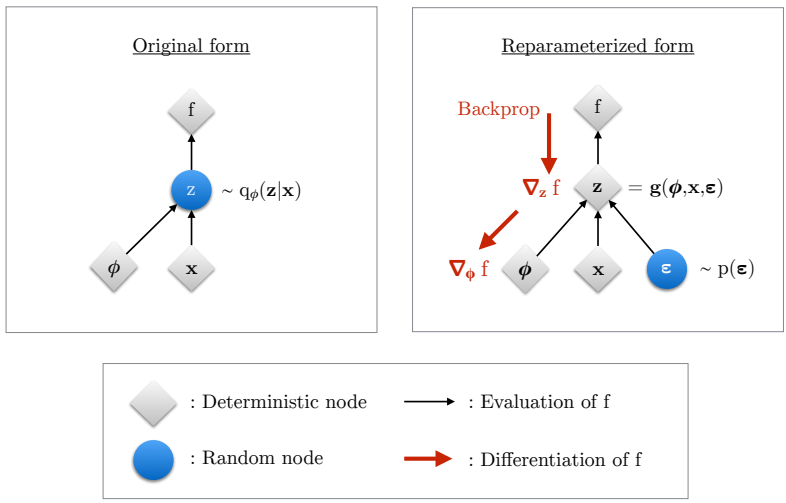

Figure 2.3: Illustration of the reparameterization trick. The variational parameters ϕ affect the objective f through the random variable $\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})$. We wish to compute gradients $\nabla_\phi f$ to optimize the objective with SGD. In the original form (left), we cannot differentiate f w.r.t. ϕ, because we cannot directly backpropagate gradients through the random variable $\mathbf{z}$. We can 'externalize' the randomness in $\mathbf{z}$ by re-parameterizing the variable as a deterministic and differentiable function of ϕ, $\mathbf{x}$, and a newly introduced random variable $\boldsymbol{\epsilon}$. This allows us to 'backprop through $\mathbf{z}$', and compute gradients $\nabla_\phi f$.