# HOMEWORK 5. STOCHASTIC GRADIENT DESCENT

**Kun Yuan**

October 25, 2023

**Attention**: Turn in your homework at the beginning of our lecture on Oct. 31, 2023

## 1 Linear neural network is non-convex

Linear neural network, e.g.,

$$f(W_1, W_2, h_1, h_2; \xi) = \frac{1}{2}\|W_2(W_1\xi - h_1) - h_2\|^2 \quad (\xi \text{ is the input data}),$$

is non-convex. To illustrate it, please prove that the following simplified linear neural network

$$f(w_1, w_2) = \frac{1}{2}(w_1 w_2)^2$$

is non-convex in terms of $(w_1, w_2)$, where $w_1 \in \mathbb{R}^d$ and $w_2 \in \mathbb{R}^d$.

## 2 Mini-batch SGD

Mini-batch SGD iterates as follows:

$$g_k = \frac{1}{B}\sum_{b=1}^{B} \nabla F(x_k; \xi_k^{(b)}), \tag{1a}$$

$$x_{k+1} = x_k - \gamma g_k. \tag{1b}$$

where $B$ is the batch size. Please introduce necessary assumptions on stochastic gradient and establish the convergence rate of the above mini-batch SGD in the smooth and non-convex scenario with both constant and decaying learning rate. Write down the detailed derivation.