Expressiveness of Multi-Neuron Convex Relaxations in Neural Network Certification

Yuhao Mao*, **Yani Zhang***, **Martin Vechev**

Department of Computer Science
Department of Information Technology and Electrical Engineering
ETH zürich SRILAB

We study multi-neuron convex relaxation in neural network certification, revealing (perhaps surprisingly):

- A universal convex barrier, that no convex relaxation can be complete verifiers
- Recovered expressiveness, that every function has a realizing network bounded exactly by multi-neuron relaxations, which was proven impossible for single-neuron relaxations

Neural Network Certification

Certification: given an input set S, prove a property holds or fails for S.

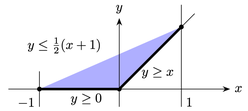
Example (robustness):
$f(x')_{i^*} - f(x')_i \geq 0, \forall i, x' \text{ s.t. } \|x' - x\|_\infty \leq \epsilon.$

Generally coNP-hard.

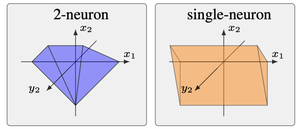
Convex Relaxation: over-approximate the embedding space to certify

Example (IBP): use a hyper-cube

Example (Triangle): use the convex hull of a single neuron



Example (Multi-neuron): over-approximate multiple neurons jointly



Relaxation Precision: given a neural network f, the relaxation error is measured by the gap between its functional upper (lower) bound and the upper (lower) bound derived by the convex relaxation.

Known Limitations of Single-Neuron Relaxations

- **Convex barrier:** Even the most precise single-neuron convex relaxation, when solved exactly, cannot produce exact bounds for some toy networks.
- **Expressiveness barrier:** There exist toy functions such that no network realizing them can be bounded exactly by the most precise single-neuron relaxation.

Open Problems Addressed:

- **Is there a convex relaxation precise enough to certify every network given sufficient resources?** ✗
- **Do multi-neuron relaxations offer expressiveness advantages over single-neuron relaxations?** ✓

A Universal Convex Barrier

Statement (informal): For every non-empty input set, every non-polynomial activation, every convex relaxation with finite resources, there exists a neural network such that the relaxation error is unbounded.

- Convex relaxations cannot be complete verifiers even with arbitrary computational resources —— a fundamental limitation.
- This discourages sacrificing computational efficiency for improved precision beyond current methods.

Resolving Expressiveness Barrier

Statement (informal): For every continuous piecewise linear function, there exists a ReLU network implementing this function and the relaxation error is zero.

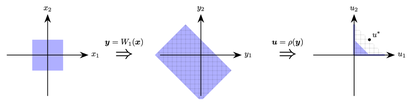
- Expressiveness of ReLU networks are not restricted.
- Special case: $\max(x_1, x_2)$ in $[0,1]^2$ fail to have this property with any single-neuron relaxation, but $\max(x_1, \ldots, x_d)$ in $[0,1]^d$ can be exactly bounded with polynomial runtime with a simple layerwise multi-neuron relaxation.

Statement (informal): For every certification problem, the worst-case number of subproblems required by multi-neuron relaxations is strictly smaller than single-neuron relaxations.

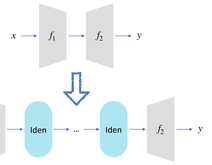
- This does not directly imply a runtime advantage.
- Special case: for the $\max$ function, multi-neuron relaxations reduce the number of subproblems exponentially, while the cost of solving each subproblem grows only polynomially, resulting in an exponential runtime advantage.

Convex Barrier Proof Sketches

Layerwise Relaxation: (i) Convex hull contains spurious points. (ii) Design the subsequent layers to have a single minimum at the spurious point. (iii) Every layerwise relaxation has no better lower bounds than the constructed minimum. (iv) The functional lower bounds are greater than the constructed minimum. (v) this gap can be made arbitrarily large by multiplying the weight of last layer.



Cross-layer Relaxation: (i) The convex hull of f_1 contains spurious points. (ii) Insert proper number of identity layers after f_1 to block information exchange enabled by cross-layer relaxation.



Expressiveness Proof Sketches

(i) Copy the input neurons by appending them to every layer and assign zero weight at the last layer. (ii) By definition of multi-neuron relaxation, there exists a relaxation that precisely captures the convex hull of the last layer, thus the convex hull of the overall network.