

Initial approximation
$$\widehat{\mathcal{H}}(q,p) = W_{L+1}\Phi^{(L)}(q,p) + b_{L+1}$$

$$\{W_l,b_l\}_{l=1}^L \quad \{W_{L+1},b_{L+1}\}$$

$$\{W_l,b_l\}_{l=1}^L \text{ are sampled (unsupervised)}$$

$$\{W_{L+1},b_{L+1}\} = \arg\min \mathcal{L}(\nabla \Phi^{(L)},\mathcal{D})$$

