Efficient and Accurate Estimation of Lipschitz Constants for Deep Neural Networks

Mahyar Fazlyab, Alexander Robey
Hamed Hassani, Manfred Morari, George J. Pappas

NeurIPS 2019

Lipschitz Constant of Neural Networks

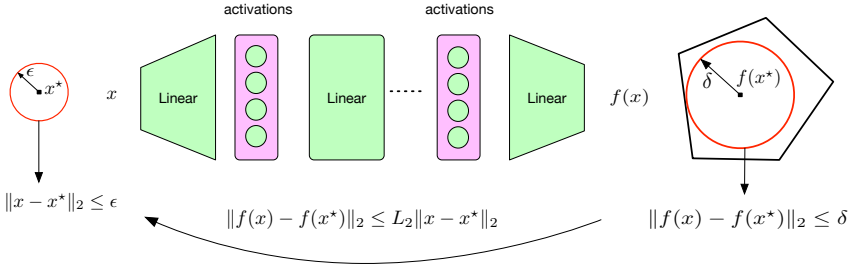
- **Definition:** the *smallest* L_2 such that

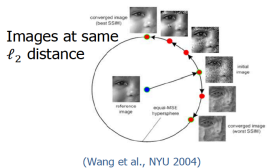
$$\|f(x) - f(y)\|_2 \leq L_2 \|x - y\|_2 \quad \forall x, y \in \mathbb{R}^{n_x}$$

  where $f \colon \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ is represented by a NN

- **Why important:** *tight* upper bound on L_2 useful in
  - Robustness certification of classifiers
  - Closed-loop stability analysis of systems with neural network controllers
  - Robust training
  - Generalization bounds
- **Challenge:** finding L_2 is NP-hard

Robustness Certification of Classifiers





(Wang et al., NYU 2004)

Lower Lipschitz constant implies more robustness

Estimation of Lipschitz Constant

- Feed-forward fully-connected neural network

$$x^{k+1} = \phi(W^k x^k + b^k) \quad k = 0, \cdots, \ell - 1 \quad f(x^0) = W^\ell x^\ell + b^\ell$$

- Product of norms ($\prod_{k=0}^{\ell} \|W^k\|_2$) is **overly conservative**

- We improve this bound by orders of magnitude using convex optimization
  - Example: a randomly generated NN with 8 layers:

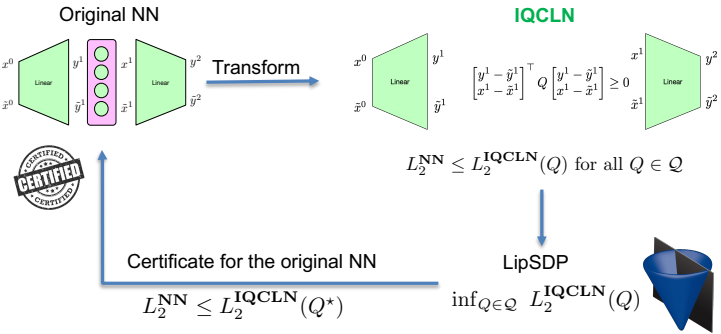
$$\text{Product of Norms} \approx 9571 (= \prod_{k=0}^{\ell} \|W^k\|_2)$$

$$\text{Our Bound} \approx 104$$

- **Current Status:** scales to small CNNs (10k neurons)
- **Future Work:** scale to large CNNs (100k neurons)

Our Main Idea

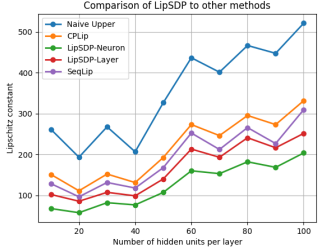
- Transform NN into a family of **I**ncrementally **Q**uadratically **C**onstrained **L**inear **N**etworks



- **LipSDP:** Estimating Lipschitz constants of NNs via Semidefinite Programming

Tightness of the Bounds

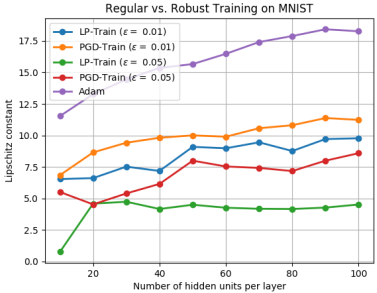
- **Platform:** MATLAB, CVX toolbox, and MOSEK on a 9-core CPU with 16GB of RAM
- **Methods:**
  - Variants of **LipSDP**: **LipSDP-Network**, **LipSDP-Neuron**, **LipSDP-Layer**
  - **CPLip:** Combettes, Patrick L., and Jean-Christophe Pesquet. "Lipschitz Certificates for Neural Network Structures Driven by Averaged Activation Operators."arXiv preprint arXiv:1903.01014(2019).
  - **SeqLip:** Virmaux, Aladin, and Kevin Scaman. "Lipschitz regularity of deep neural networks: analysis and efficient estimation."Advances in Neural Information Processing Systems. 2018.

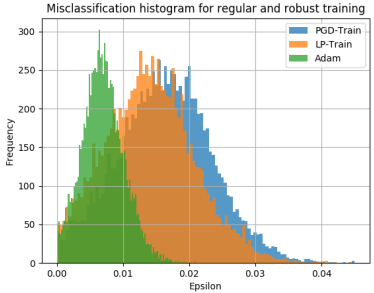


- Our bounds are the tightest in the literature

Experiments on MNIST

- **Training Methods:**
  - **Adam:** Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980(2014).
  - **LP-Train:** Wong, Eric, and J. Zico Kolter. "Provable defenses against adversarial examples via the convex outer adversarial polytope." arXiv preprint arXiv:1711.00851(2017).
  - **PGD-Train:** Madry, Aleksander, et al. "Towards deep learning models resistant to adversarial attacks." arXiv preprint arXiv:1706.06083(2017).





Thank You!