# Question 1.4

Tuesday, April 21, 2020    6:00 PM

$$g_{0,k} = \nabla_{\theta_0} L(\theta)$$

$$= -2\epsilon_k^T \frac{\partial f_\theta(y_k)}{\partial \theta_0}$$

$$= -2\epsilon_k^T \frac{\partial f_{1,\theta_1}(z_1)}{\partial z_1} \cdot \frac{\partial z_1}{\partial \theta_0}$$

$$= -2\epsilon_k^T A_{1,k} \cdot \frac{\partial f_{0,\theta_0}(z_0)}{\partial \theta_0}$$

$$= -2\epsilon_k^T A_{1,k} \cdot B_{0,k}$$

$$\Rightarrow g_{0,k} = \underbrace{-2\epsilon_k^T A_{1,k} B_{0,k}}_{\text{Fwd grad}} \Longleftrightarrow \underbrace{B_{0,k}^t A_{1,k}^t (-2\epsilon_k)}_{\text{Adjoint grad}}$$

Gradient of Total loss function:

$$\Rightarrow g_0 = \sum_{k=0}^{K-1} g_{0,k}$$

$$\boxed{g_0 = \sum_{k=0}^{K-1} B_{0,k}^t A_{1,k}^t (-2\epsilon_k)}$$