Reachability Analysis of Neural Networks with Uncertain Parameters

Pierre-Jean Meyer

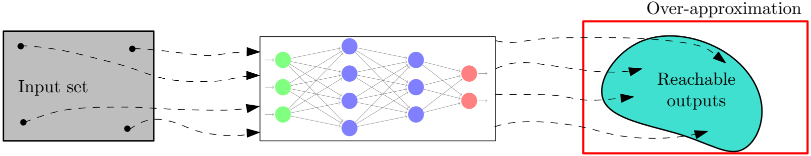
Université Gustave Eiffel

11th of July 2023

Motivations

Formal verification of pre-trained NN

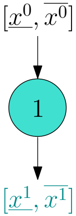
- Many tools based on reachability analysis: ReluVal, Neurify, VeriNet, CROWN, ...



This paper: reachability analysis of uncertain NN

- MMRANN (Meyer, LCSS 2022)
  - Mixed-Monotonicity reachability analysis
  - Requirements: bounded activation function derivative
- ESIP (Neurify; VeriNet; Henriksen, ECAI 2020)
  - Linear bounding functions propagated through the layers
  - Requirements: linear relaxation of nonlinear activation function

Mixed-Monotonicity



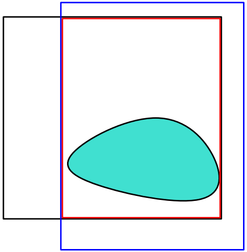
Reachability on each partial network ending at layer *i*

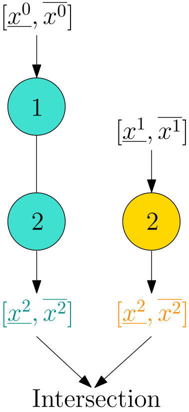
- 1 for layer 1

Mixed-Monotonicity

Reachability on each partial network ending at layer *i*

- 1 for layer 1
- 2 for layer 2





Intersection of two over-approximations

$\rightarrow$ tighter over-approximation

Mixed-Monotonicity

Reachability on each partial network ending at layer *i*

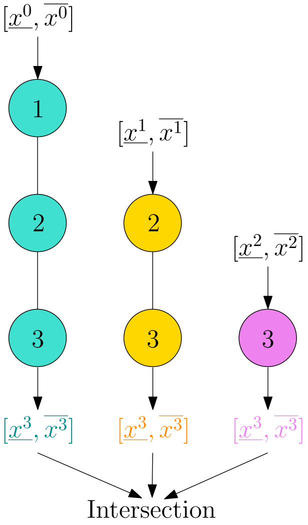
- 1 for layer 1
- 2 for layer 2
- 3 for layer 3
- ...

Stength: generality

- applicable to NN with any continuous activation function

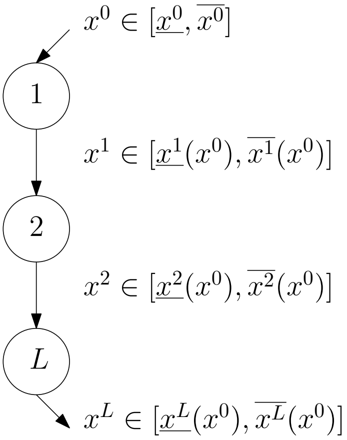
Weakness: polynomial complexity in the network depth L

- $\frac{L(L+1)}{2}$ reachability computations



Error-based Symbolic Interval Propagation

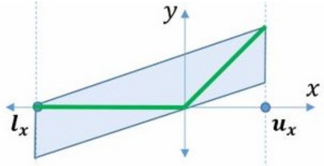
Symbolic Interval Propagation
→ 2 linear functions of x^0 bounding the layers output



Error-based Symbolic Interval Propagation
→ Single linear function and an error matrix

Linear bounding functions

- Deals well with the NN's linear transformations
- Needs **linear relaxation** of the nonlinear activation functions



Stength: low complexity
Weakness: limited to few activation functions (ReLU, sigmoid)

Comparison on pre-trained networks

Comparisons on MNIST benchmarks and random networks[1]

- **Generality**: MM natively handles all continuous activation function

Activation	ReLU PW affine	TanH S-shaped	ELU Monotone	SiLU Non-monotone
ESIP				
MM				

- **Tightness**
  - Complementarity on small ReLU networks
  - ESIP better on large ReLU networks
  - ESIP fails on large non-ReLU networks
- **Complexity**: ESIP 10 to 1000 times faster

[1]Meyer, LCSS 2022

Mixed-monotonicity on uncertain NN

Larger set of uncertain variables

- Network's input x^0
- Weight matrices W^i and bias vectors b^i for all layers i

Minor algorithm updates

- Affine transformation computed with **interval arithmetics**

$$[\underline{W^i}, \overline{W^i}] * [\underline{x^{i-1}}, \overline{x^{i-1}}] + [\underline{b^i}, \overline{b^i}]$$

- Bounding the derivative of the partial network
  $\rightarrow$ derivative **with respect to all uncertainties**

ESIP on uncertain NN

Uncertainty only on input x^0

- Linear function of x^0
- Error matrix E

Uncertainty also on W^i and b^i for all i

- **Multi-linear** function with **input size growing** after each layer

$$[x^0, W^1, b^1, \dots, W^i, b^i]$$

- Error **interval** matrix $[\underline{E}, \overline{E}]$

Implementation of the symbolic equation

- Layer i: stored in $n_i \times (n_0 + 1)$ matrix

Implementation of the symbolic equation

- Store the factor for each multi-linear term

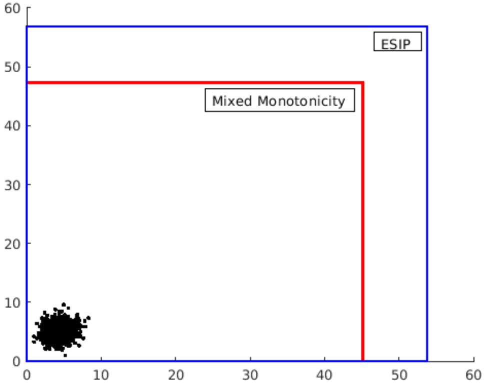
$$n \times \left(\frac{1 - n^{2L+2}}{1 - n}\right)$$

- **Exponential** complexity in the depth L
- **Polynomial** complexity in the width n

Numerical comparisons on random uncertain ReLU NN

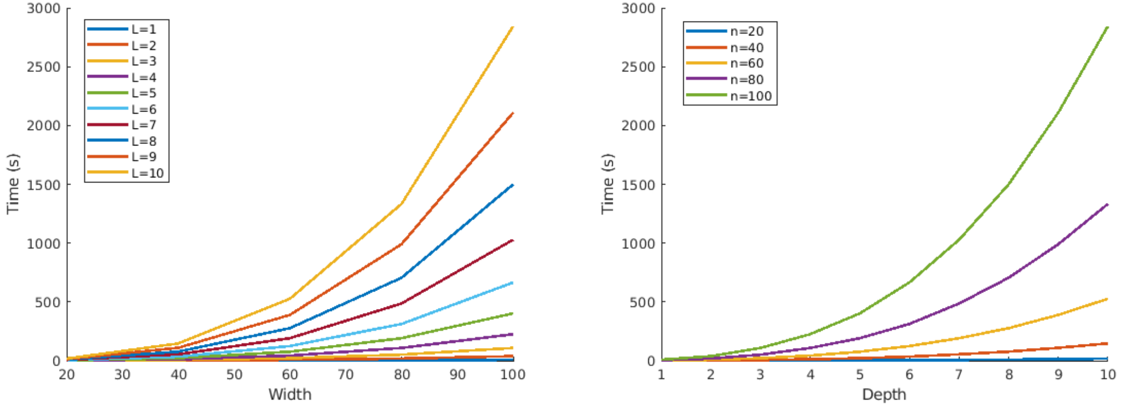
ReLU networks with 20 neurons per layer

		MM	ESIP
$L = 1$	**Time** (s)	0.067	0.79
	Memory (MB)	0.16	0.22
$L = 2$	**Time** (s)	0.25	368
	Memory (MB)	0.46	81
$L = 3$	**Time** (s)	0.64	–
	Memory (MB)	0.90	> 16000



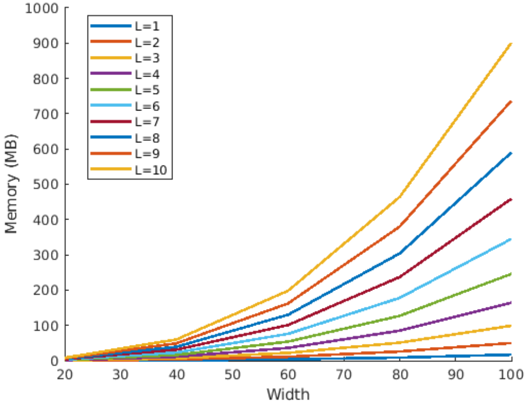
Over-approximation comparison for $L = 2$

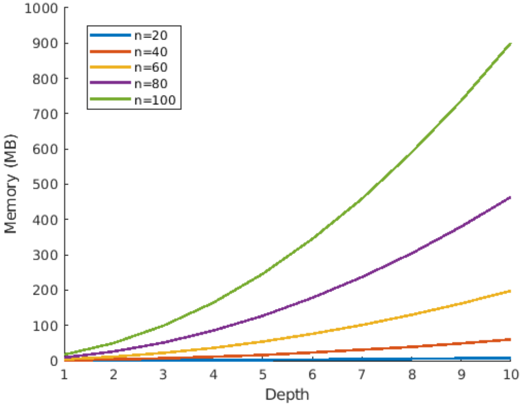
Mixed-monotonicity on random SiLU networks



Computation time: polynomial complexity $O(n^3 * L^3)$ in the width n and depth L

Mixed-monotonicity on random SiLU networks





Memory usage: polynomial complexity $O(n^3 * L^2)$ in the width n and depth L

Conclusion

	Uncertain input	**Uncertain weight/bias**
Complexity (time, memory)	ESIP	Mixed Monotonicity
Generality (activation functions)	Mixed Monotonicity	Mixed Monotonicity
Tightness of over-approximations	Complementary	Mixed Monotonicity

Future work

- Safe training
- Network repair

Contact: `pierre-jean.meyer@univ-eiffel.fr`