Exercise 05 - Solution

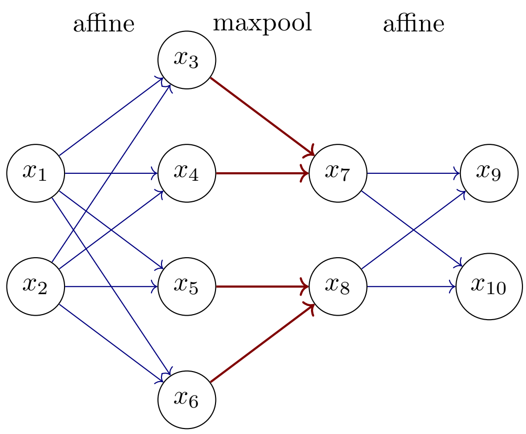
Certification with Box and MILP

Reliable and Interpretable Artificial Intelligence
ETH Zurich

Problem 1 (Box Verification for Maxpool)**.** Consider the maxpool operation defined as $y := \max(x_1, x_2)$, which computes the maximum of two input neurons $x_1, x_2 \in \mathbb{R}$. This operation is typically used in neural networks to reduce dimensionality. In this task, you are going to extend box verification to the maxpool operation.

1. Derive a sound abstract transformer $\max^\sharp$ for the maxpool operation in the box domain. 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 $a_1, b_1, a_2, b_2 \in \mathbb{R}$. Your transformer should be as precise as possible.

2. Consider the neural network defined below. The network takes inputs x_1, x_2 and produces outputs x_9, x_{10}. It consists of both affine and maxpool layers.



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

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 transformer from above, try to prove the property by performing verification in the box domain. Does the proof succeed?

Solution 1.

1. The most precise sound transformer is:

$$[y_1, y_2] = \max{}^\sharp([a_1, b_1], [a_2, b_2]) = [\max(a_1, a_2),\ \max(b_1, b_2)].$$

2. The intervals for the different neurons in the network are:

$$\begin{aligned} x_1 &\in [0,1] & x_6 &\in [0,1] \\ x_2 &\in [0,1] & x_7 &\in [0,2] \\ x_3 &\in [0,2] & x_8 &\in [0,1] \\ x_4 &\in [-2,-1] & x_9 &\in [0,2] \\ x_5 &\in [-1,1] & x_{10} &\in [-2.5, 0.5] \end{aligned}$$

From this, we cannot conclude that $x_9 > x_{10}$. In particular, the lower bound for $x_9 - x_{10}$ is -0.5, which is not sufficient to prove the property.

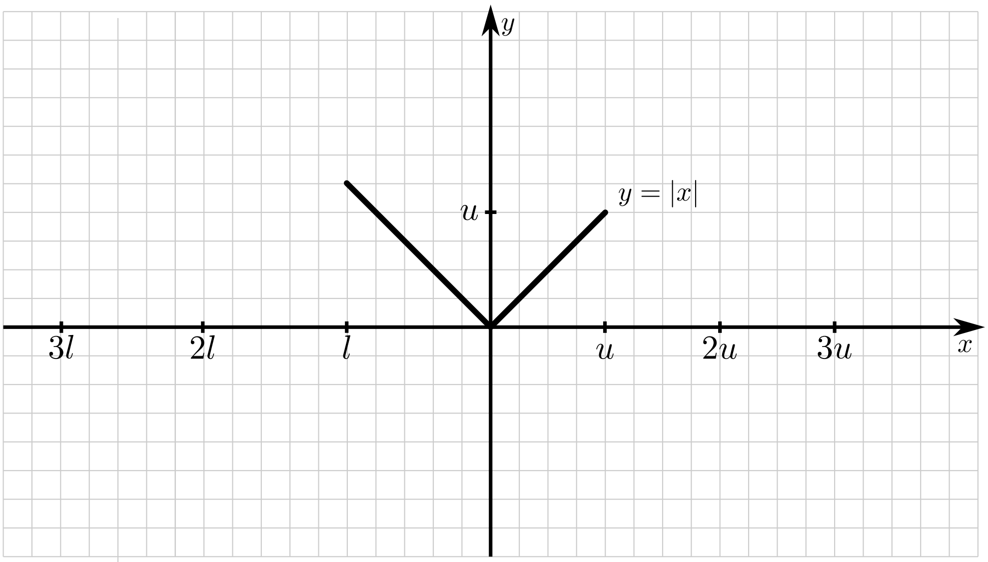
Problem 2 (MILP for Absolute Function—*from a previous exam*)**.** Consider the absolute function $y = |x|$, which computes the absolute value of a neuron $x \in \mathbb{R}$. Assume we know that x takes values in the range $l \le x \le u$ (e.g., computed using box verification).

1. In the coordinate system below (where $l \le 0 \le u$), 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 $l \le x \le u$):

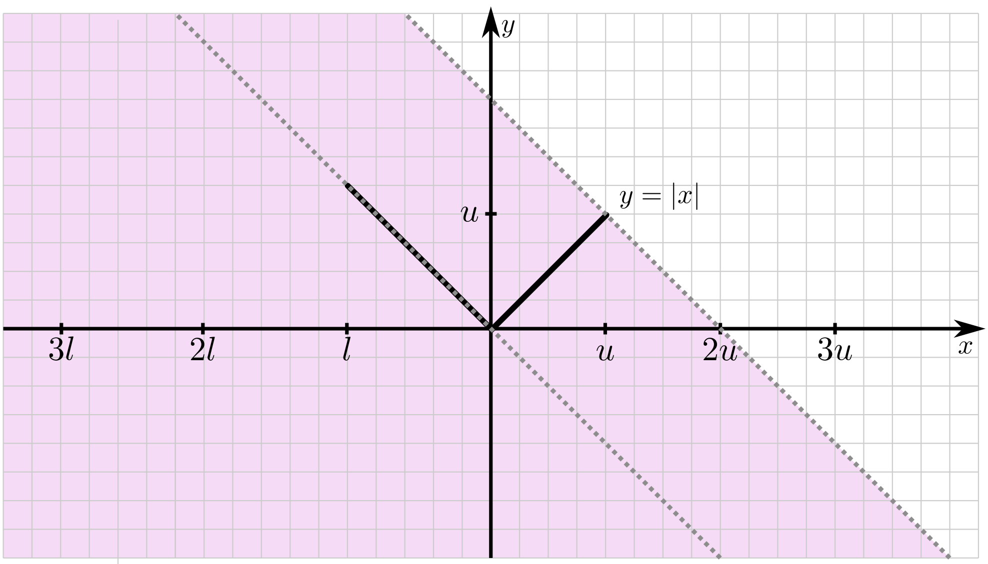
$$\frac{y}{2} \le -\frac{x}{2} + u \cdot a, \qquad a \in \{0, 1\}.$$



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

Solution 2.

1. See the following figure.



2. We can use an analogous construction as in the previous question to create a line which (i) for $a = 1$ coincides with the line segment at $x \geq 0$, and (ii) for $a = 0$ matches the lower end of the line segment at $x = l$. We construct the following inequality constraints, which bound the values of y from above:

$$\frac{y}{2} \leq \frac{x}{2} - l \cdot (1 - a), \qquad a \in \{0, 1\}$$

As a last step, we also need to bound y from below according to the "v-shape" of the absolute function. This can easily be achieved using the constraints

$$y \geq x \quad \text{and} \quad y \geq -x.$$

Combining these items leads to the following set of MILP constraints, which exactly represents the bold line segments in the figure.

$$\begin{aligned}
&\frac{y}{2} \leq \frac{x}{2} - l \cdot (1 - a), \\
&y \geq x, \\
&\frac{y}{2} \leq -\frac{x}{2} + u \cdot a, \\
&y \geq -x, \\
&a \in \{0, 1\}.
\end{aligned}$$