Neural Network Verification as Piecewise Linear Optimization: A Composition of Staircase Function Formulation

Tu Anh-Nguyen Joey Huchette
Rice University

Abstract

In this work, we provide a strong (ideal) MIP formulation for neural network verification task, which extends the work of Anderson et al. [1]:

- We derive a polynomial algorithm to obtain the facet-defining hyperplane for the Cayley embedding of the graph of activation functions.
- Empirically, our formulation are shown to give tighter LP relaxations for relaxed verifier and improves the performance of exact verifier.

Model

Given a function $g_i : \mathbb{R}^{n_i} \to \mathbb{R}$ as activation function of a neuron, and the graph of g_i is defined to be $\text{gr}(g_i(x)) \coloneqq \{(x, y) \in \mathbb{R}^{n_i+1} | y = g(x)\}$. Formally, the verification problem can be reformulated as

$$\begin{aligned}
&\max\ c_1 y_1 + \cdots + c_{n_2} y_{n_2} && \text{(1a)}\\
&\quad (x_1, \ldots, x_i) \in \text{gr}(g_i(x_1, \ldots, x_{i-1})) \quad \forall i \in \{n_1 + 1, \ldots, N\} && \text{(1b)}\\
&\quad (x_1, \ldots, x_{n_1}) \in X_\epsilon(x_0) && \text{(1c)}\\
&\quad y = (x_{N+1-n_2}, \ldots, x_N), && \text{(1d)}
\end{aligned}$$

where x_0 is the input, and $X_\epsilon(x_0)$ denote a neighborhood of x_0.

Staircase Functions

A univariate piecewise linear function $f : \mathbb{R} \to \mathbb{R}$ with k pieces is a **staircase** function if there exists $s \in \mathbb{R}$ such that every pieces' slope $a_i \in \{0, s\}$.

A piecewise linear function $f : \mathbb{R}^n \to \mathbb{R}$ is a k-piece **staircase** function if $f = g(w \cdot x)$ where g is a univariate staircase function.

Let $D^i \coloneqq \{x \in D |\ h_{i-1} \le w \cdot x + b \le h_i\}$, the *Cayley embedding* [2] for the closure of graph of f is:

$$S_{\text{Cayley}}(g) \coloneqq \bigcup_{i=1}^{k} \{(x, y, z) | x \in D^i,\ y = f(x),\ z = e^i\},$$

Theoretical Results

Lemma 1. *Given $(\hat{x}, \hat{y}, \hat{z})$, if the optimal value of the following problem is greater than $\hat{y}$ then $(\hat{x}, \hat{y}, \hat{z})$ is feasible, otherwise, the optimal solution α^* corresponds to a hyperplane that cut off $(\hat{x}, \hat{y}, \hat{z})$*

min $s\sum_{i=1}^{k} z_i(\sum_{j=1}^{n} u_j|w_j|\bar{\beta}_j^i - \sum_{j=1}^{n} l_j|w_j|\bar{\gamma}_j^i + (h_i - b)\bar{\theta}_1^i - (h_{i-1} - b)\bar{\theta}_2^i) + s\sum_{j=1}^{n} x_j|w_j|\bar{\alpha}_j$

$$\textit{subject to} \underbrace{\begin{bmatrix} A & 0 & \ldots & 0 & I_n \\ 0 & A & \ldots & 0 & I_n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & A & I_n \end{bmatrix}}_{\hat{A}} \begin{bmatrix} \bar{\beta}^1 \\ \bar{\gamma}^1 \\ \bar{\theta}^1 \\ \vdots \\ \bar{\beta}^k \\ \bar{\gamma}^k \\ \bar{\theta}^k \\ \bar{\alpha} \end{bmatrix} = \begin{bmatrix} \frac{a_1}{s}\bar{w} \\ \frac{\hat{a}_2}{s}\bar{w} \\ \frac{\hat{a}_3}{s}\bar{w} \\ \vdots \\ \frac{a_k}{s}\bar{w} \end{bmatrix}, \textit{ and } \begin{bmatrix} \bar{\beta}^1 \\ \bar{\gamma}^1 \\ \bar{\theta}^1 \\ \vdots \\ \bar{\beta}^k \\ \bar{\gamma}^k \\ \bar{\theta}^k \end{bmatrix} \ge \mathbf{0}. \qquad (2)$$

Theorem 1. *The separation procedure can be done in $O(n\log(n) + \max(k, n))$ time complexity*

Composition of Staircase Functions

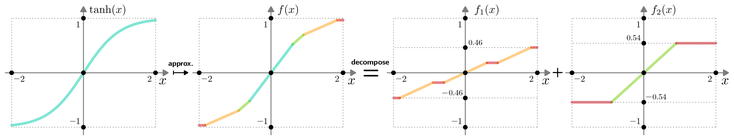

Fig. 1: An example

Lemma 2. *Let $g = g_1 + \cdots + g_m$ where $g_1, \ldots, g_k$ are staircase function, then conv$(C(g))$ is the solutions of the following system*

$$\begin{aligned}
y \le \min_{\bar{\alpha}_1 \in \mathbb{R}^n} (\alpha_1 \cdot x + \sum_{i=1}^{k} (\max_{x^i \in D^i} (a_i^1 w - \alpha_1) \cdot x^i + b_i) z_i) + \cdots + \\
\min_{\bar{\alpha}_m \in \mathbb{R}^n} (\alpha_m \cdot x + \sum_{i=1}^{k} (\max_{x^i \in D^i} (a_i^m w - \alpha_m) \cdot x^i + b_i) z_i) && \text{(3a)}\\
y \ge \max_{\underline{\alpha}_1 \in \mathbb{R}^n} (\underline{\alpha}_1 \cdot x + \sum_{i=1}^{k} (\min_{x^i \in D^i} (a_i^1 w - \underline{\alpha}_1) \cdot x^i + b_i) z_i) + \cdots + \\
\max_{\underline{\alpha}_m \in \mathbb{R}^n} (\underline{\alpha}_m \cdot x + \sum_{i=1}^{k} (\min_{x^i \in D^i} (a_i^m w - \underline{\alpha}_m) \cdot x^i + b_i) z_i) && \text{(3b)}\\
(x, y, z) \in D \times \mathbb{R} \times \Delta^k. && \text{(3c)}
\end{aligned}$$

Computational Experiments

We perform the verification task on neural network with binarized and quantized activation functions. In the following experiments, we train neural networks of different architecture on the MNIST dataset, and compare our separation method with state-of-the-art verification technique.

Relaxed Verifiers

All neural networks are training using the quantized network training open-source package Larq. The activation Dorefa κ is a constant piecewise function with 2^κ pieces.

Exact Verifiers

Scalability

References

[1] Ross Anderson et al. "Strong mixed-integer programming formulations for trained neural networks". In: *Mathematical Programming* (2020), pp. 1–37.

[2] Juan Pablo Vielma. "Embedding formulations and complexity for unions of polyhedra". In: *Management Science* 64.10 (2018), pp. 4721–4734.