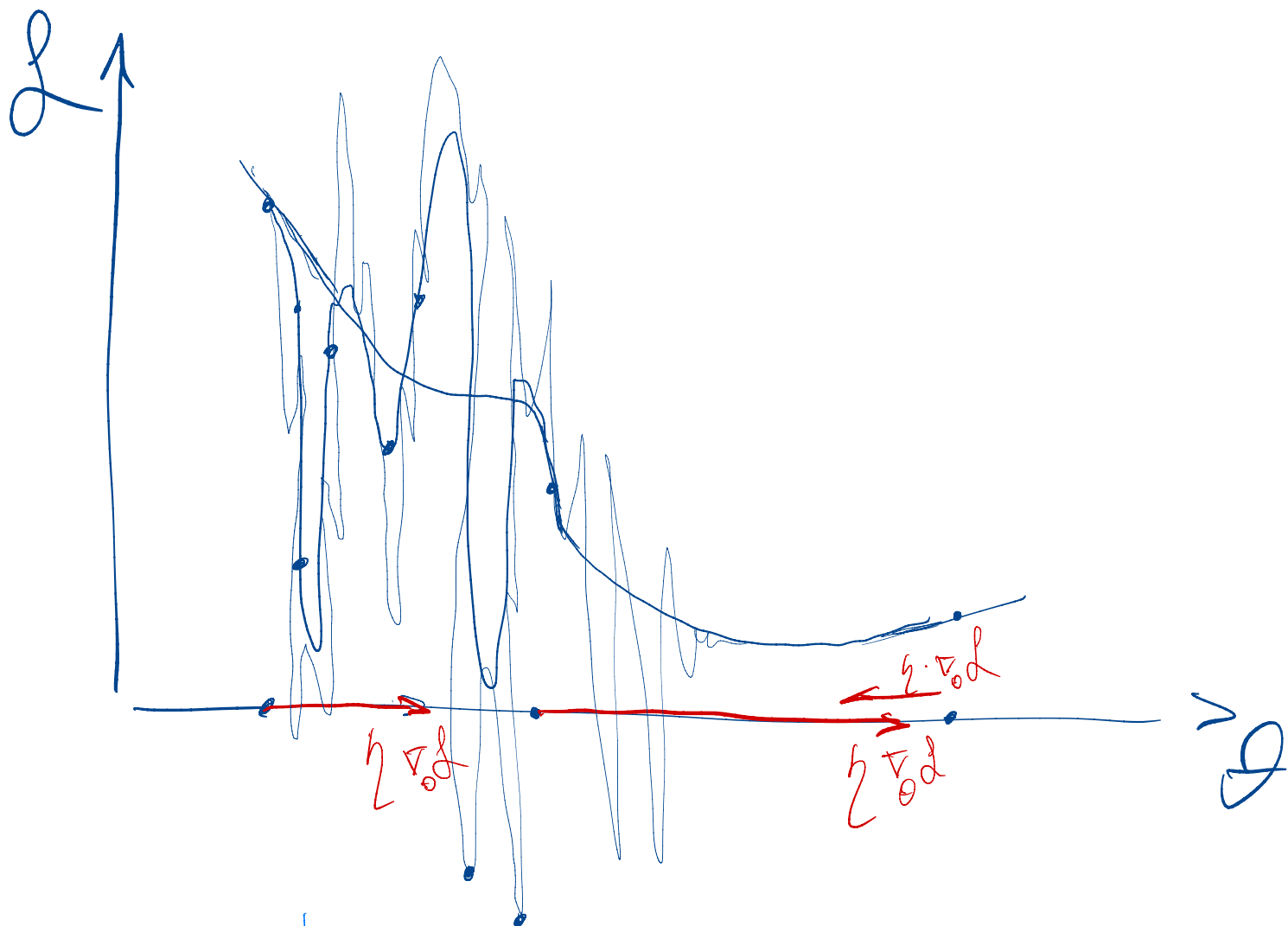


Локальный минимум:

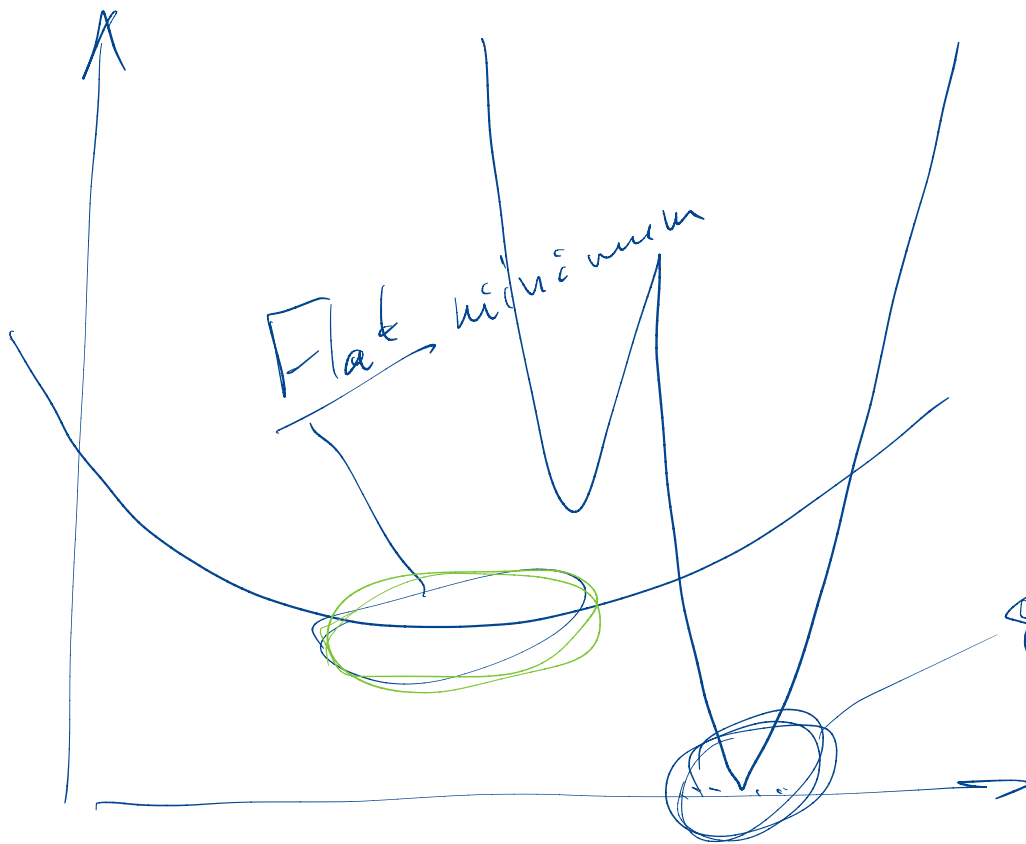
$$\nabla L = 0$$

$$\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}$$





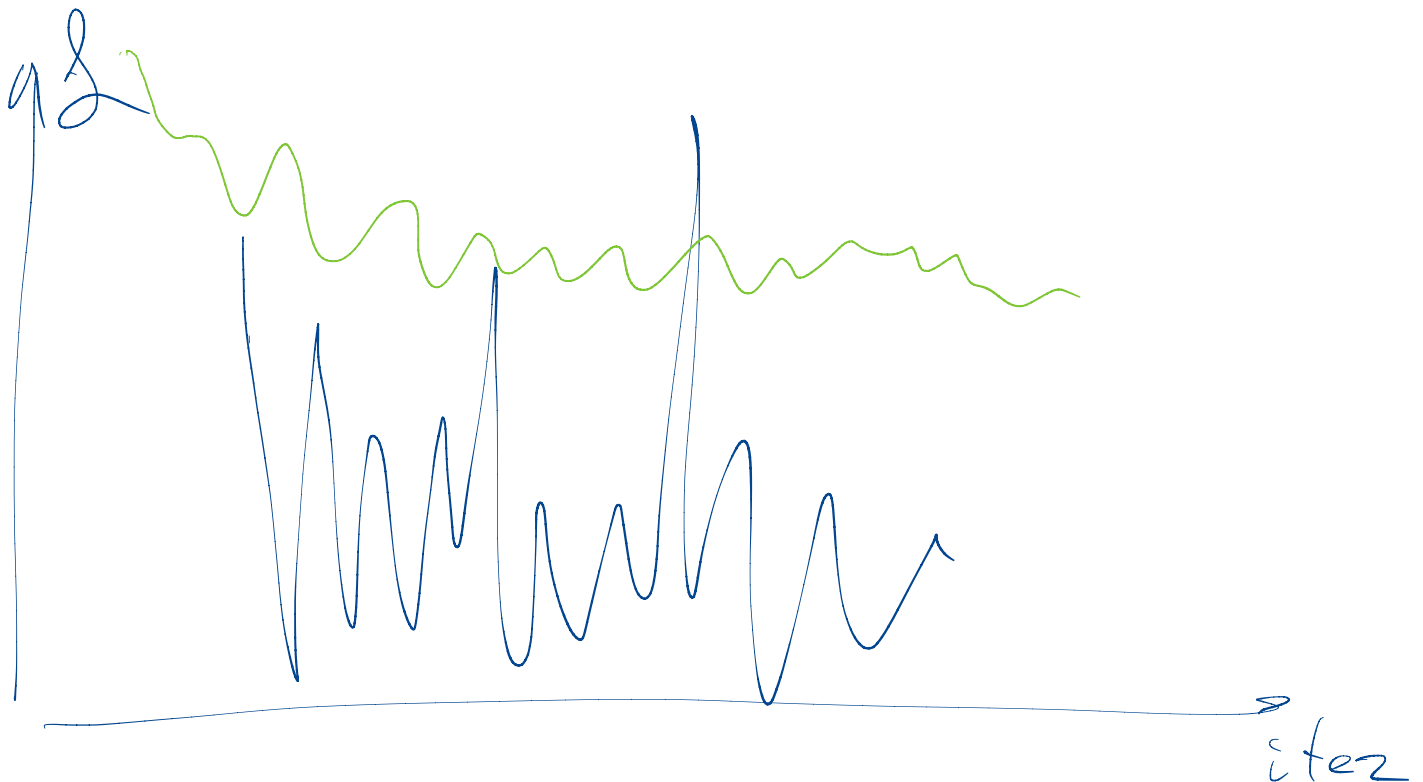
Sharp minima VS. Flat minima

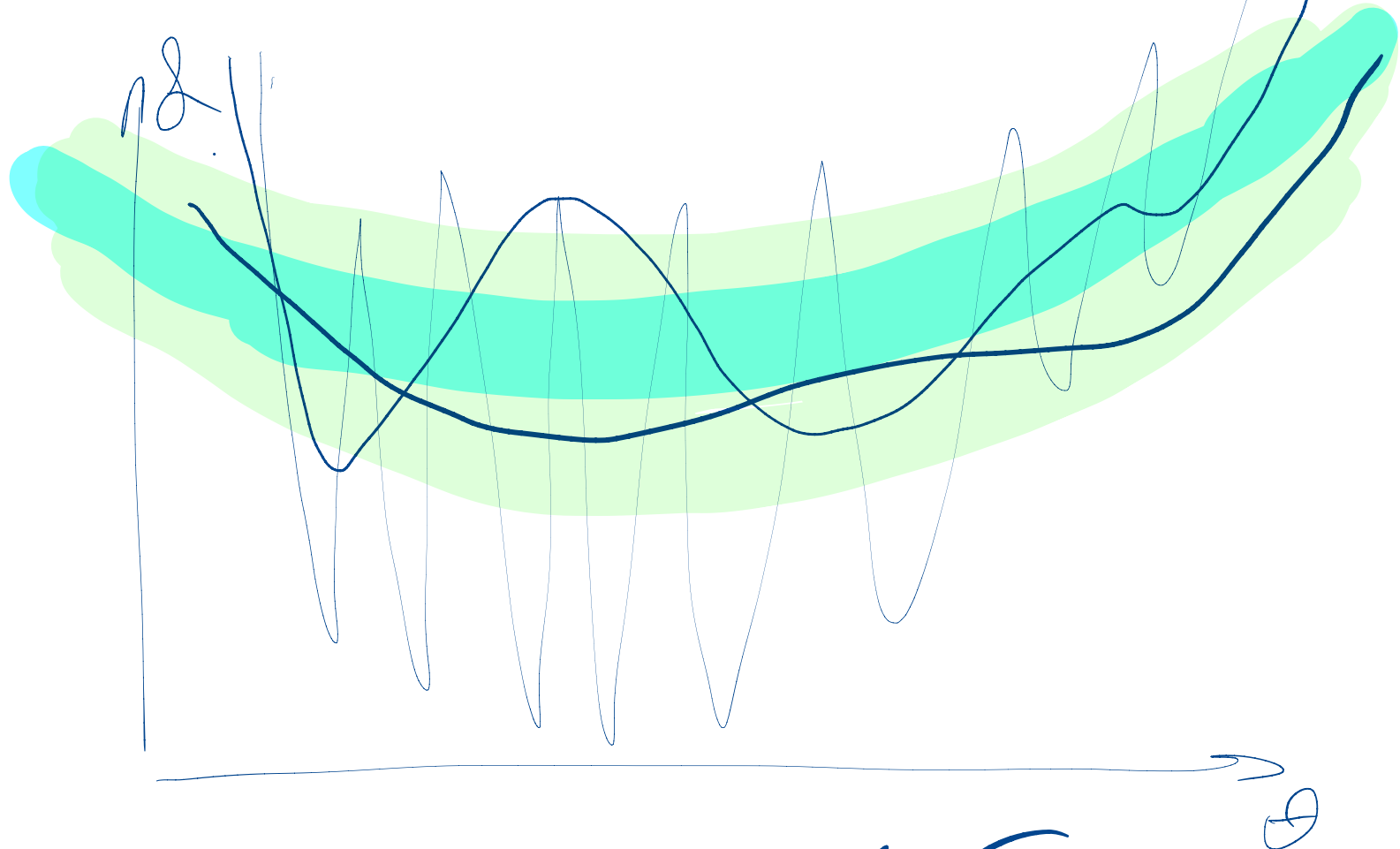


Flat minimum
↑ oscillations en.

Sharp minimum
↓ oscillations en.

Sharp minimum





Эффекты шума в обучении
ЦНС

① Величина ζ



② Величина batch size

