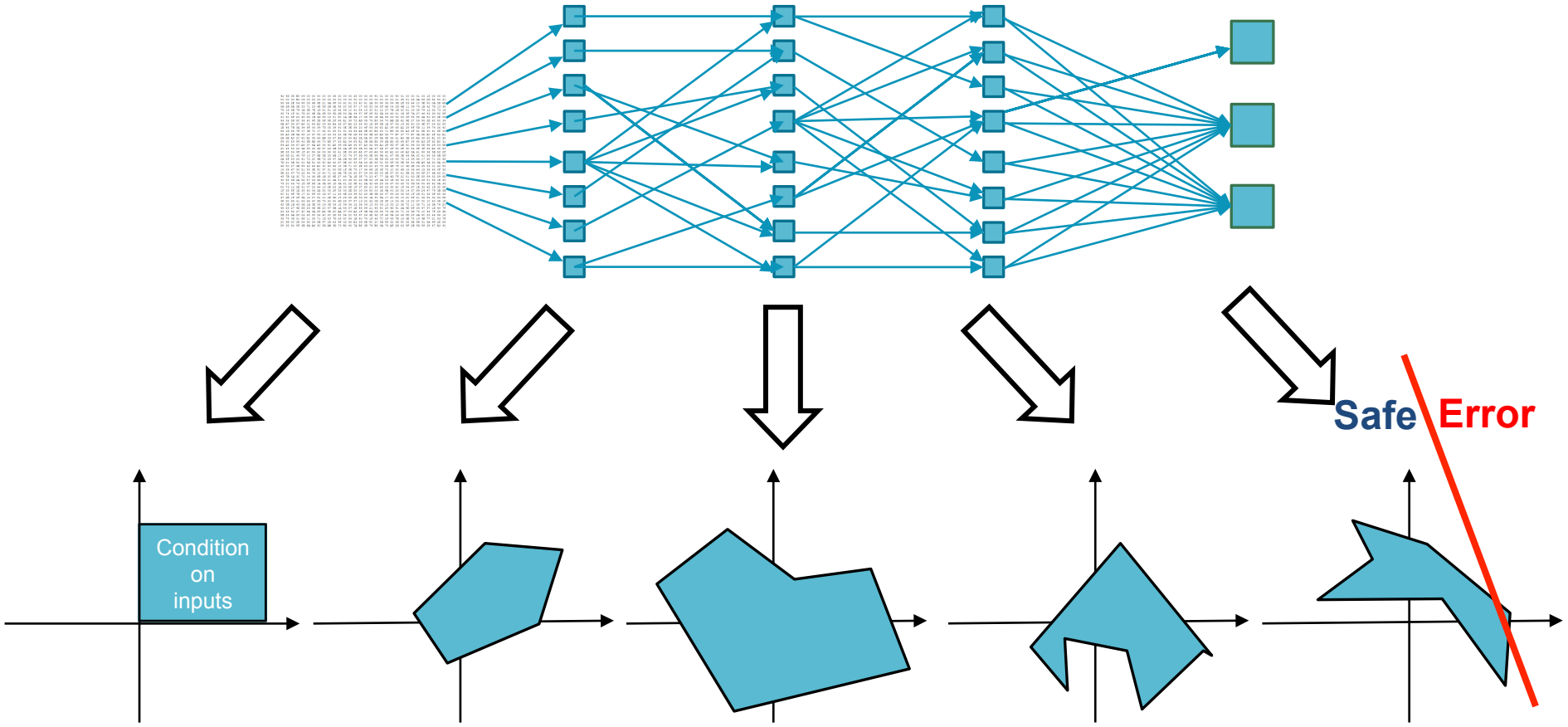
Neural Network Verification

Part 4: *Incomplete Methods*

Some true properties can be proved true

Robust Deep Learning

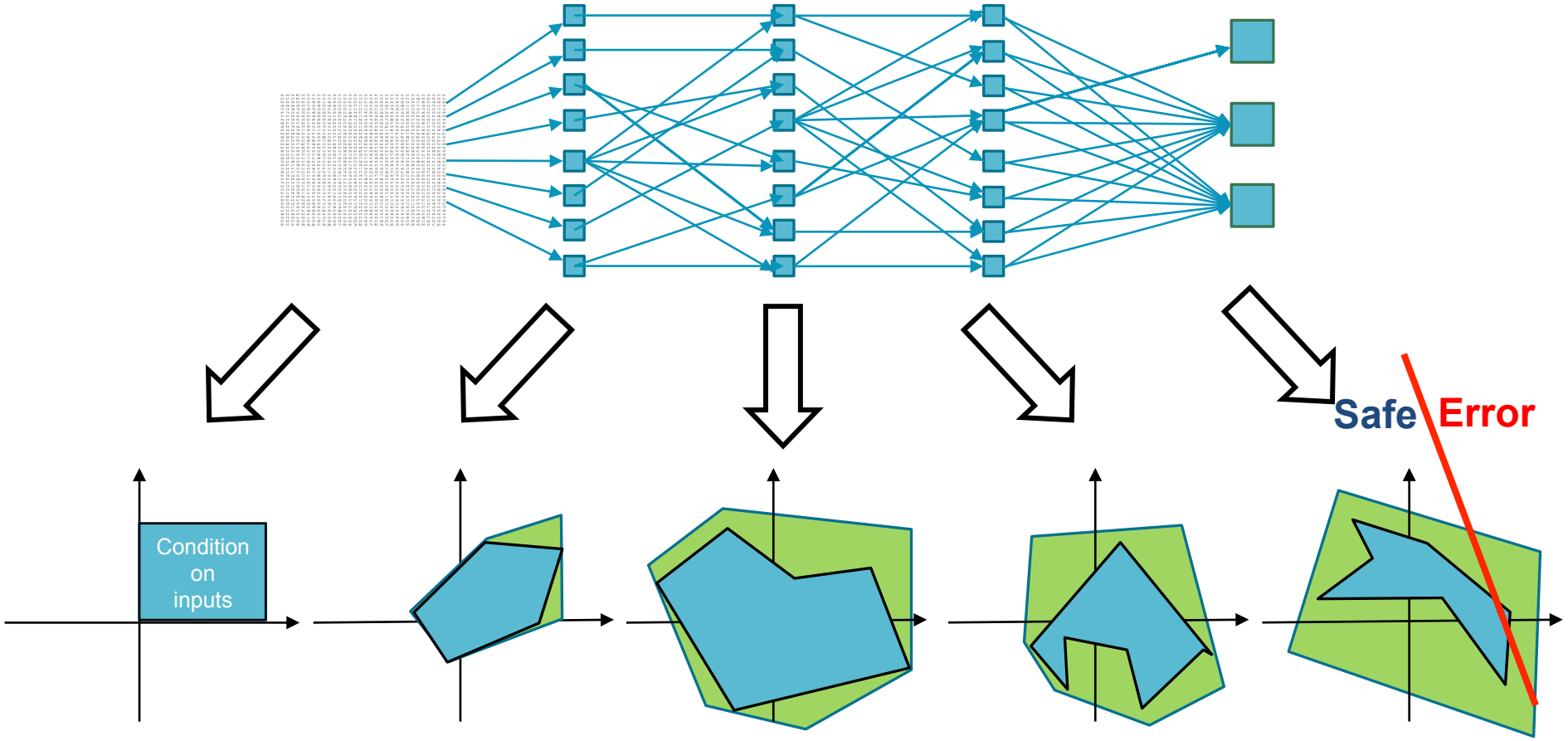
Is there an erroneous output?



Non-convexity makes the problem NP-hard

Robust Deep Learning

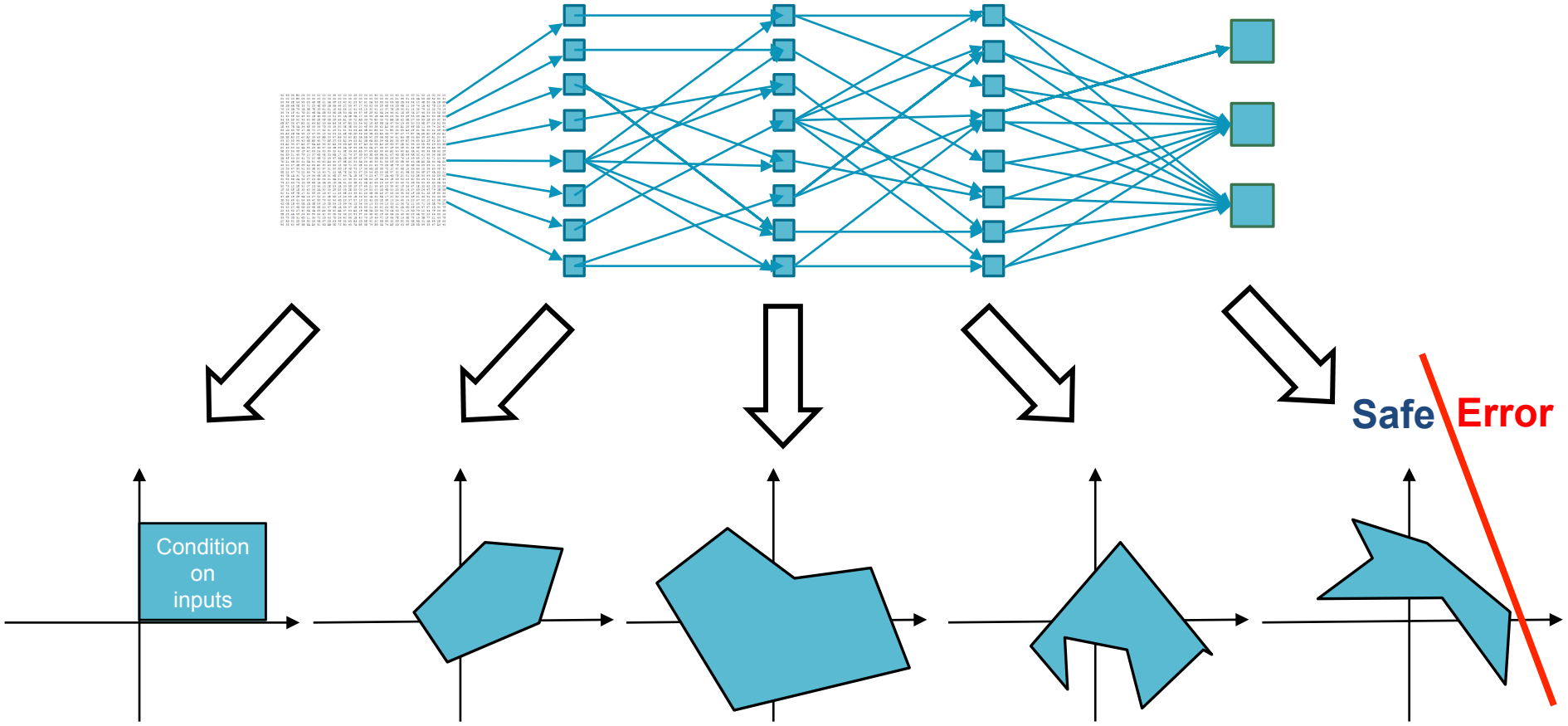
Is there an erroneous output?



Replace by a convex superset

Robust Deep Learning

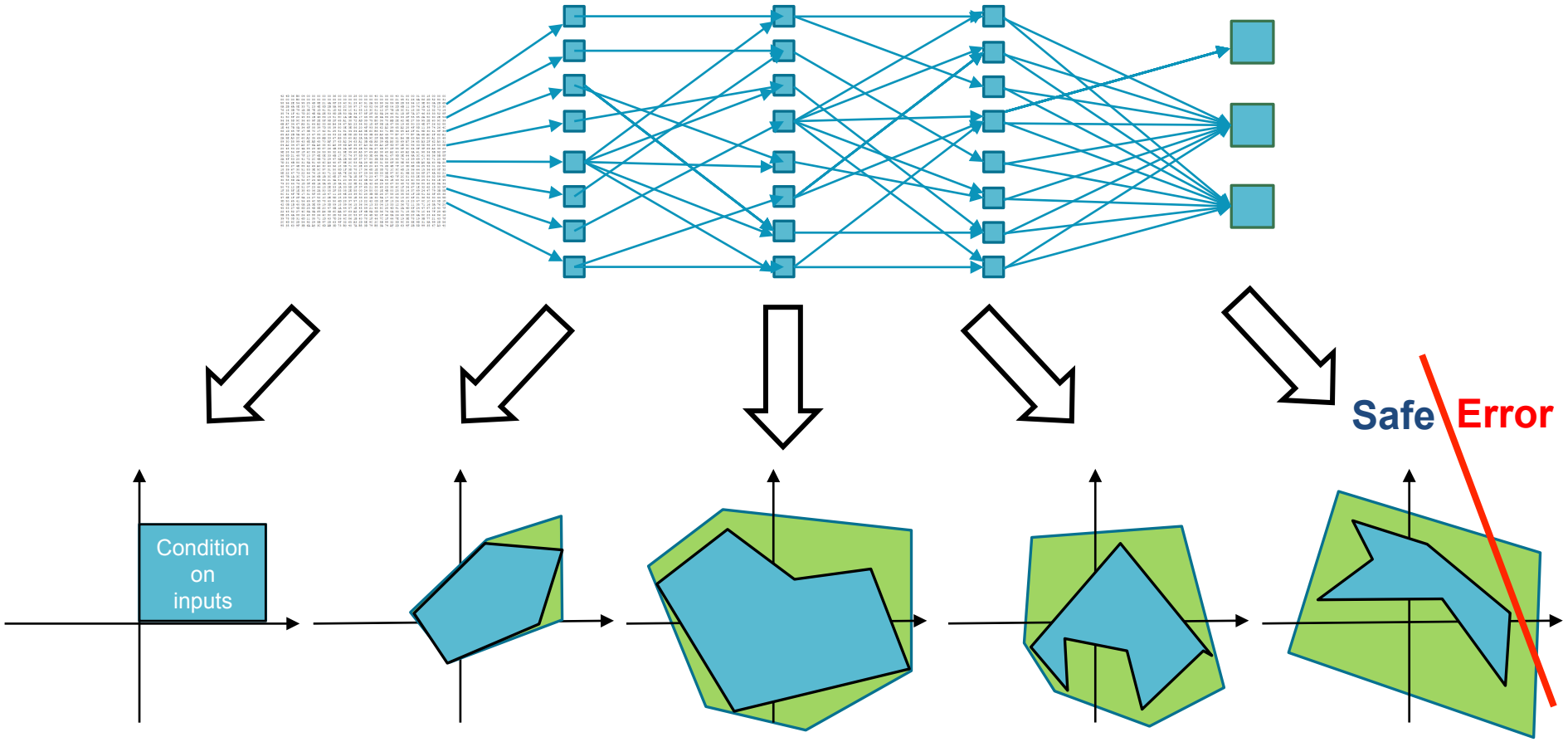
Is there an erroneous output?



Say, non-convex set has no erroneous output

Robust Deep Learning

Is there an erroneous output?



Convex superset might give incorrect answer

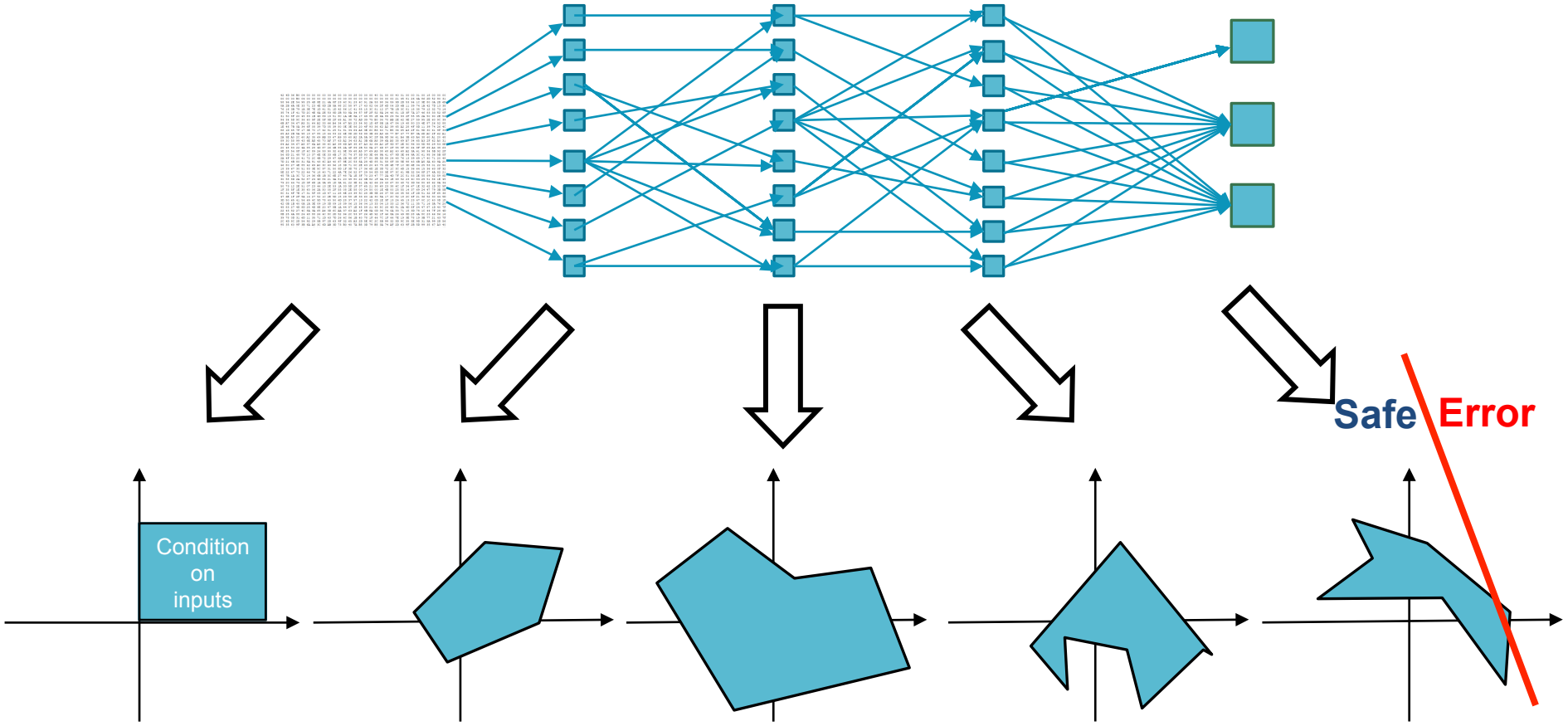
Outline

- Interval Bound Propagation
- Linear Programming Relaxation
- Results

Mirman et al., 2018; Gowal et al., 2018

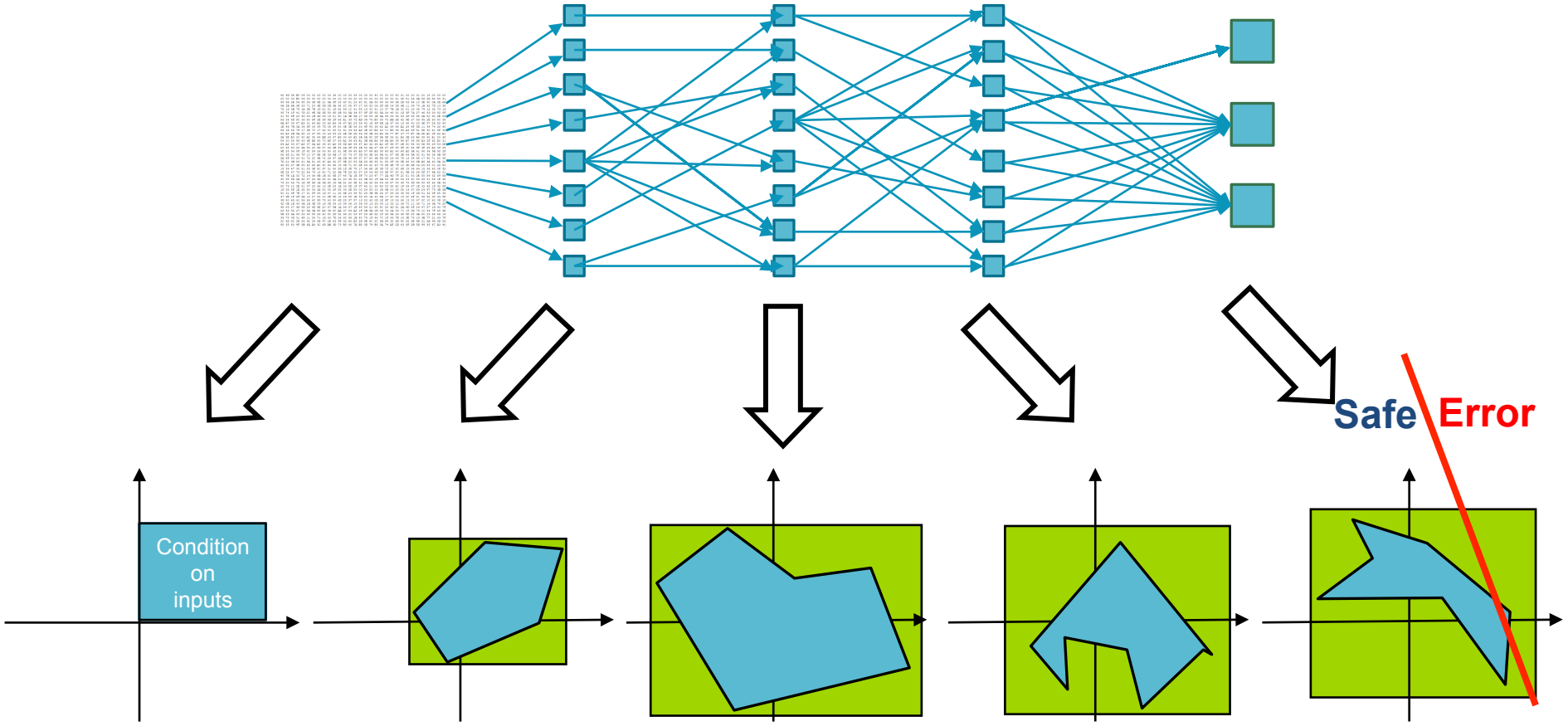
Robust Deep Learning

Is there an erroneous output?



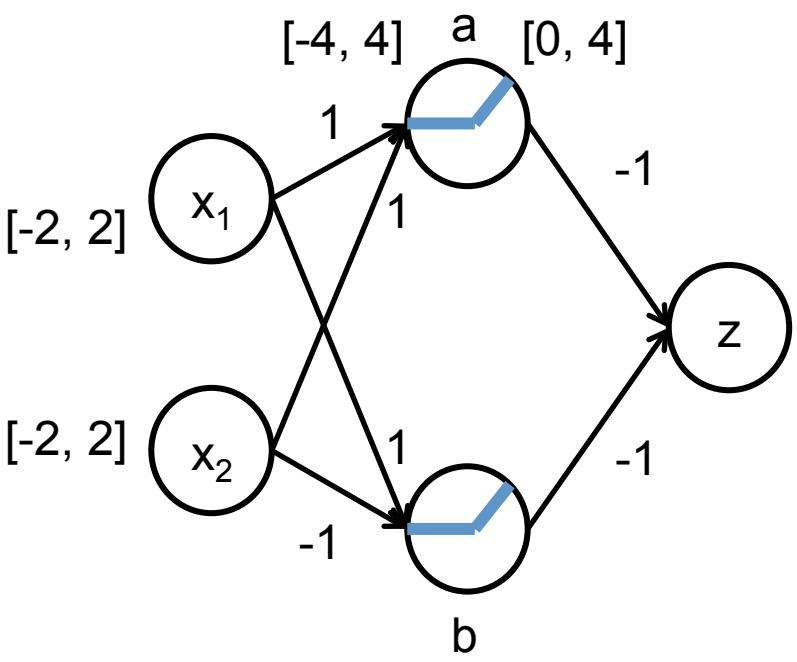
Robust Deep Learning

Is there an erroneous output?



Axis aligned convex superset

Example



$-2 \le x_1 \le 2$

$-2 \le x_2 \le 2$

$a_{in} = x_1 + x_2$

$a_{out} = \max\{a_{in}, 0\}$

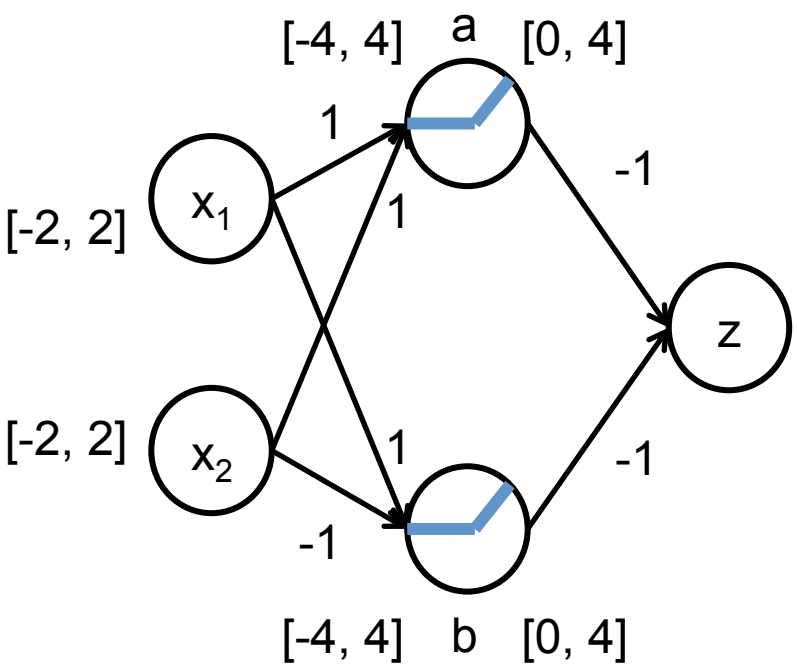
Minimum value of a_{in}? -4

Minimum value of a_{out}? 0

Maximum value of a_{in}? 4

Maximum value of a_{out}? 4

Example



$-2 \le x_1 \le 2$

$-2 \le x_2 \le 2$

$b_{in} = x_1 - x_2$

$b_{out} = \max\{b_{in}, 0\}$

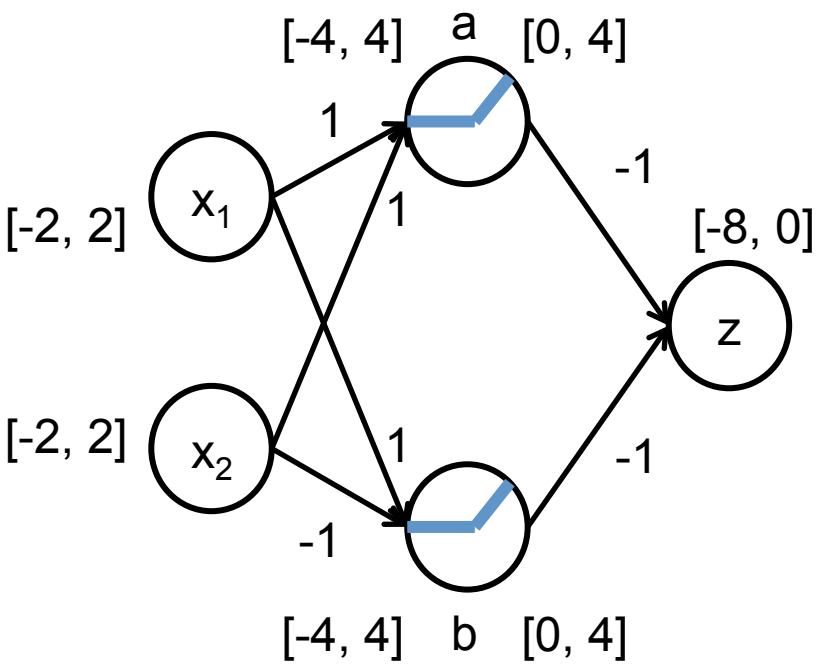
Minimum value of b_{in}? -4

Minimum value of b_{out}? 0

Maximum value of b_{in}? 4

Maximum value of b_{out}? 4

Example



$-2 \le x_1 \le 2$

$-2 \le x_2 \le 2$

$b_{in} = x_1 - x_2$

$b_{out} = \max\{b_{in}, 0\}$

$z = -a_{out} - b_{out}$

Minimum value of z? -8

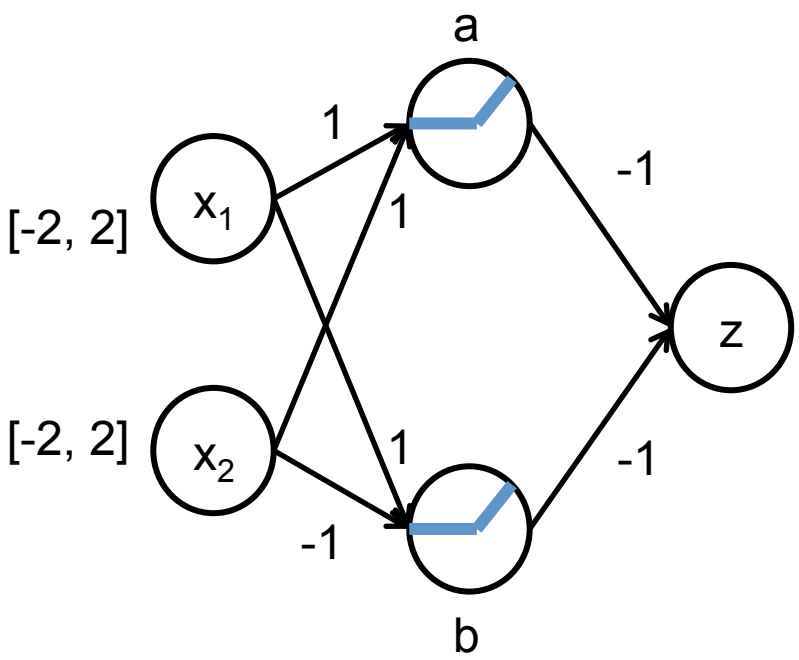
Maximum value of z? 0

Outline

- Interval Bound Propagation
- **Linear Programming Relaxation**
- Results

Wong and Kolter, 2018

Example



min z

s.t. $-2 \le x_1 \le 2$

$-2 \le x_2 \le 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

$a_{out} = \max\{a_{in}, 0\}$

$b_{out} = \max\{b_{in}, 0\}$

$z = - a_{out} - b_{out}$

Example

Linear constraints

Easy to handle

$$\min \quad z$$

$$\text{s.t.} \quad \mathbf{-2 \le x_1 \le 2}$$

$$\mathbf{-2 \le x_2 \le 2}$$

$$\mathbf{a_{in} = x_1 + x_2}$$

$$\mathbf{b_{in} = x_1 - x_2}$$

$$a_{out} = \max\{a_{in}, 0\}$$

$$b_{out} = \max\{b_{in}, 0\}$$

$$\mathbf{z = - a_{out} - b_{out}}$$

Example

min z

s.t. $-2 \le x_1 \le 2$

$-2 \le x_2 \le 2$

$a_{in} = x_1 + x_2$

$b_{in} = x_1 - x_2$

$a_{out} = \max\{a_{in}, 0\}$

$b_{out} = \max\{b_{in}, 0\}$

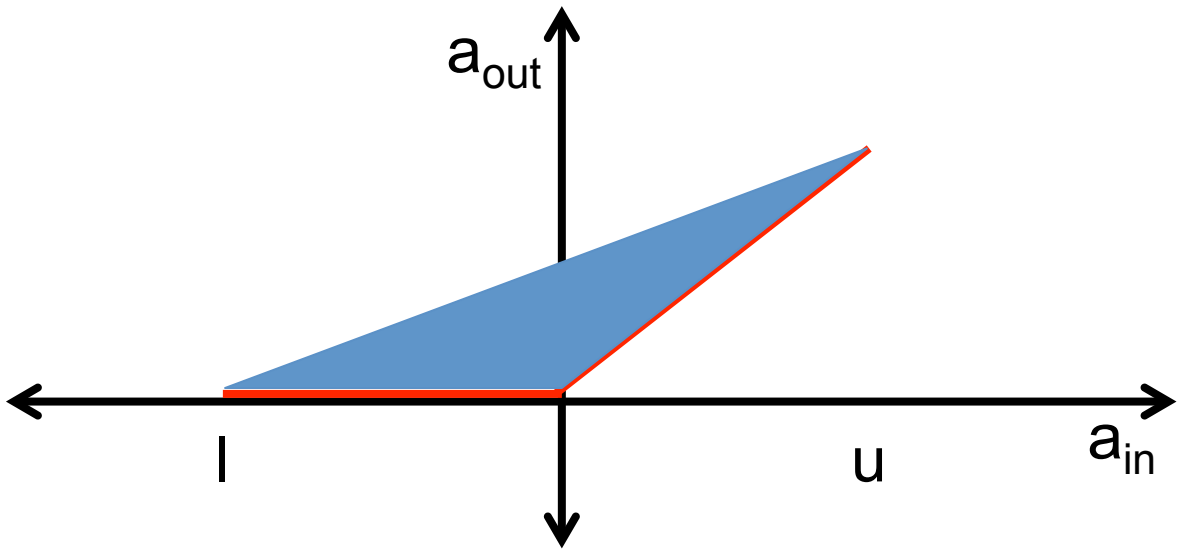
$z = - a_{out} - b_{out}$

Non-linear constraints

NP-hard problem

Relaxation

$a_{out} = \max\{a_{in}, 0\}$ $\quad\quad$ $a_{in} \in [l,u]$



Ehlers 2017

Replace with convex superset

Example

$$\min \quad z$$

$$\text{s.t.} \quad -2 \le x_1 \le 2$$

$$-2 \le x_2 \le 2$$

$$a_{in} = x_1 + x_2$$

$$b_{in} = x_1 - x_2$$

$$\mathbf{a_{out} = \max\{a_{in}, 0\}}$$

$$\mathbf{b_{out} = \max\{b_{in}, 0\}}$$

$$z = - a_{out} - b_{out}$$

Example

Linear Program

Several **"efficient"** solvers

$$\min \quad z$$

$$\text{s.t.} \quad -2 \le x_1 \le 2$$

$$-2 \le x_2 \le 2$$

$$a_{in} = x_1 + x_2$$

$$b_{in} = x_1 - x_2$$

$$a_{out} \ge 0,\ a_{out} \ge a_{in},\ a_{out} \le 0.5a_{in} + 2$$

$$b_{out} \ge 0,\ b_{out} \ge b_{in},\ b_{out} \le 0.5b_{in} + 2$$

$$z = - a_{out} - b_{out}$$

Outline

- Interval Bound Propagation
- Linear Programming Relaxation
  - **LP Duality**
- Results

Example

$$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$$

2 x 7 x

s.t. $-x_1 \le 0$, $-x_2 \le 0$, $-x_3 \le 0$

3 x $x_1 + x_2 + 3x_3 \le 30$

$2x_1 + 2x_2 + 5x_3 \le 24$

$4x_1 + x_2 + 2x_3 \le 36$

Scale the constraints, add them up

$3x_1 + x_2 + 2x_3 \le 90$

Example

$$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$$

2 x 7 x

s.t. $-x_1 \le 0$, $-x_2 \le 0$, $-x_3 \le 0$

3 x $x_1 + x_2 + 3x_3 \le 30$

$2x_1 + 2x_2 + 5x_3 \le 24$

$4x_1 + x_2 + 2x_3 \le 36$

Scale the constraints, add them up

$-3x_1 - x_2 - 2x_3 \ge -90$ Lower bound on solution

Example

$$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$$

1 x

s.t. $-x_1 \le 0$, $-x_2 \le 0$, $-x_3 \le 0$

$$x_1 + x_2 + 3x_3 \le 30$$

$$2x_1 + 2x_2 + 5x_3 \le 24$$

1 x $4x_1 + x_2 + 2x_3 \le 36$

Scale the constraints, add them up

$$3x_1 + x_2 + 2x_3 \le 36$$

Example

$$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$$

1 x

s.t. $-x_1 \leq 0$, $-x_2 \leq 0$, $-x_3 \leq 0$

$$x_1 + x_2 + 3x_3 \leq 30$$

$$2x_1 + 2x_2 + 5x_3 \leq 24$$

1 x $4x_1 + x_2 + 2x_3 \leq 36$

Scale the constraints, add them up

$-3x_1 - x_2 - 2x_3 \geq -36$ Lower bound on solution

Example

$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$

1 x

s.t. $-x_1 \leq 0$, $-x_2 \leq 0$, $-x_3 \leq 0$

$x_1 + x_2 + 3x_3 \leq 30$

$2x_1 + 2x_2 + 5x_3 \leq 24$

1 x $4x_1 + x_2 + 2x_3 \leq 36$

Scale the constraints, add them up

$-3x_1 - x_2 - 2x_3 \geq -36$ **Tightest lower bound?**

Example

$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$

$y_1 \qquad y_2 \qquad y_3$

s.t. $-x_1 \le 0, -x_2 \le 0, -x_3 \le 0$

y_4 $x_1 + x_2 + 3x_3 \le 30$

y_5 $2x_1 + 2x_2 + 5x_3 \le 24$

y_6 $4x_1 + x_2 + 2x_3 \le 36$

We should be able to add up the inequalities

$y_1, y_2, y_3, y_4, y_5, y_6 \ge 0$

Example

$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$

$y_1 \qquad y_2 \qquad y_3$

s.t. $-x_1 \le 0, -x_2 \le 0, -x_3 \le 0$

y_4 $x_1 + x_2 + 3x_3 \le 30$

y_5 $2x_1 + 2x_2 + 5x_3 \le 24$

y_6 $4x_1 + x_2 + 2x_3 \le 36$

Coefficient of x_1 should be 3

$-y_1 + y_4 + 2y_5 + 4y_6 = 3$

Example

$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$

$y_1 \quad y_2 \quad y_3$

s.t. $-x_1 \le 0, -x_2 \le 0, -x_3 \le 0$

$y_4 \quad x_1 + x_2 + 3x_3 \le 30$

$y_5 \quad 2x_1 + 2x_2 + 5x_3 \le 24$

$y_6 \quad 4x_1 + x_2 + 2x_3 \le 36$

Coefficient of x_2 should be 1

$-y_2 + y_4 + 2y_5 + y_6 = 1$

Example

$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$

$y_1 \quad y_2 \quad y_3$

s.t. $-x_1 \le 0, -x_2 \le 0, -x_3 \le 0$

$y_4 \quad x_1 + x_2 + 3x_3 \le 30$

$y_5 \quad 2x_1 + 2x_2 + 5x_3 \le 24$

$y_6 \quad 4x_1 + x_2 + 2x_3 \le 36$

Coefficient of x_3 should be 2

$-y_3 + 3y_4 + 5y_5 + 2y_6 = 2$

Example

$\min_{\mathbf{x}} -3x_1 - x_2 - 2x_3$

$y_1 \qquad y_2 \qquad y_3$

s.t. $-x_1 \le 0, -x_2 \le 0, -x_3 \le 0$

$y_4 \quad x_1 + x_2 + 3x_3 \le 30$

$y_5 \quad 2x_1 + 2x_2 + 5x_3 \le 24$

$y_6 \quad 4x_1 + x_2 + 2x_3 \le 36$

Lower bound should be tightest

$\max_{\mathbf{y}} -30y_4 - 24y_5 - 36y_6$

Dual

$$\max_y -30y_4 - 24y_5 - 36y_6$$

s.t. $y_1, y_2, y_3, y_4, y_5, y_6 \geq 0$

$$-y_1 + y_4 + 2y_5 + 4y_6 = 3$$

$$-y_2 + y_4 + 2y_5 + y_6 = 1$$

$$-y_3 + 3y_4 + 5y_5 + 2y_6 = 2$$

Original problem is called primal

Dual of dual is primal

$$\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}$$

$$\text{s.t. } A\,\mathbf{x} \leq \mathbf{b}$$

Primal

$$\max_{\mathbf{y}\geq\mathbf{0}} \mathbf{b}^T\mathbf{y}$$

$$\text{s.t. } A^T\mathbf{y} = \mathbf{c}$$

Dual

Weak Duality

Any feasible primal solution **x**

Any feasible dual solution **y**

Primal value at **x** ≥ Dual value at **y**

Dual provides a lower bound

Strong Duality

Optimal primal solution $\mathbf{x}^*$

Optimal dual solution $\mathbf{y}^*$

Primal value at $\mathbf{x}^*$ = Dual value at $\mathbf{y}^*$

Some mild conditions have to be satisfied

Example

Linear Program

$$\min \quad z$$

$$\text{s.t.} \quad -2 \le x_1 \le 2$$

$$-2 \le x_2 \le 2$$

$$a_{in} = x_1 + x_2$$

$$b_{in} = x_1 - x_2$$

Dual feasible solution

Evaluated using dual network

$$a_{out} \ge 0,\ a_{out} \ge a_{in},\ a_{out} \le 0.5a_{in} + 2$$

$$b_{out} \ge 0,\ b_{out} \ge b_{in},\ b_{out} \le 0.5b_{in} + 2$$

$$z = - a_{out} - b_{out}$$

Outline

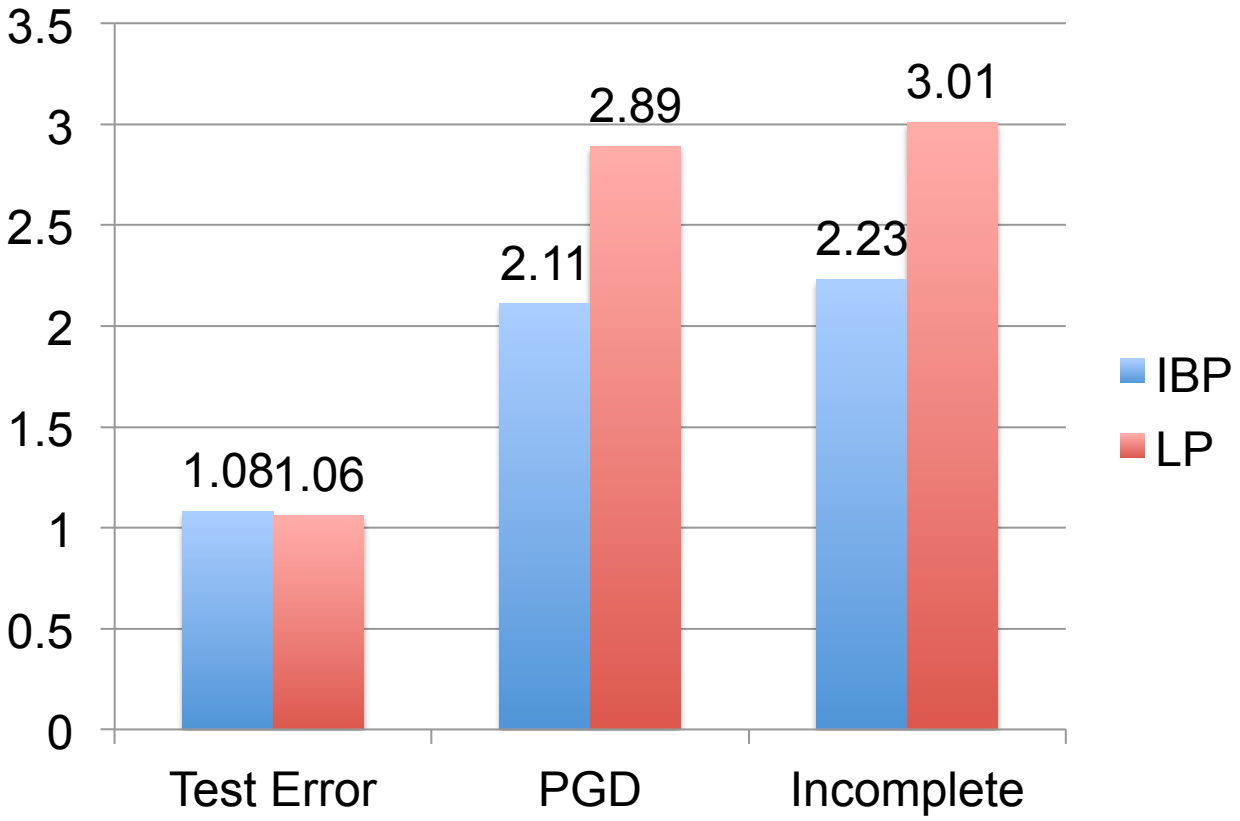
- Interval Bound Propagation
- Linear Programming Relaxation
- **Results**

Experimental Setup

Input model is ε-perturbation of an image

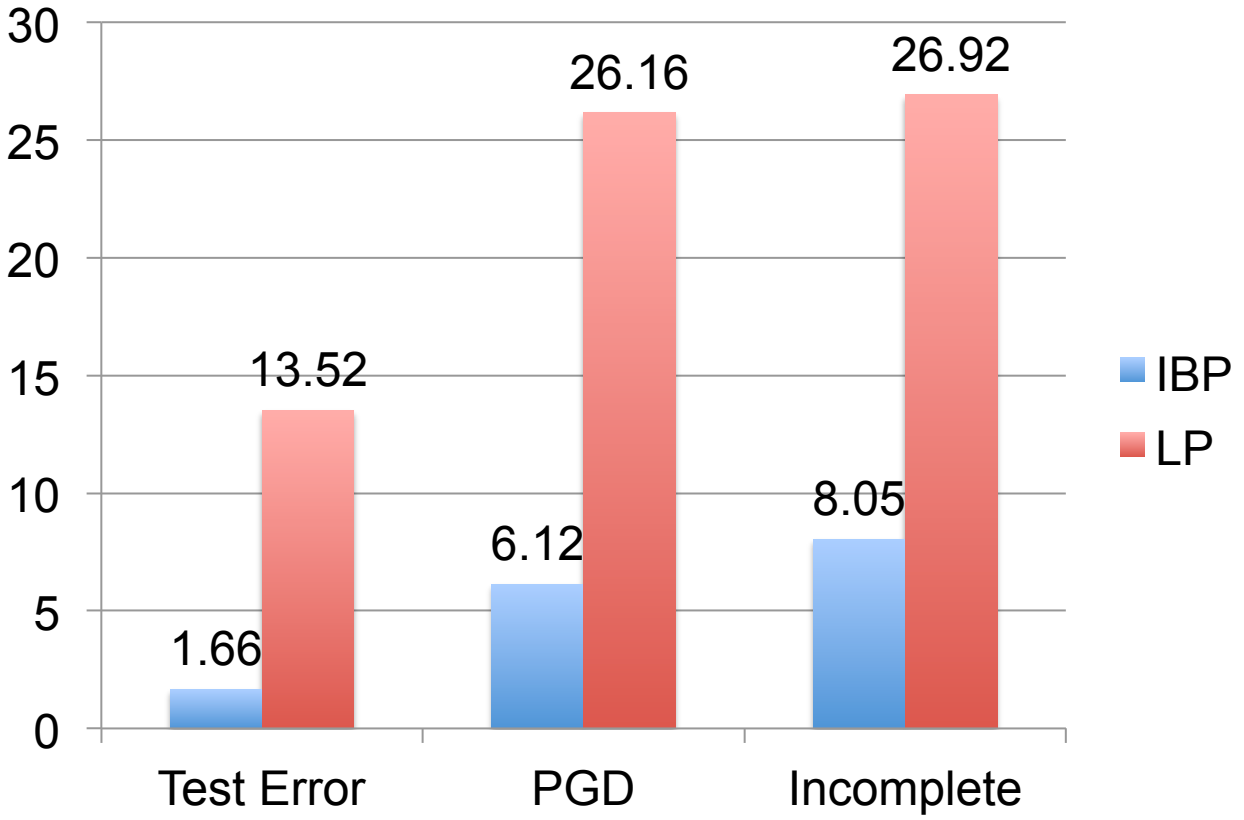
Robust deep learning

MNIST with ε = 0.1



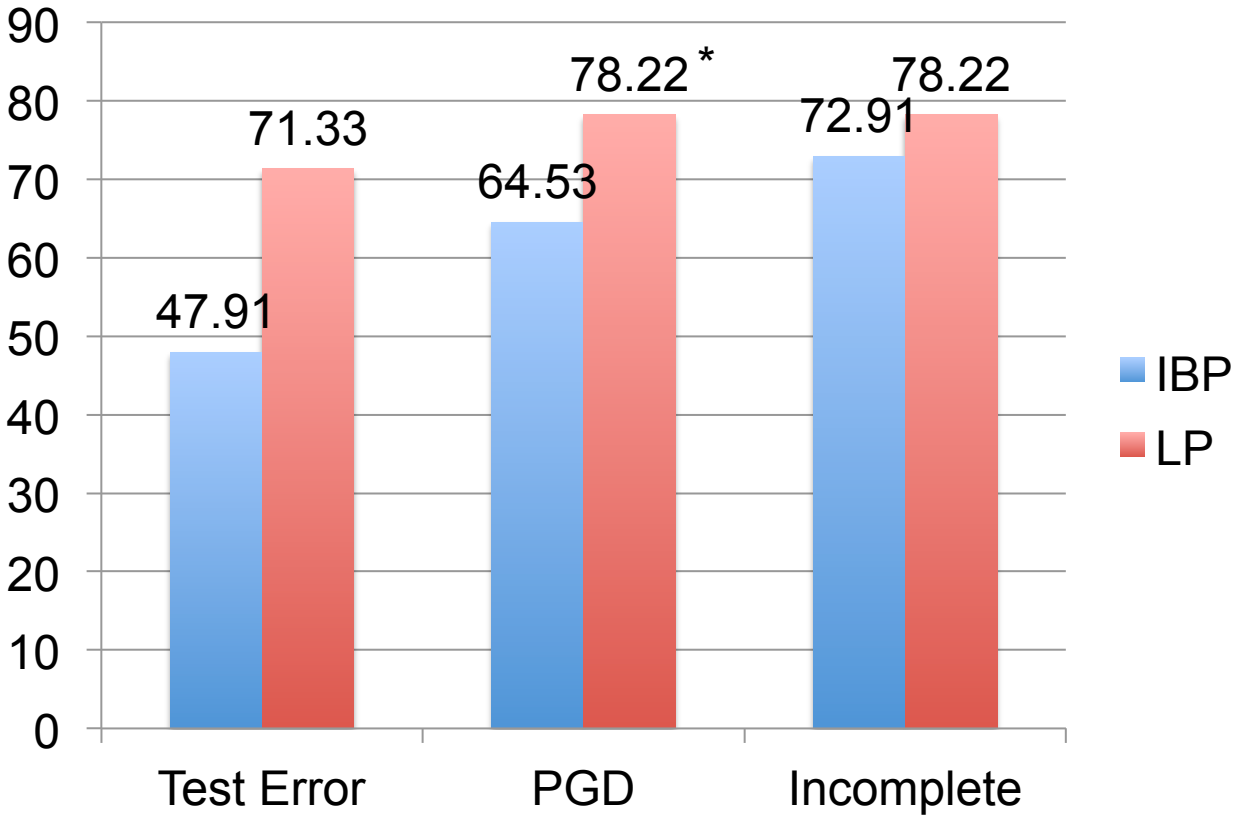
Nominal: Test error = 0.65%, PGD = 27.72%

MNIST with ε = 0.3



Nominal: Test error = 0.65%, PGD = 99.63%

CIFAR-10 with ε = 8/255



Nominal: Test error = 16.66%, PGD = 100%

* PGD error not available (replaced by Incomplete)

Results

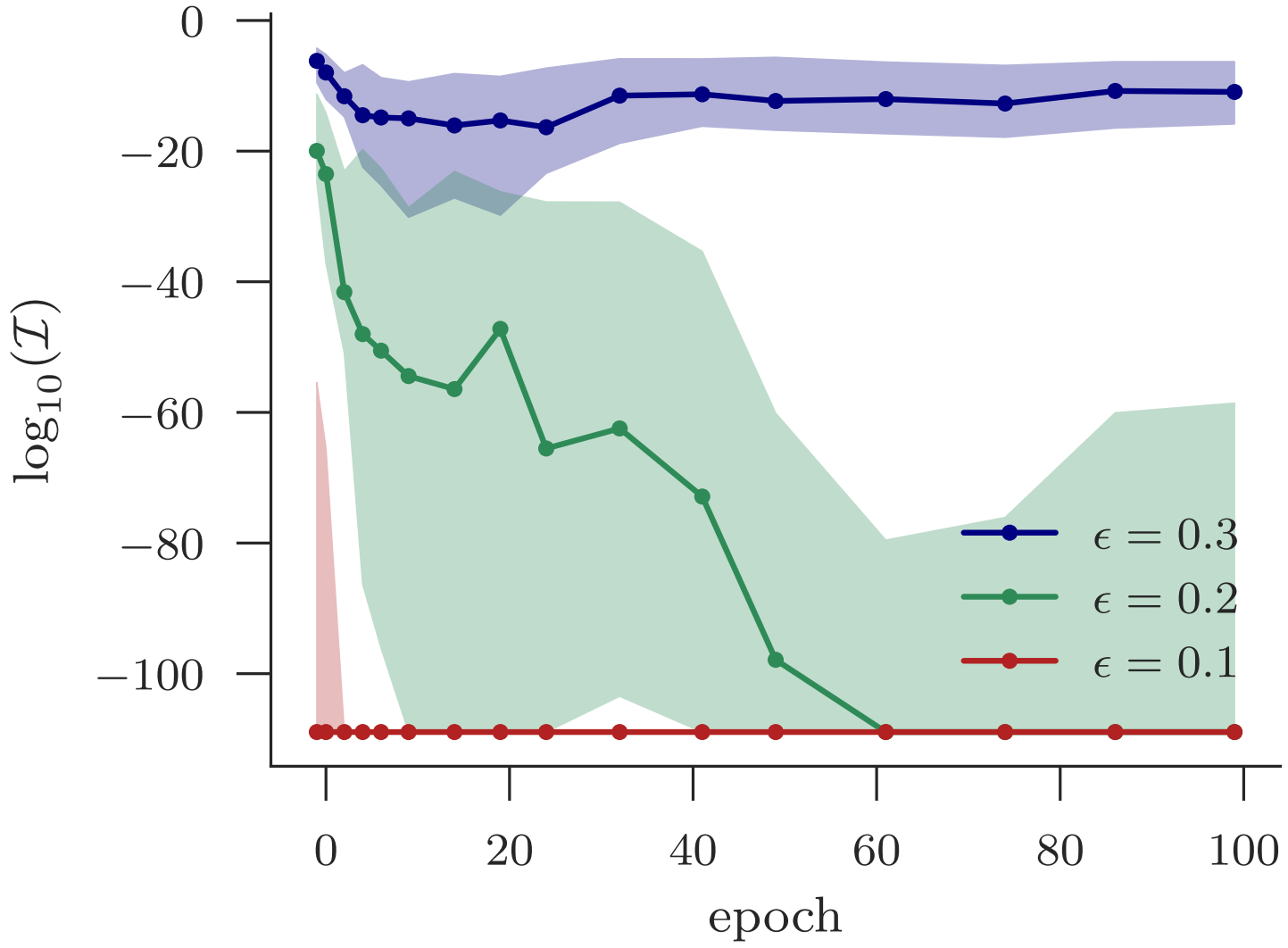
Interval bounds are useful in practice

Significantly faster than LP relaxation methods

Requires careful tuning of more hyperparameters

Still a long way to go to reach “real” networks

Robustness Metric for MNIST



Questions?

References for other incomplete methods on tutorial webpage