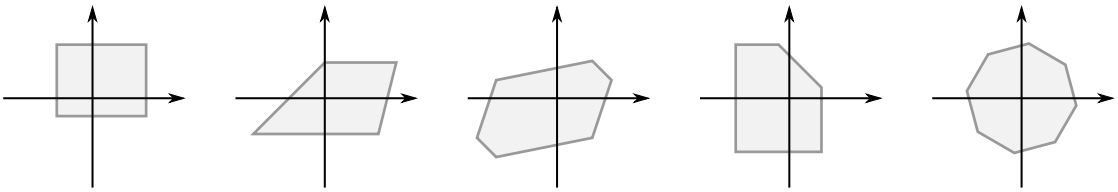
Exercise 06 - Solution

Zonotopes and Abstract Interpretation

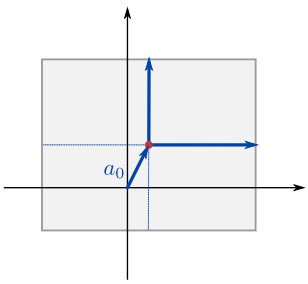
Reliable and Interpretable Artificial Intelligence
ETH Zurich

Problem 1 (Zonotope Concretizations)**.** Which of the following 2D regions (a–e) represent concretizations of a zonotope? For all such regions, sketch a set of 2D magnitude vectors $a_0, \ldots, a_k$ describing the zonotope (select k as small as possible).



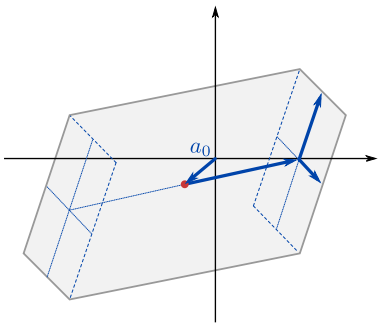
(a) (b) (c) (d) (e)

Solution 1. The regions (b) and (d) are not point symmetric and hence can not be concretizations of zonotopes. The regions (a), (c) and (e) are concretizations of zonotopes with the following magnitude vectors:



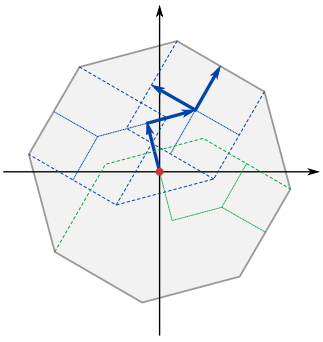
(a)

Needs three magnitudes ($k = 2$), a_0 points to the center



(c)

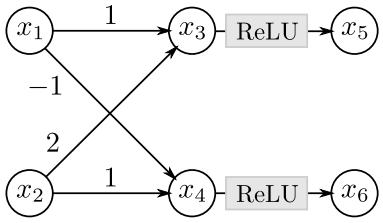
Needs four magnitudes ($k = 3$), a_0 points to the center



(e)

Needs five magnitudes ($k = 4$). The zonotope is centered and we have $a_0 = \mathbf{0}$ (not shown).

Problem 2 (Certification using Zonotopes)**.** Consider the following small neural network with two input neurons x_1, x_2 and two output neurons x_5, x_6. The network consists of an affine layer followed by a ReLU layer.



$$x_5 = \mathrm{ReLU}(x_1 + 2x_2)$$
$$x_6 = \mathrm{ReLU}(x_2 - x_1)$$

You are given the following zonotope ψ over the input neurons:

$$\psi: \quad \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \cdot \varepsilon_1 + \begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \varepsilon_2 + \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

Your goal is to prove that $x_5 \geq x_6$ for all inputs x_1, x_2 in the zonotope ψ.

1. Draw the concretization $\gamma(\psi)$ of ψ. What shape does it have?

2. Using the transformers for affine and ReLU layers discussed in the lecture, transform ψ to a zonotope ϕ over the output neurons of the network above.

3. Draw the concretization $\gamma(\phi)$ of ϕ. Can you use ϕ to prove the desired property?

Solution 2.

1. The 4 corners of the zonotope's concretization can be obtained by setting ε_1 and ε_2 to the extreme values $\{-1, 1\}$. The corners are

$$\begin{pmatrix} 7 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

   and the concretization is the parallelogram with these corners.

   Note: For $k > 2$, finding corners is a bit more involved as not all such extreme points are corners (see for example region (c) in the solution of problem 1).

2. We can write $\hat{x}_1 = 2\varepsilon_1 + \varepsilon_2 + 4$ and $\hat{x}_2 = \varepsilon_1 + 2\varepsilon_2 + 3$.

   Step 1: Affine layer. It is

$$\begin{aligned} \hat{x}_3 &= (2 + 2 \cdot 1) \cdot \varepsilon_1 + (1 + 2 \cdot 2) \cdot \varepsilon_2 + (4 + 2 \cdot 3) &&= 4\varepsilon_1 + 5\varepsilon_2 + 10 \\ \hat{x}_4 &= ((-1) \cdot 2 + 1) \cdot \varepsilon_1 + ((-1) \cdot 1 + 2) \cdot \varepsilon_2 + ((-1) \cdot 4 + 3) &&= -\varepsilon_1 + \varepsilon_2 - 1 \end{aligned}$$

Step 2: ReLU layer. First, we compute the bounds for $\hat{x}_3$. The lower bound $l_{\hat{x}_3}$ for $\hat{x}_3$ can be obtained by setting $\varepsilon_1 = -1$ and $\varepsilon_2 = -1$. The upper bound $u_{\hat{x}_3}$ can be obtained by setting $\varepsilon_1 = 1$ and $\varepsilon_2 = 1$. This gives the bounds $[l_{\hat{x}_3}, u_{\hat{x}_3}] = [1, 19]$. Because the bounds are above 0, the ReLU has no effect and it is:

$$\hat{x}_5 = \hat{x}_3 = 4\varepsilon_1 + 5\varepsilon_2 + 10 \tag{1}$$

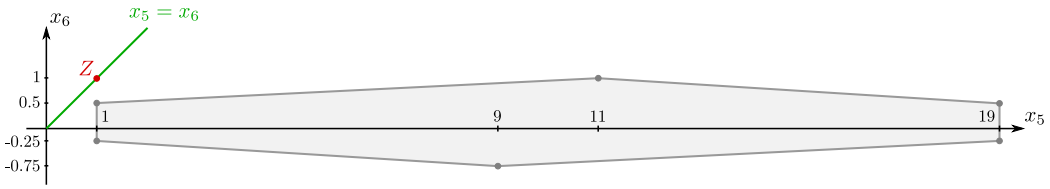
Next, we compute the bounds for $\hat{x}_4$. The lower bound is obtained for $\varepsilon_1 = 1$ and $\varepsilon_2 = -1$, the upper bound for $\varepsilon_1 = -1$ and $\varepsilon_2 = 1$. It is $[l_{\hat{x}_4}, u_{\hat{x}_4}] = [-3, 1]$, hence we are in the "crossing boundary" case. We compute the slope λ as

$$\lambda = \frac{u_{\hat{x}_4}}{u_{\hat{x}_4} - l_{\hat{x}_4}} = \tfrac{1}{4}$$

and obtain the following expression for $\hat{x}_6$, where ε_3 is a new noise term:

$$\begin{aligned} \hat{x}_6 &= \lambda \cdot \hat{x}_4 - \varepsilon_3 \cdot \frac{\lambda \cdot l_{\hat{x}_4}}{2} - \frac{\lambda \cdot l_{\hat{x}_4}}{2} = \tfrac{1}{4}\hat{x}_4 - \varepsilon_3 \cdot \left(\tfrac{1}{4} \cdot \tfrac{-3}{2}\right) - \left(\tfrac{1}{4} \cdot \tfrac{-3}{2}\right) \\ &= \tfrac{1}{8} - \tfrac{1}{4}\varepsilon_1 + \tfrac{1}{4}\varepsilon_2 + \tfrac{3}{8}\varepsilon_3 \end{aligned} \tag{2}$$

The zonotope ϕ is given by (1) and (2). Its concretization is:



3. By (1), the minimum value (lower bound) for x_5 is $l_{\hat{x}_5} = -4 - 5 + 10 = 1$. Independently, by (2), the maximum value (upper bound) for x_6 is $u_{\hat{x}_6} = \frac{1}{8} + \frac{1}{4} + \frac{1}{4} + \frac{3}{8} = 1$. Hence, it is guaranteed that $x_5 \geq l_{\hat{x}_5} = 1 = u_{\hat{x}_6} \geq x_6$, which proves the property.

   Note: Here, we have analyzed the lower and upper bounds for x_5 and x_6 independently. This is an overapproximation because x_5 and x_6 are not independent due to shared noise terms (ε_1, ε_2). In particular, $l_{\hat{x}_5}$ and $u_{\hat{x}_6}$ can not be attained at the same time (see how the point Z is outside the zonotope in the figure above). While our argument here is sound, it may be too imprecise in certain cases. For example, with this argument we can not prove the *strict* inequality $x_5 > x_6$, even though it holds: See how the zonotope lies strictly below the line $x_5 = x_6$ in the figure above. Thus, proving properties given an output zonotope may require further analysis (e.g., inspecting the zonotope's shape).

Problem 3 (hybrid Zonotopes)**.** In this problem, we consider a new convex relaxation, fusing the zonotope with the interval relaxation. Specifically, we generalize the standard zonotope $x = c + \sum_i a_i \epsilon_i$ for $c \in \mathbb{R}$ and $a_i \in \mathbb{R}$ for all i to $x = [c_l, c_u] + \sum_i a_i \epsilon_i$, where we replaced the center c with an interval $[c_l, c_u]$. The design goal for this exercise is to not increase the number of error-terms.

1. Derive a formula for the addition of two hybrid Zonotopes.
2. Derive a formula for the unary "-" operation applied to a hybrid Zonotope.
3. Derive a formula for the subtracting one hybrid Zonotope from another one.
4. Derive a formula for concretizing a hybrid Zonotope to an interval.
5. Derive a formula for multiplying an interval with a hybrid Zonotope.

Bonus: Derive a formula for the multiplication of two hybrid Zonotopes.

Solution 3. Let $x_1 := [c_l^1, c_u^1] + \sum_i a_i^1 \epsilon_i$, $x_2 := [c_l^2, c_u^2] + \sum_i a_i^2 \epsilon_i$ and $x = [c_l, c_u] + \sum_i a_i \epsilon_i$.

1. Addition:
$$x_1 + x_2 = [c_l^1, c_u^1] + \sum_i a_i^1 \epsilon_i + [c_l^2, c_u^2] + \sum_i a_i^2 \epsilon_i = [c_l^1 + c_l^2, c_u^1 + c_u^2] + \sum_i (a_i^1 + a_i^2) \epsilon_i$$

2. Unary "-":
$$-x = -[c_l, c_u] - \sum_i a_i \epsilon_i = [-c_u, -c_l] + \sum_i (-a_i) \epsilon_i$$

3. Subtraction:
$$x_1 - x_2 = [c_l^1, c_u^1] + \sum_i a_i^1 \epsilon_i - [c_l^2, c_u^2] - \sum_i a_i^2 \epsilon_i = [c_l^1 - c_u^2, c_u^1 - c_l^2] + \sum_i (a_i^1 - a_i^2) \epsilon_i$$

4. Interval-concretization:
$$\begin{aligned} \text{Interval}(x) &= \text{Interval}\left([c_l, c_u] + \sum_i a_i \epsilon_i \right) \\ &= \left[\min([c_l, c_u] + \sum_i a_i \epsilon_i), \max([c_l, c_u] + \sum_i a_i \epsilon_i) \right] \\ &= \left[c_l - \sum_i \|a_i\|, c_u + \sum_i \|a_i\| \right] \end{aligned}$$

5. Interval - hybrid Zonotope multiplication:

$$
\begin{aligned}
[v,w]\cdot x &= [v,w]\cdot\left([c_l,c_u]+\sum_i a_i\epsilon_i\right)\\
&= [v,w]\cdot[c_l,c_u]+[v,w]\cdot\sum_i a_i\epsilon_i\\
&= [v,w]\cdot[c_l,c_u]+(\tfrac{v+w}{2}+\tfrac{w-v}{2}\epsilon')\cdot\sum_i a_i\epsilon_i\\
&= [v,w]\cdot[c_l,c_u]+\tfrac{v+w}{2}\cdot\sum_i a_i\epsilon_i+\tfrac{w-v}{2}\epsilon'\cdot\sum_i a_i\epsilon_i\\
&= [v,w]\cdot[c_l,c_u]+\sum_i \tfrac{v+w}{2}a_i\epsilon_i+\sum_i \tfrac{w-v}{2}a_i\epsilon'\cdot\epsilon_i\\
&\to [v,w]\cdot[c_l,c_u]+\sum_i \tfrac{v+w}{2}a_i\epsilon_i+\sum_i \tfrac{w-v}{2}a_i\epsilon_i''\\
&\to [v,w]\cdot[c_l,c_u]+\sum_i \tfrac{v+w}{2}a_i\epsilon_i+[-1,1]\cdot\sum_i \|\tfrac{w-v}{2}a_i\|\\
&= [\min(vc_l,vc_u,wc_l,wc_u),\max(vc_l,vc_u,wc_l,wc_u)]\\
&\quad+\sum_i \tfrac{v+w}{2}a_i\epsilon_i+[-1,1]\cdot\sum_i \|\tfrac{w-v}{2}a_i\|\\
&= \left[\min(vc_l,vc_u,wc_l,wc_u)-\sum_i \|\tfrac{w-v}{2}a_i\|,\max(vc_l,vc_u,wc_l,wc_u)+\sum_i \|\tfrac{w-v}{2}a_i\|\right]\\
&\quad+\sum_i \tfrac{v+w}{2}a_i\epsilon_i
\end{aligned}
$$

Bonus:

$$
\begin{aligned}
x_1\cdot x_2 &= ([c_l^1,c_u^1]+\sum_i a_i^1\epsilon_i)\cdot([c_l^2,c_u^2]+\sum_j a_j^2\epsilon_j)\\
&= [c_l^1,c_u^1]\cdot[c_l^2,c_u^2]+[c_l^1,c_u^1]\cdot\sum_j a_j^2\epsilon_j+[c_l^2,c_u^2]\cdot\sum_i a_i^1\epsilon_i+\sum_i a_i^1\epsilon_i\cdot\sum_j a_j^2\epsilon_j
\end{aligned}
$$

The first term is standard interval multiplication, the second and third term are interval - hybrid Zonotope multiplications, so we just show how to proceed with the last term:

$$
\begin{aligned}
\sum_i a_i^1\epsilon_i\cdot\sum_j a_j^2\epsilon_j &= \sum_{i,j} a_i^1a_j^2\epsilon_i\epsilon_j=\sum_{i,j} a_i^1a_j^2\epsilon'_{i,j}\\
&\to\left[-\sum_{i,j}\|a_i^1a_j^2\|,\sum_{i,j}\|a_i^1a_j^2\|\right]
\end{aligned}
$$