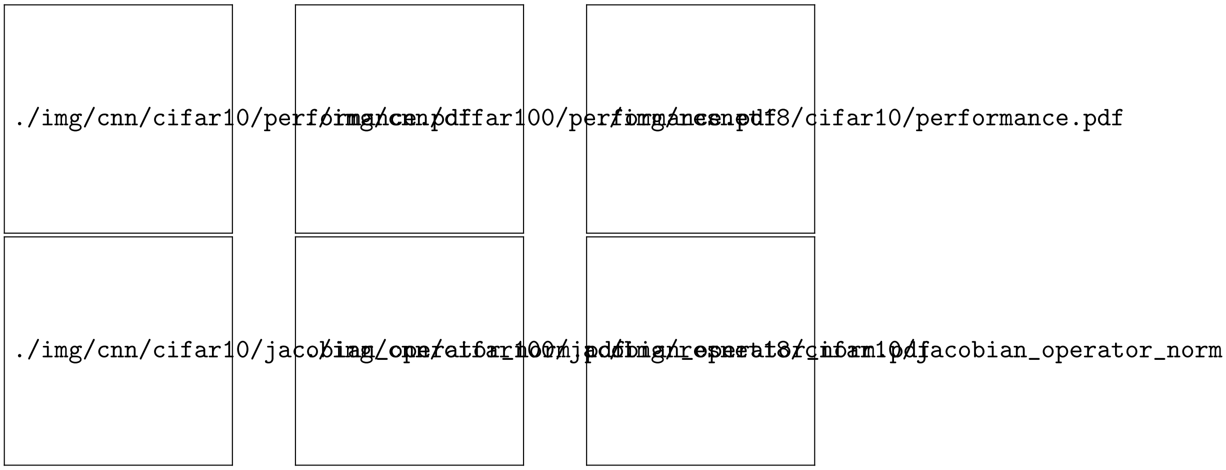

Figure A: (Top) **Train error** (dashed) and **test error** (solid) for our experimental setting, with the test error undergoing double descent as model size increases. (Left to right) ConvNets trained on CIFAR-10 (left) and CIFAR-100 (middle), and ResNets trained on CIFAR-10 (right). (Bottom) **Empirical Lipschitz constant** for the same models. For each setting, the Lipschitz depends non-monotonically on model size, strongly correlating with double descent. This finding questions the utility and validity of uniformly bounded Lipschitz assumptions in representing the hypothesis class of trained networks.

The present document contains revised figures from the main submission. Figure **??** extends Figure 1 from the main paper, presenting an upper bound on the true Lipschitz constant, together with the lower bound studied in the main paper.

Figure **??** extends the results presented in Figure 2 of the main paper, presenting the trend of the smallest loss Hessian eigenvalue $\lambda_r(H)$ as model size increases. Figure **??** also tracks the loss input-space gradient $\mathbb{E}_{\mathcal{D}}\|\nabla_{\mathbf{x}}\mathcal{L}(\theta, x, y)\|$, which is shown in Theorem 2 to be upper bounded by the mean curvature $\operatorname{tr} H$.

Figure B: (Top) **Mean loss curvature** (Hessian trace) in parameter space. (Bottom) **Maximum curvature** for the loss in parameter space. From left to right: ConvNets trained on CIFAR-10 (left), CIFAR-100 (middle) and ResNets trained on CIFAR-10 (right). In all settings, mean and maximum parameter-space curvature strongly correlate with double descent, peaking at the interpolation threshold, and highlighting a nonlinear dependence on network width. All values are reported in log-y scale to better separate models in the interpolating regime.