

Gradients

$$\nabla_{\phi} f(w, \phi) = \sum_i \mathbb{E}_{p(\varepsilon_i)} \nabla_{\phi} \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

$$p(\varepsilon_i) = \mathcal{N}(0, I)$$

$$t_i = \varepsilon_i \odot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

