$$\mathbf{g} := \mathbb{E}_{\mathbf{x}, y \in B_i}\left[\frac{\partial \mathscr{L}(\mathbf{x}, y \mid \theta)}{\partial \theta}\right]$$

$$\mathbf{v} := \beta_1 \mathbf{v} + (1 - \beta_1)\mathbf{g}$$

$$\mathbf{m} := \beta_2 \mathbf{m} + (1 - \beta_2)\mathbf{g}^2$$

$$s := \epsilon \frac{\sqrt{1 - \beta_2^{\text{step}}}}{1 - \beta_1^{\text{step}}}$$

$$\theta := \theta - s \frac{\mathbf{v}}{\sqrt{\mathbf{m}} + \varepsilon}$$