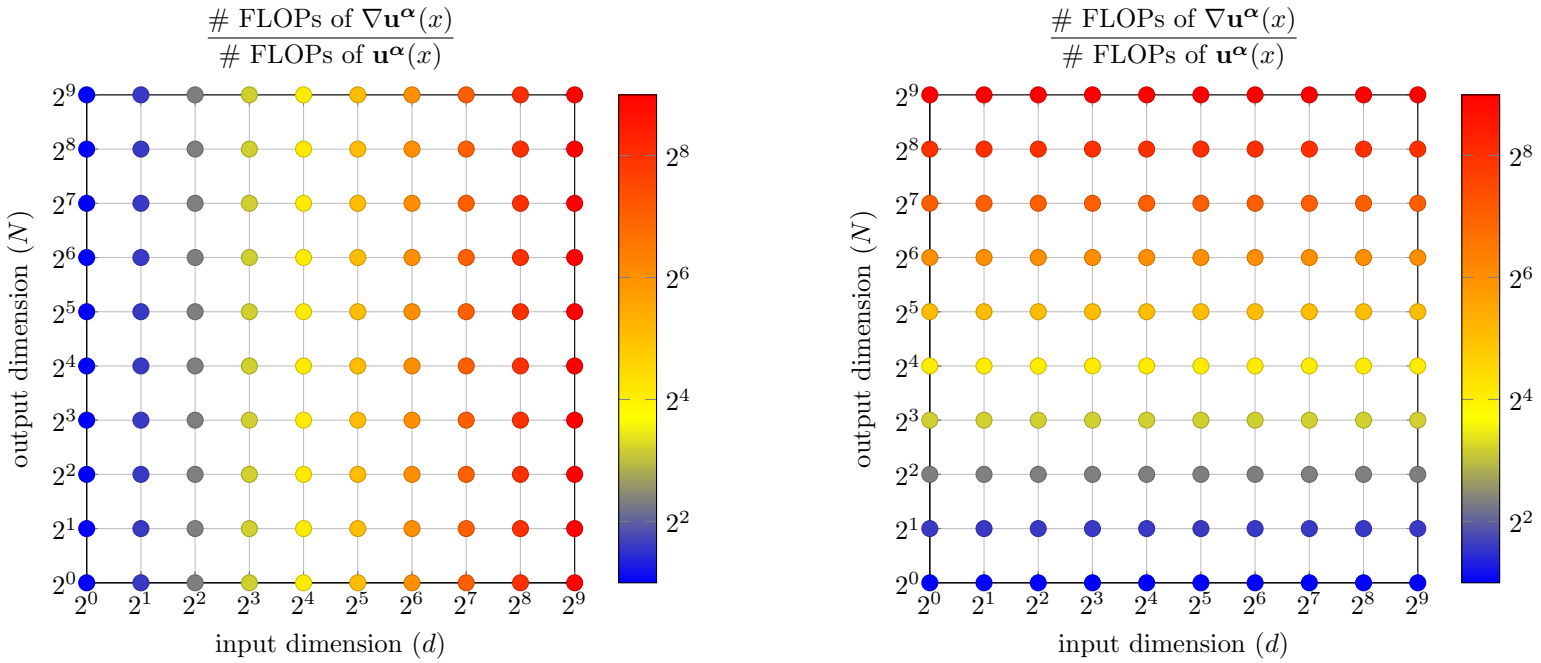

(a) Forward AD.

(b) Backward AD.

Figure 1: Costs of computing the gradient of a vector-valued neural network $\mathbb{R}^d \ni x \mapsto \mathbf{u}^{\boldsymbol{\alpha}}(x) \in \mathbb{R}^N$ with input dimension d, five hidden layers of width 2^{10}, and a final activated layer of width N. Experimentation is performed for $d, N \in \{2^0, 2^2, \ldots, 2^9\}$.