

The score function method or also known as the log-derivative trick:

$$\nabla_x \log(f(x)) = \frac{\nabla_x f(x)}{f(x)}$$

*Proof.*

$$\begin{aligned}\nabla_\psi \mathbb{E}_{q_\psi(\mathbf{z})}[f(\mathbf{z})] &= \nabla_\psi \mathbb{E}_{p_\psi(\boldsymbol{\epsilon})}[f(T(\boldsymbol{\epsilon}, \psi))] \\&= \nabla_\psi \int_{\Omega} p_\psi(\boldsymbol{\epsilon}) f(T(\boldsymbol{\epsilon}, \psi)) d^n \boldsymbol{\epsilon} \\&= \int_{\Omega} p_\psi(\boldsymbol{\epsilon}) \nabla_\psi f(T(\boldsymbol{\epsilon}, \psi)) d^n \boldsymbol{\epsilon} + \int_{\Omega} \nabla_\psi p_\psi(\boldsymbol{\epsilon}) f(T(\boldsymbol{\epsilon}, \psi)) d^n \boldsymbol{\epsilon} \\&= \mathbf{g}^{rep} + \int_{\Omega} \nabla_\psi p_\psi(\boldsymbol{\epsilon}) f(T(\boldsymbol{\epsilon}, \psi)) d^n \boldsymbol{\epsilon} \\&= \mathbf{g}^{rep} + \int_{\Omega} \frac{\nabla_\psi p_\psi(\boldsymbol{\epsilon})}{p_\psi(\boldsymbol{\epsilon})} p_\psi(\boldsymbol{\epsilon}) f(T(\boldsymbol{\epsilon}, \psi)) d^n \boldsymbol{\epsilon} \\&= \mathbf{g}^{rep} + \int_{\Omega} p_\psi(\boldsymbol{\epsilon}) f(T(\boldsymbol{\epsilon}, \psi)) \nabla_\psi \log(p_\psi(\boldsymbol{\epsilon})) d^n \boldsymbol{\epsilon} \\&= \mathbf{g}^{rep} + \mathbf{g}^{corr}\end{aligned}$$

□