SDP-CROWN: Efficient Bound Propagation for Neural Network Verification with Tightness of Semidefinite Programming

Hong-Ming Chiu[1], Hao Chen[1], Huan Zhang[1], Richard Y. Zhang[1]

[1] ECE Department, University of Illinois at Urbana–Champaign.

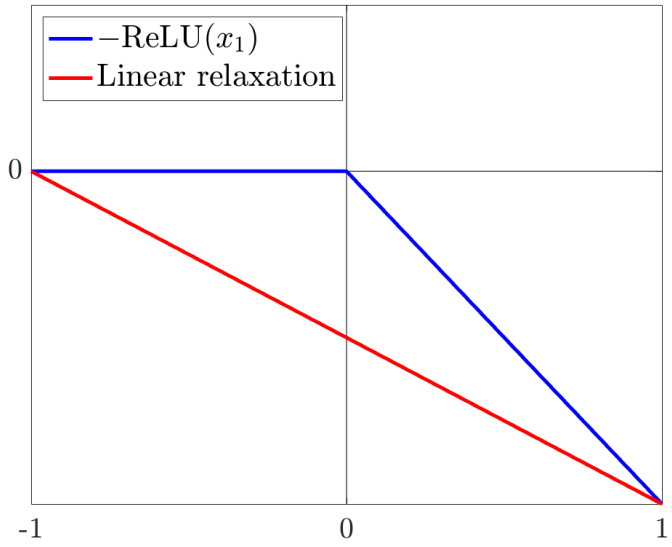
Neural network verification

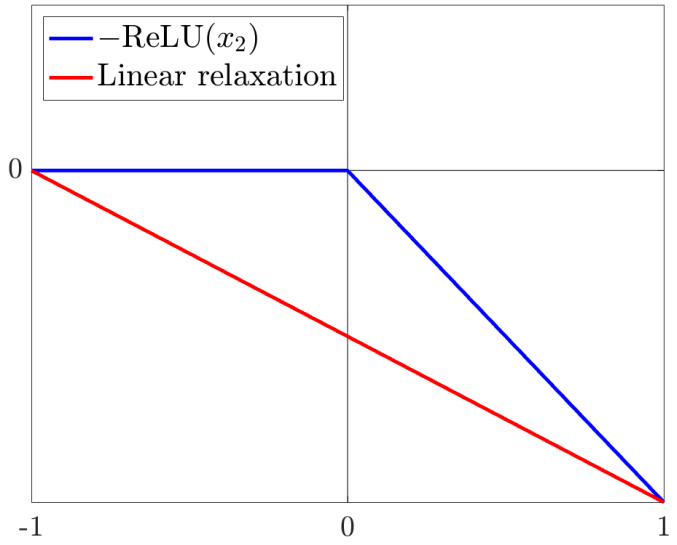
- Neural network (NN) verification aims to guarantee consistent model behavior in the presence of small perturbations to the input.
- **Bound propagation** is one of the leading methods for NN verification because it is highly scalable.
  - **For elementwise perturbations**, where each neuron is perturbed **independently**, bound propagation works remarkably well.
  - **However, for ℓ_2-norm perturbations**, which introduce **inter-neuron coupling**, bound propagation becomes loose and overly conservative.

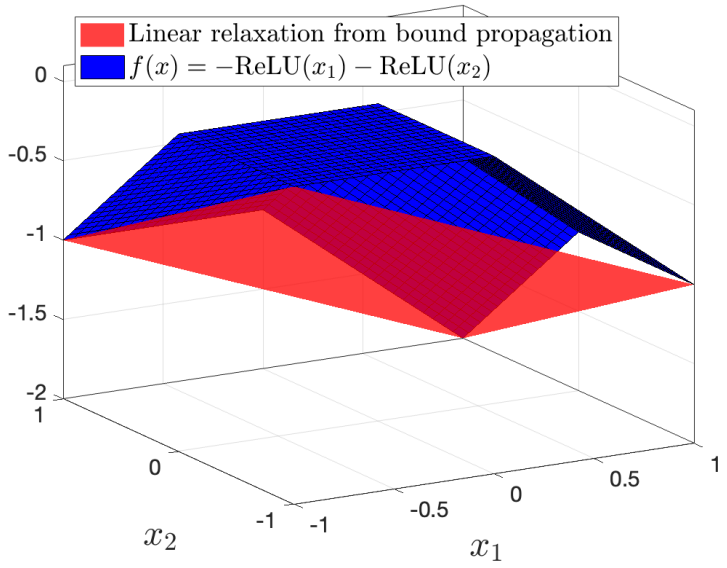
Bound propagation

Neural network: $\mathrm{f}(x) = -\mathrm{ReLU}(x_1) - \mathrm{ReLU}(x_2)$.

- **Elementwise perturbations:** $-1 \leq x_1 \leq 1$ and $-1 \leq x_2 \leq 1$.



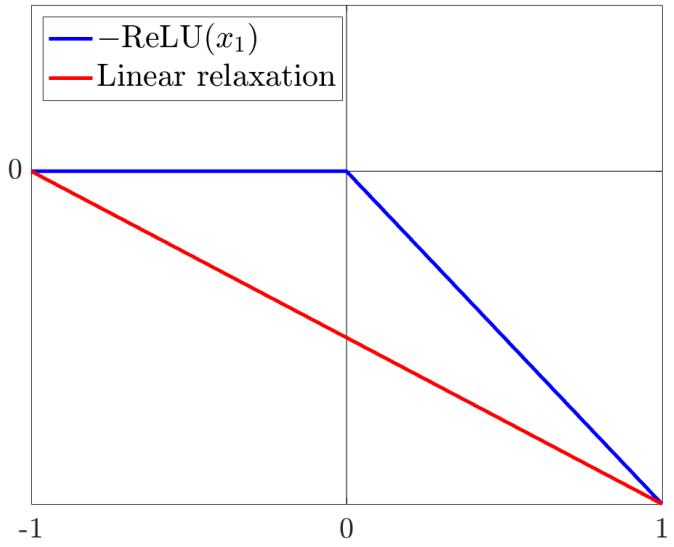


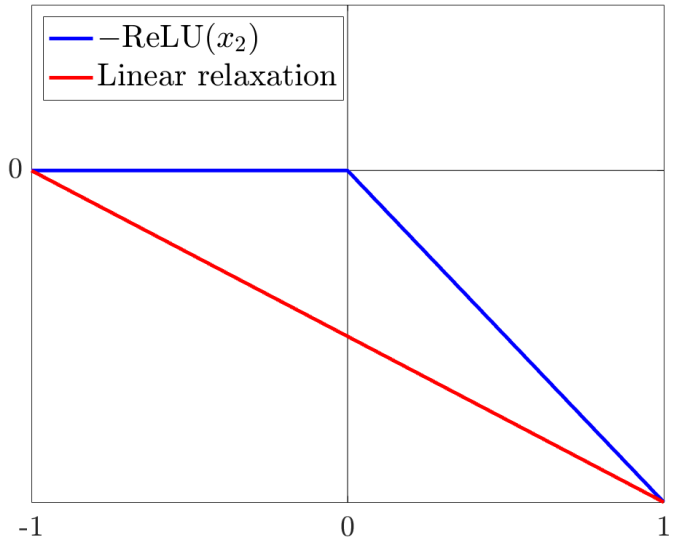


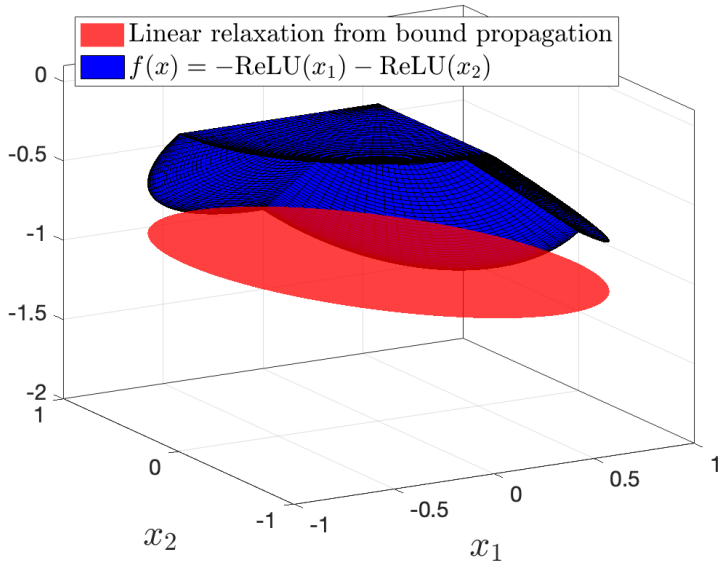
Why does bound propagation fail?

Neural network: $\mathrm{f}(x) = -\mathrm{ReLU}(x_1) - \mathrm{ReLU}(x_2)$.

- **ℓ_2-norm perturbations:** $x_1^2 + x_2^2 \leq 1$.





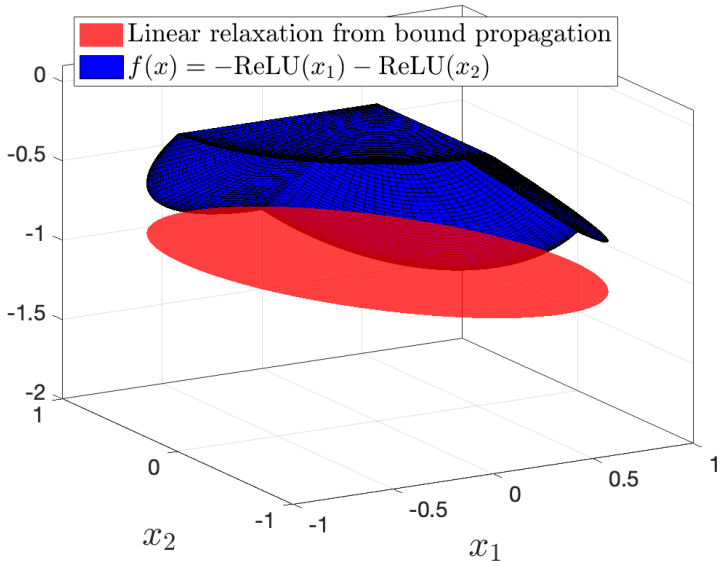


Our method: SDP-CROWN

Neural network: $\mathrm{f}(x) = -\mathrm{ReLU}(x_1) - \mathrm{ReLU}(x_2)$

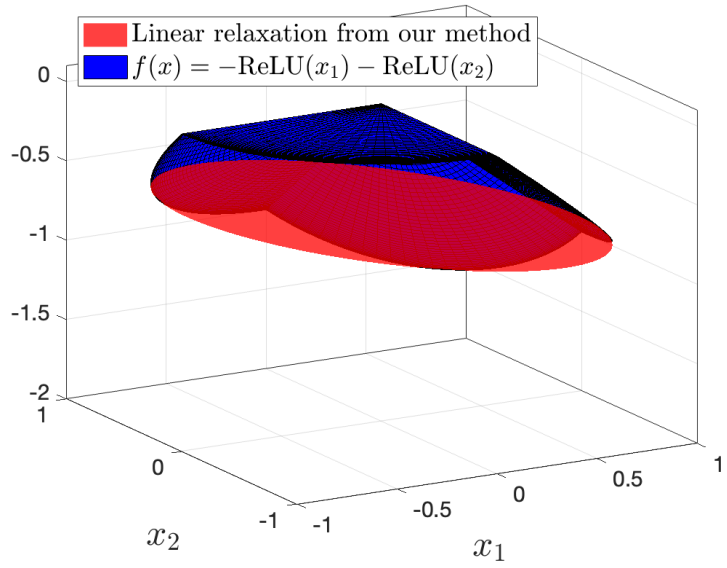
- **ℓ_2-norm perturbations: $x_1^2 + x_2^2 \leq 1$.**

Bound propagation



Model inter-neuron coupling by optimizing a scalar variable per layer

SDP-CROWN



Results summary

- **Significantly tighter** than existing bound propagation methods for certifying ℓ_2 adversaries.
- **Enjoying the same level of scalability** as bound propagation methods.

For more details, please talk a look at our paper!

Thank you for your attention!