2. problem:

$\mathcal{L} = (1 - [1, 0.1, 0.1]\theta)^2 \Rightarrow$ solution is all θ that satisfy $[1, 0.1, 0.1]\theta = 1$

$\downarrow$

a flat set of minimizers.

$\Downarrow$

Q:
which one of these infinitely many solutions does each optimizer converge to?

$M_t : [M_t]_{ii} = \frac{1}{\sqrt{\mathbb{E}(g_{t,i}^2)}}$ $\nabla_{\theta_i} f_t$ $\frac{1}{T}\sum_t^T g_{t,i}^2$

(a) vanilla SGD ($M_t = I$, small constant $\alpha > 0$)

$\nabla f(\theta) = -2x(1 - x^T\theta)$. $\Rightarrow \theta = \frac{x}{x^T x} = \frac{1}{1.0101}[1, 0.1, 0.01]$

lies along the span of x.

(b) simplified Adam

$\Delta\theta_i = -\alpha (M_t)_{ii} g_i$. $\Longrightarrow \Delta\theta_i = -\alpha \frac{1}{\sqrt{\mathbb{E}(g_i^2)}} \cdot (-2x_i)(1 - x^T\theta)$

plug in $g_i = -2x_i(1 - x^T\theta)$

$= -\alpha \cdot \frac{\sqrt{T}}{2x_i \sqrt{\sum_t^T (1 - x^T\theta)^2}} \cdot (-2x_i)(1 - x^T\theta)$.

$g_i^2 = 4x_i^2(1 - x^T\theta)^2$.

$\sqrt{\mathbb{E}(g_i^2)} = \sqrt{\frac{1}{T}\sum_t^T g_{t,i}^2} = \frac{2x_i}{\sqrt{T}}\sqrt{\sum_t^T (1 - x^T\theta_t)^2}$. $= \alpha\sqrt{T}\sqrt{\sum_t^T (1 - x^T\theta_t)^2}\,(1 - x^T\theta)$.

$g_i = -2v_i(1 - v^T\theta)$ $|g_i| = 2|v_i||1 - v^T\theta|$

$\therefore [M_t]_{ii} = \frac{1}{|g_i|} = \frac{1}{2|v_i||1 - v^T\theta|}$

$\therefore \Delta\theta_i = -\alpha_t [M_t]_{ii} g_{t,i} = -\alpha_t \frac{1}{2|v_i||1 - v^T\theta|}(-2v_i)(1 - v^T\theta)$

$= \alpha_t \,\mathrm{sign}(v_i)$.

$\therefore \Delta\vec{\theta} = -\alpha_t M_t g_t = 2\alpha_t \,\mathrm{sign}(v) = 2\alpha_t \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Update are all in the direction $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.