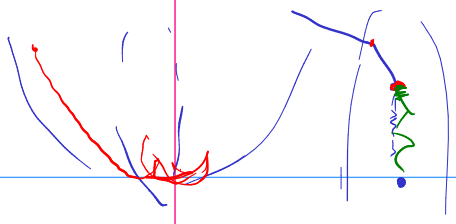
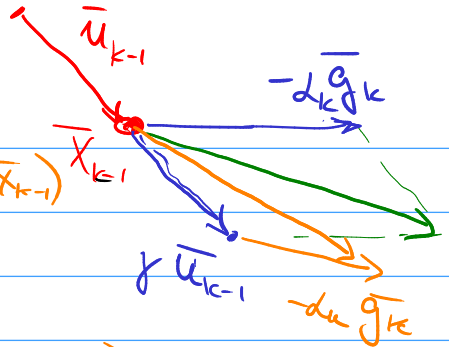


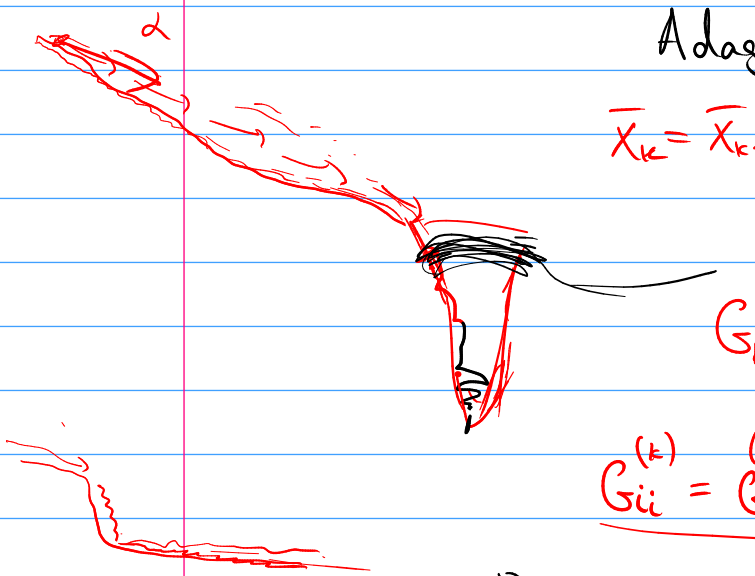
Momentum

$$\bar{u}_k = \gamma \bar{u}_{k-1} - \alpha_k \bar{g}_k$$



NAG Nesterov accelerated gradient

$$\bar{u}_k = \gamma \bar{u}_{k-1} - \alpha_k \nabla_{\bar{x}} F\left(\bar{x}_{k-1} + \gamma \bar{u}_{k-1}\right)$$



Adagrad

$$\bar{x}_k = \bar{x}_{k-1} - \frac{\alpha_0}{\sqrt{G_k + \varepsilon}} \cdot \bar{g}_k$$

$$G_k = \begin{pmatrix} G_{ii} & & \\ & \ddots & \\ & & \end{pmatrix}$$

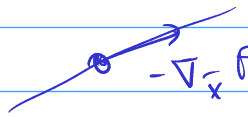
$$G_{ii}^{(k)} = G_{ii}^{(k-1)} + g_{k-1}^2 .$$

$$\left\| F(\bar{x}^*) - F(\bar{x}^{(k)}) \right\| \le \frac{R^2 + G^2 \sum_k \alpha_i^2}{2 \sum_k \alpha_i}$$

$$\frac{R^2}{2kh} + \frac{G^2 h}{2}$$

RMSprop

$$G_{ii}^{(k)} = (1-p)\, G_{ii}^{(k-1)} + p\, g_{k-1,i}^2$$



$$\bar{x}_k = \bar{x}_{k-1} - \frac{\alpha_0}{\sqrt{G^{(k)} + \varepsilon}} \cdot \nabla_{\bar{x}} F(\bar{x}_{k-1})$$

c c M/c

M^2/c^2

$$\bar{x}_k = \bar{x}_{k-1} - H_{k-1}^{-1} \cdot \nabla_{\bar{x}} F(\bar{x}_{k-1})$$

c c $(M/c^2)^{-1}$ M/c

$$\bar{x}_k = \bar{x}_{k-1} - \alpha_k \nabla_{\bar{x}} F(\bar{x}_{k-1})$$

c c M/c

Adadelta

$$\bar{x}_k = \bar{x}_{k-1} - \alpha \cdot \frac{\sqrt{E[\Delta \bar{x}^2]^{(k)} + \varepsilon}}{\sqrt{G^{(k)} + \varepsilon}} \cdot \bar{g}_{k-1}$$

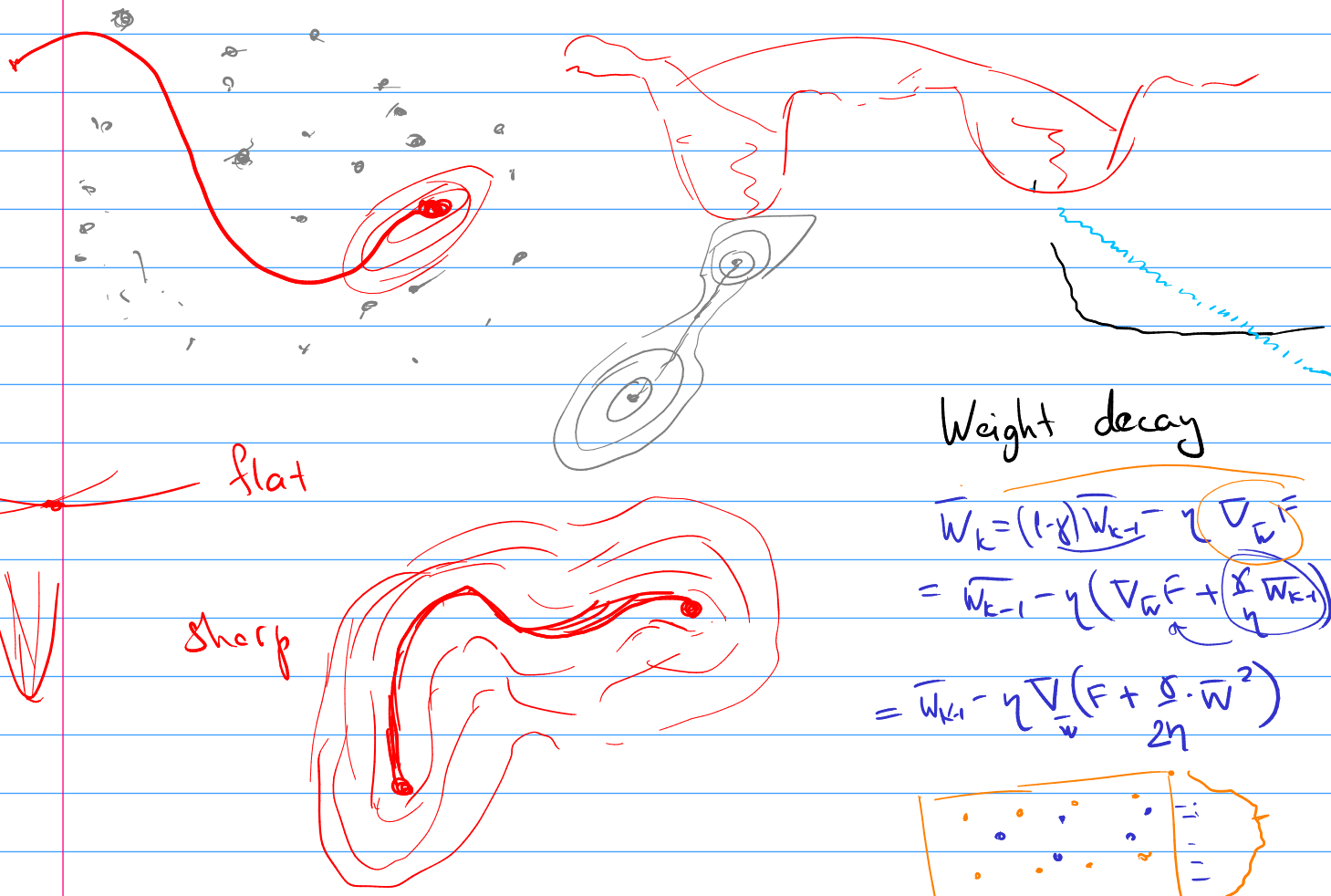
$$E[\bar{u}^2]^{(k)} = (1-p)\, E[\bar{u}^2]^{(k-1)} + p \cdot \bar{u}^{(k-1)^2}$$

Adam

$$\bar{x}_t = \bar{x}_{t-1} - \frac{\alpha}{\sqrt{\bar{v}_t + \varepsilon}} \cdot \bar{m}_t$$

$$\bar{v}_t = \beta_2 \bar{v}_{t-1} + (1-\beta_2)\bar{g}_t^2$$

$$\bar{m}_t = \beta_1 \bar{m}_{t-1} + (1-\beta_1)\bar{g}_t$$



flat

sharp

Weight decay

$$\bar{w}_k = (1-\gamma)\bar{w}_{k-1} - \eta \nabla_{\bar{w}} f$$

$$= \bar{w}_{k-1} - \eta\left(\nabla_{\bar{w}} F + \frac{\gamma}{\eta}\bar{w}_{k-1}\right)$$

$$= \bar{w}_{k-1} - \eta \nabla_{\bar{w}}\left(F + \frac{\gamma}{2\eta}\cdot \bar{w}^2\right)$$