Exercise 03

Box, MILP and DeepPoly Certification

Reliable and Trustworthy Artificial Intelligence
ETH Zurich

Problem 1 (Box Transformers)**.** Recall the box domain for numerical analysis. For vectors $a, b \in \mathbb{R}^m$ with $\forall i.\ a_i \leq b_i$, the box $[a, b]$ is a hypercube in $\mathbb{R}^m$. Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and an input box $[a, b] \subset \mathbb{R}^n$, a sound abstract transformer $f^\sharp$ finds $[c, d] \subset \mathbb{R}^m$ such that $\forall x \in [a, b].\ f(x) \in [c, d]$.

For example, let $x \in [1, 3]$ and $y \in [2, 4]$, and assume we want to approximate the result of $2x - y$. Using the basic abstract transformers from the lecture, we can compute

$$2 \cdot^\sharp [1, 3] -^\sharp [2, 4] \;=\; [2, 6] +^\sharp [-4, -2] \;=\; [-2, 4]$$

and conclude that $2x - y \in [-2, 4]$.

(a) Show that the box transformers lose precision, by approximating the outcome of $x - x$ for $x \in [0, 1]$ using the transformers $+^\sharp$ and $-^\sharp$ from the lecture.

(b) Prove or disprove: The alternative transformer $[a, b] +' [c, d] = [a + c,\ \ b + |d|]$ for addition is sound.

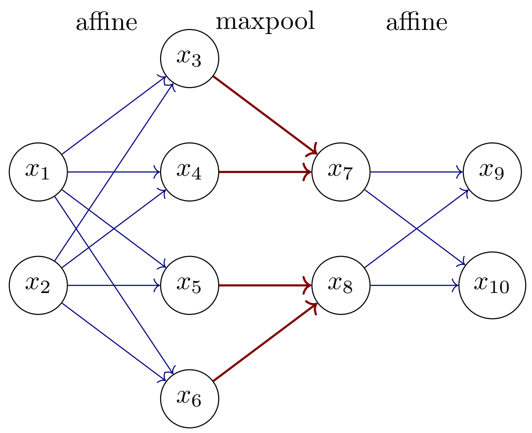
(c) Prove or disprove: The alternative transformer $[a, b] +'' [c, d] = [-\infty,\ \ a + b + d]$ for addition is sound.

(d) Derive a sound abstract transformer $f^\sharp$ for the function $f(x) := x^2$. That is, derive expressions for g, h such that $[g, h] = f^\sharp([a, b])$ for $a, b \in \mathbb{R}$.

(e) Derive a sound abstract transformer $\cdot^\sharp$ for multiplication. That is, derive expressions for g, h such that $[g, h] = [a, b] \cdot^\sharp [c, d]$, where $a, b, c, d \in \mathbb{R}$.

(f) The maxpool operation defined as $y := \max(x_1, x_2)$ computes the maximum of two input neurons $x_1, x_2 \in \mathbb{R}$. Derive a sound abstract transformer $\max^\sharp$ for maxpool. That is, derive expressions for y_1, y_2 such that $[y_1, y_2] = \max^\sharp([a_1, b_1], [a_2, b_2])$ for $a_1, b_1, a_2, b_2 \in \mathbb{R}$.

Problem 2 (Verification using Box)**.** Consider the neural network defined below, which takes inputs x_1, x_2 and produces outputs x_9, x_{10}. It consists of affine and maxpool layers.



$$
\begin{aligned}
x_3 &:= x_1 + x_2 & x_7 &:= \max(x_3, x_4) \\
x_4 &:= x_1 - 2 & x_8 &:= \max(x_5, x_6) \\
x_5 &:= x_1 - x_2 & x_9 &:= x_7 \\
x_6 &:= x_2 & x_{10} &:= -x_7 + x_8 - 0.5
\end{aligned}
$$

Assume we want to prove that for all values of $x_1, x_2 \in [0, 1]$, the output satisfies $x_9 > x_{10}$. Using your abstract transformers from Problem 1, try to prove the property by performing verification in the box domain. Does the proof succeed?

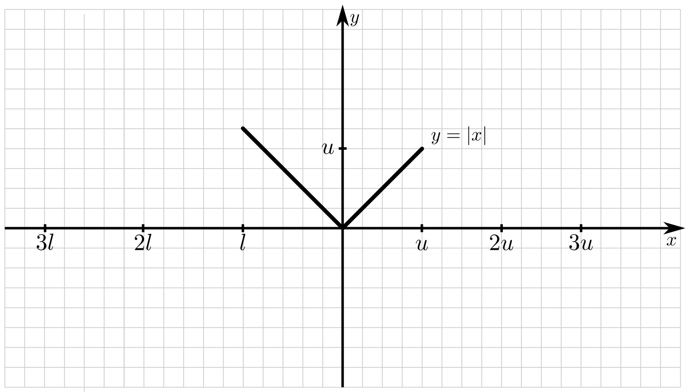
Problem 3 (MILP for Absolute Function—*from a previous exam*)**.** Consider the absolute function $y = |x|$, which computes the absolute value of a neuron $x \in \mathbb{R}$. Assume we know that x takes values in the range $l \leq x \leq u$ (e.g., computed using box verification).

(a) In the coordinate system below (where $l \leq 0 \leq u$), draw the two lines indicated by

$$\frac{y}{2} = -\frac{x}{2} + u \cdot a \quad \text{ for } a \in \{0, 1\}.$$

Indicate which points satisfy the following Mixed Integer Linear Program (MILP) constraints (here, ignore that $l \leq x \leq u$):

$$\frac{y}{2} \leq -\frac{x}{2} + u \cdot a, \qquad a \in \{0, 1\}.$$



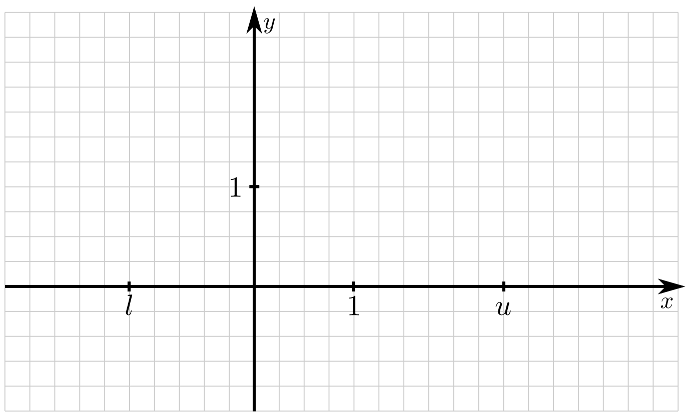
(b) Starting from the constraints above, find an exact MILP encoding of the absolute function. That is, provide a set of MILP constraints with solution $y = |x|$.

Problem 4 (MILP for ReLU1—*from a previous exam*). Consider the alternative activation function ReLU1: $\mathbb{R} \to \mathbb{R}$ defined as $\text{ReLU1}(x) := \min(1,\ \text{ReLU}(x))$. Assume we know that x takes values in the range $l \leq x \leq u$ (e.g., computed using box verification).

(a) Consider the following set of MILP constraints.

$$\begin{aligned} & y \leq a \\ & y \leq x - l \cdot (1 - a) \\ & y \geq 0 \\ & y \geq x \\ & a \in \{0, 1\} \end{aligned}$$

Draw the line segments representing the solution of this constraint set in the coordinate system below.



(b) To arrive at an exact MILP encoding of ReLU1, we adapt the constraint system from subtask 1). In particular, we extend it by an additional integer variable b and replace two inequalities as indicated below.

$$
\begin{aligned}
&y \le a \\
&y \le x - l \cdot (1 - a) \\
&\boxed{\quad \text{(i)} \quad} \\
&\boxed{\quad \text{(ii)} \quad} \\
&a, b \in \{0, 1\}
\end{aligned}
$$

Provide the two missing inequalities (i) and (ii) such that the resulting MILP constraint system has the solution $y = \text{ReLU1}(x)$ with $l \le x \le u$, as shown above. The inequalities must be linear in the variables x, y, a, b. You are *not* allowed to introduce any further constraints or variables.

Hint: Leverage the variable b to extend your solution from subtask (a). For $b = 0$, the constraint system should reduce to the system in (a).

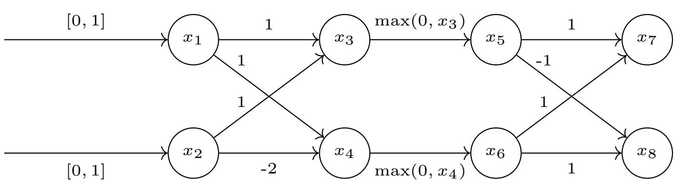
Problem 5 (DeepPoly). Recall that DeepPoly decides between two options for relaxing the result of $y = \text{ReLU}(x)$ based on the area, shown in Fig. 1.

(a) Option 1 (b) Option 2

Figure 1: Options for triangle relaxations in DeepPoly.

(a) Derive a decision procedure depending on l and u which decides when Option 1 results in a smaller area. Break ties in favor of Option 1.

(b) Consider the fully connected neural network shown below. The network has two input neurons (x_1, x_2) and two output neurons (x_7, x_8).



Analyze this network using DeepPoly with respect to the input region spanned by $x_1 \in [0, 1]$ and $x_2 \in [0, 1]$. Use the smaller area transformer as derived in subtask (a). Then, use the result to show that $x_7 \geq x_8$.