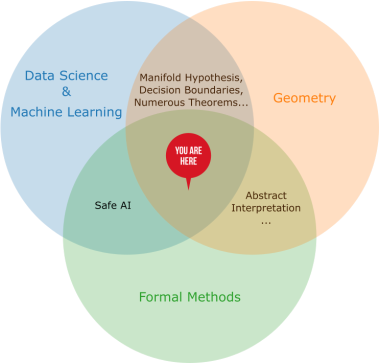
Tropical AI for Safe AI

A Story of Geometrical Data Science
Tropical Abstract Interpretation for Verified Neural Networks

Louis Rustenholz

MMSD – G1 seminar
9 June 2021

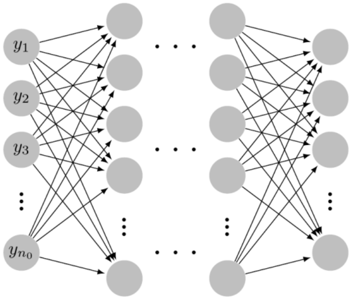


Disclaimer

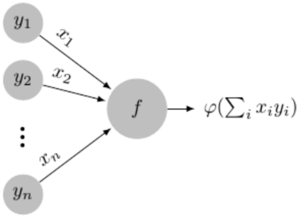
- Miscellaneous stories of Geometrical Data Science.
- Background on tropical geometry, verification of neural networks, abstract interpretation.
- M1 supervised by Éric Goubault and Sylvie Putot at École Polytechnique.

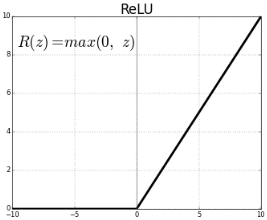
- Our pipeline : work in progress, ideas worth exploring, but complexity yet unsatisfactory (exponential operation hidden between cubical ones).

Neural Networks



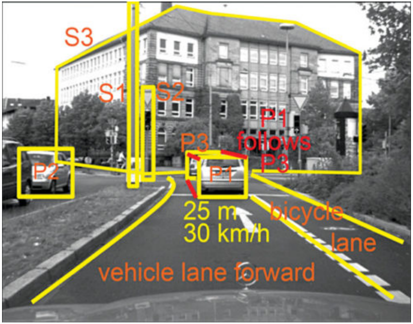
$\mathbb{R}^n \to \mathbb{R}^m$



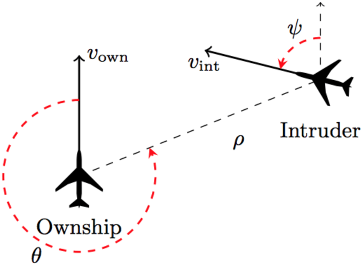


Many applications...

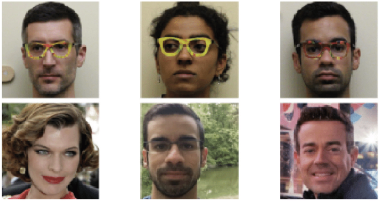
- Vision, self-driving cars

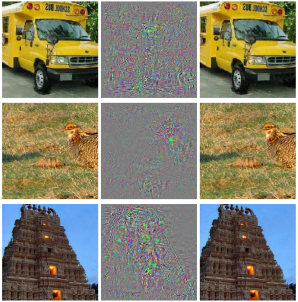


- Aviation : ACAS Xu



...but Neural Networks are unsafe





"Ostrich"

(a) Input 1 (b) Input 2 (darker version of 1)

Machine Learning is geometric !

- In many ways, Data Science and Machine Learning happen where Geometry and Statistics meet.
- Tap into Geometry for understanding, theorems, and solutions !
- ... Geometric ideas for verification problems ?

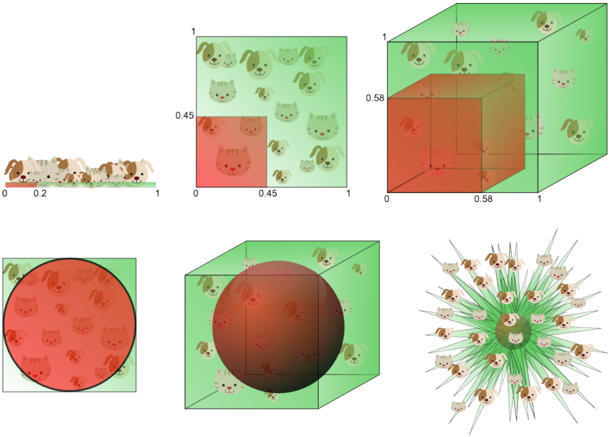
Geometrical Data Science

Geometrical Data Science

Many of ML problems & solutions lie in geometry.

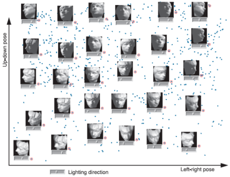
- Curse of dimensionality
- Manifold hypothesis
- Decision boundaries, Kernel trick, Universal Approximation Theorem, ...
- Tropical geometry of Neural Networks

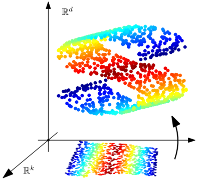
Curse of dimensionality

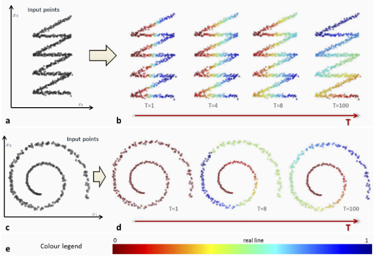


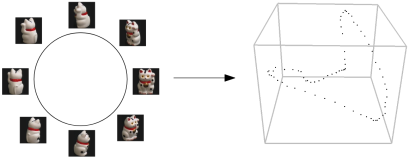
Manifold Hypothesis

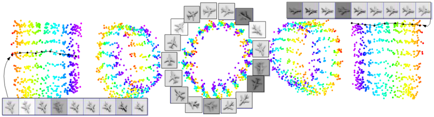
Data density concentrates towards low dimension manifolds



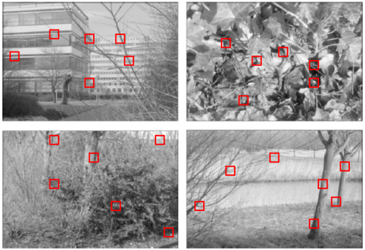




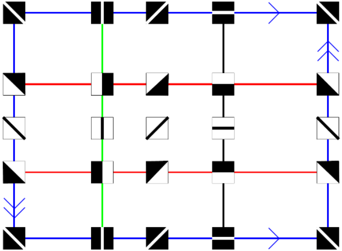


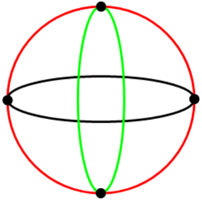


Manifold Hypothesis – TDA (Topological Data Analysis)



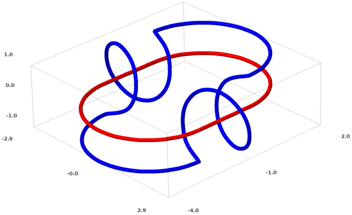
(source: [Lee, Pederson, Mumford 03])



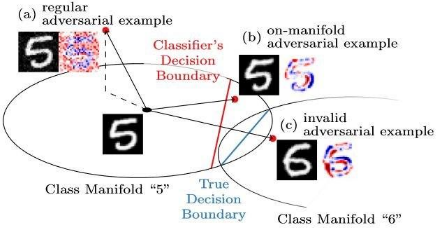


Manifold Hypothesis – Neural Networks

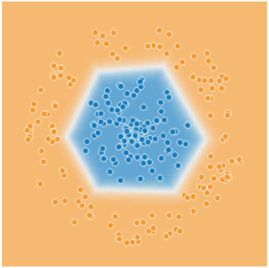
Unlinks

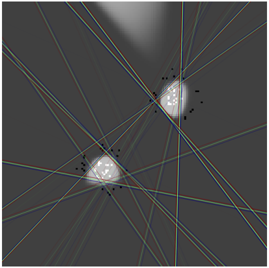


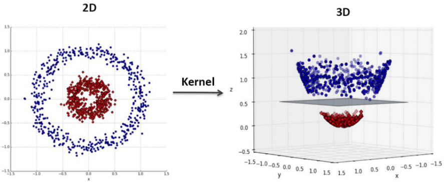
Adversarial examples

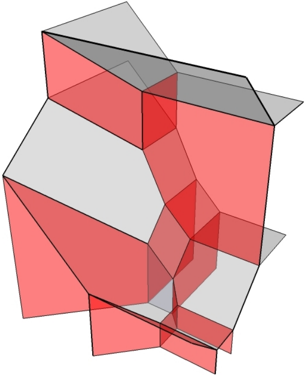


Decision Boundaries

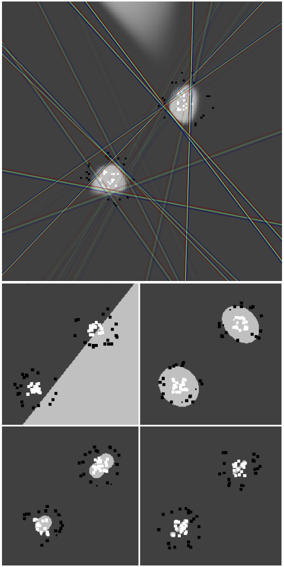


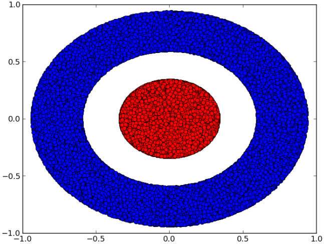


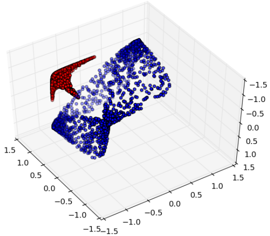




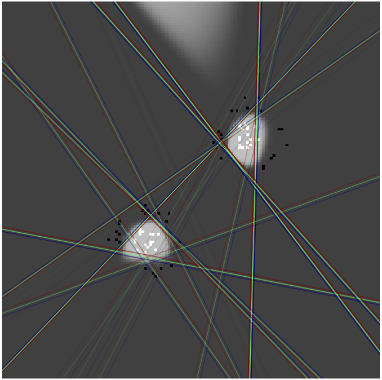
Decision Boundaries – SVM, Kernel trick – Implicit spaces

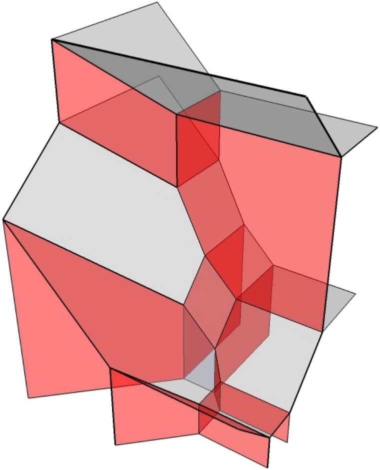






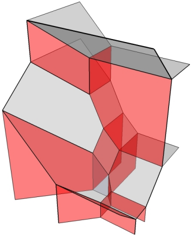
Decision Boundaries – Universal approximation theorems

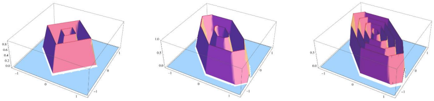




Tropical Decision Boundaries

(ReLU) Neural Networks are, in some sense, tropical objects !





(a) $H_{\frac{1}{2},\frac{1}{2}} \circ N_{\ell_1}$ (b) $H_{\frac{1}{2},\frac{1}{2}} \circ \gamma_{Z(\mathbf{b}^1,\mathbf{b}^2,\mathbf{b}^3,\mathbf{b}^4)}$ (c) $H_{\frac{1}{2},\frac{1}{2},\frac{1}{2}} \circ \gamma_{Z(\mathbf{b}^1,\mathbf{b}^2,\mathbf{b}^3,\mathbf{b}^4)}$

Proposition (Zhang–Naitzat–L 2018)

Let $\nu : \mathbb{R}^d \to \mathbb{R}$ be an L-layer neural network. Write $\nu = f \oslash g$ then

(i) *A decision boundary $\mathcal{B} = \{x \in \mathbb{R}^d : \nu(x) = c\}$ divides $\mathbb{R}^d$ into at most $\operatorname{lin}(f)$ connected regions above c and at most $\operatorname{lin}(g)$ connected regions below c;*

(ii) *The decision boundary is contained in the tropical hypersurface of the tropical polynomial $(c \odot g(x)) \oplus f(x)$, i.e.,*

$$\mathcal{B} \subseteq \mathcal{T}((c \odot g) \oplus f).$$

Corollary (Raghu et al. 2017, Zhang–Naitzat–L 2018)

Assume $n_i \geq d, i = 1, \ldots, L-1$ and $n_L = 1$. The number of linear regions of an L-layer ReLU neural network does not exceed

$$\prod_{i=1}^{L-1} \sum_{j=0}^{d} \binom{n_i}{j} \sim \mathcal{O}(n^{d(L-1)}) \text{ when } n_1 = \cdots = n_{L-1} = n.$$

Tropical geometry

Tropical Geometry

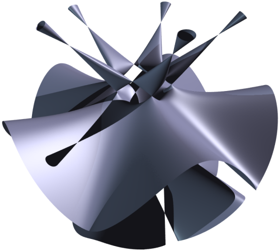
The tropical semiring

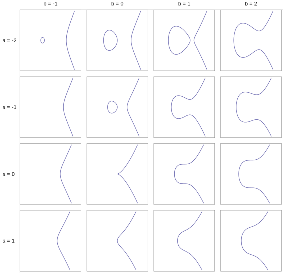
$$(\mathbb{R}_{\max}, \oplus, \otimes)$$

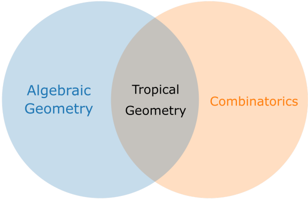
$$\begin{aligned}
\mathbb{R}_{\max} &:= \mathbb{R} \cup \{-\infty\} \\
x \oplus y &:= \max(x, y) \\
x \otimes y &:= x + y \\
\mathbb{1} &:= 0 \\
\mathbb{0} &:= -\infty
\end{aligned}$$

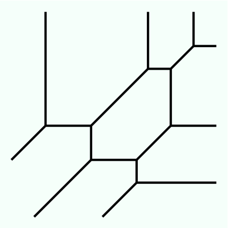
Operations using only $+$ and max are linear in the tropical world.
ReLU layers are linear in the tropical world!

A subject this is rich for pure mathematicians...









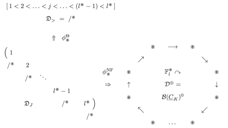
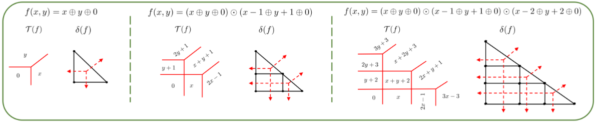
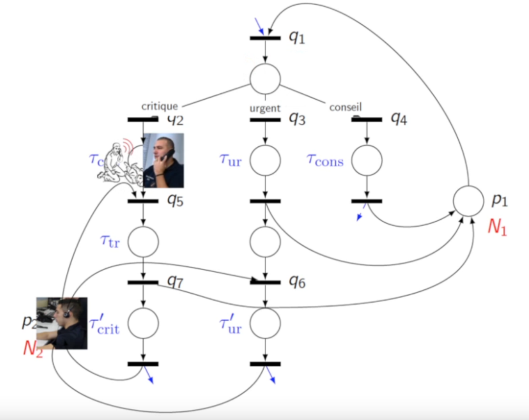


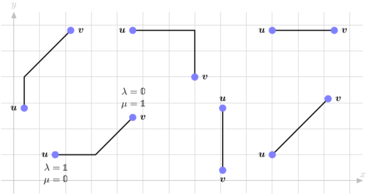
Fig. 4.4: The combinatorial structure of a $\mathcal{D}$-ΘNF-Hodge theater

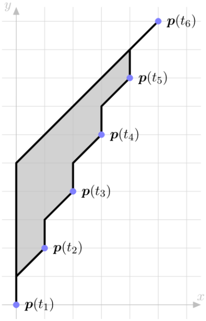


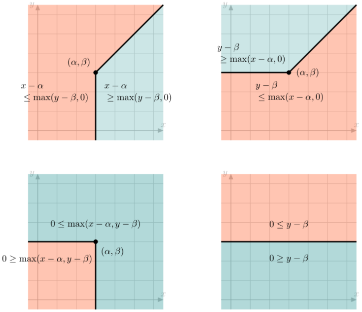
...and has many applications

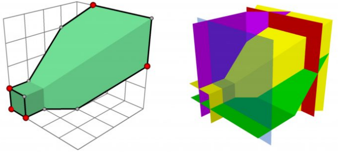


Tropical convex geometry





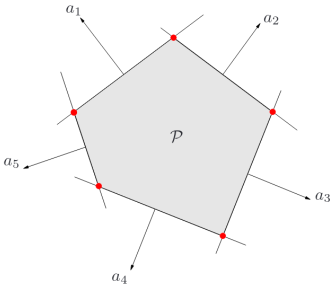


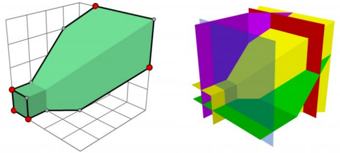


A tropical polytope (left) and the associated arrangement of tropical hyperplanes (right).
Source: X. Allamigeon, P. Benchimol, S. Gaubert, and M. Joswig, Tropicalizing the Simplex Algorithm *SIAM J. Discrete Math.*, 29(2), 751–795.

Tropical Convex Polyhedra – Double Representation

Double representation for classical and tropical polyhedra





A tropical polytope (left) and the associated arrangement of tropical hyperplanes (right).
Source: X. Allamigeon, P. Benchimol, S. Gaubert, and M. Joswig, Tropicalizing the Simplex Algorithm *SIAM J. Discrete Math.*, 29(2), 751–795.

Conversion is expensive !

Tropical Convex Polyhedra – Double Representation

- External description with constraints

$$\left\{X \in (\mathbb{R}_{\max})^d \,\middle|\, AX \leq BX \text{ in the tropical sense}\right\}$$

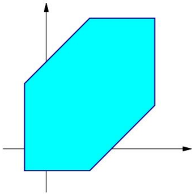
- Internal description with generators

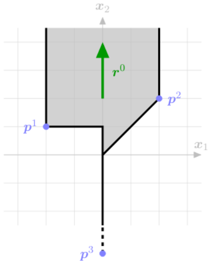
$$co(P) \oplus cone(R), \text{where}$$

$$co(P) = \left\{ \bigoplus_i \lambda_i \odot p_i \,\middle|\, \lambda_i \in \mathbb{R}_{\max},\ p_i \in P,\ \bigoplus_i \lambda_i = 0 \right\},$$

$$cone(R) = \left\{ \bigoplus_i \lambda_i \odot p_i \,\middle|\, \lambda_i \in \mathbb{R}_{\max},\ p_i \in P \right\}.$$

An open problem : Classical Zones and Tropical Polyhedra

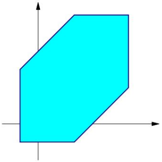


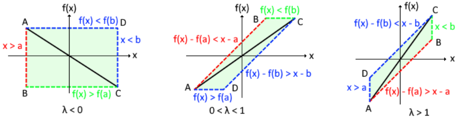


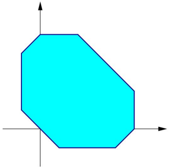
- Conversion Zone $\rightarrow$ TropPoly
  - Immediate for constraint representation
  - Can be done cheaply for internal representation ($n+1$ extreme points, $O(n^2)$)
- TropPoly $\rightarrow$ Zone
  - We can compute a tight overapproximation ($O(n^3)$).
  - TropPoly are unions of Zones.
  - *Which unions of zones are tropical polyhedra?*

MinMaxPoly : into higher-order geometry

Enrich with negative slopes.
Convex geometry is with polynomial of degree 1. Add degree -1.







Use 2 dimensions for each variable x_i.

$$\left\{ X \in (\mathbb{R}_{\max})^d \;\middle|\; A \begin{pmatrix} +X \\ -X \end{pmatrix} \leq B \begin{pmatrix} +X \\ -X \end{pmatrix} \right\}$$

Warning : $x_i \otimes (-x_i) = (x_i) + (-x_i) = 0$
is non-linear in the tropical world.

Safe Neural Networks

Verification of Neural Networks

$$\mathrm{NN} : \mathbb{R}^n \to \mathbb{R}^m \text{ Fully connected ReLU network}$$

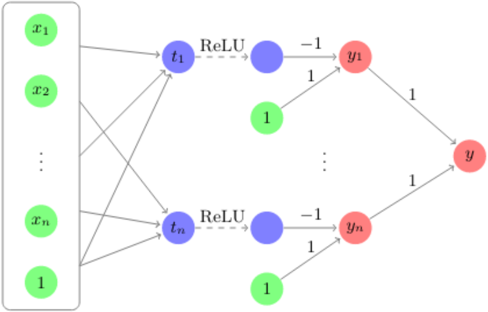
- Φ a linear property between input and output.
- Do we have $\forall x\, \Phi(x, \mathrm{NN}(x))$, i.e.

$$\mathrm{NN} \vDash \Phi\,?$$

- Decidable, but NP-hard.
- Other kinds of properties : local robustness, fairness, ...

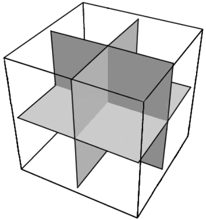
NP-hardness

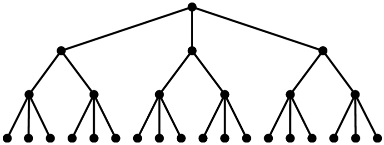
Reduction to 3-SAT by Reluplex authors (Guy Katz et al., 2017)



Some techniques

- First methods : Linear Programming + Branch-and-Bound
  MILP, SMT, ...
- A recent problem !
  Reluplex (2017) can deal with 20 neurons, using an extended simplex algorithm, encoded in SMT.
- Other lines of research : (extended ?) polyhedra, abstract interpretation...

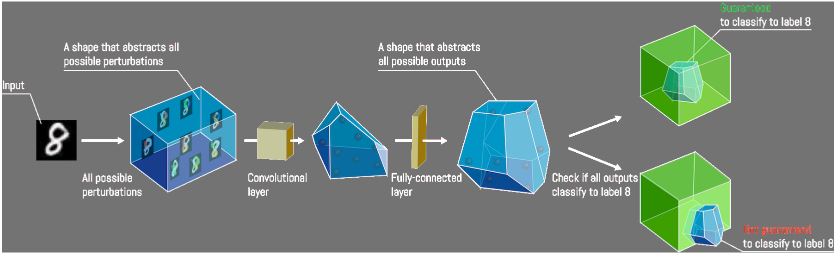




Combinatorial explosion

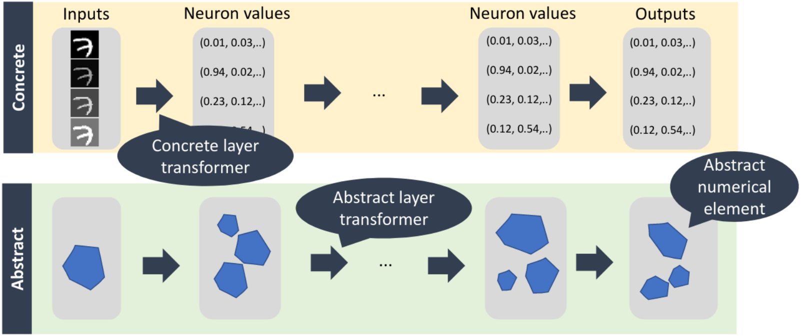
Geometric methods : AI^2

Abstract Interpretation for Artificial Intelligence
[SafeAI, ETHZürich, Martin Vechez et al., 2018].



- Local protection against adversarial examples.
- Cubes ? Zonotopes ? Polyhedra ?

Geometric methods : AI^2

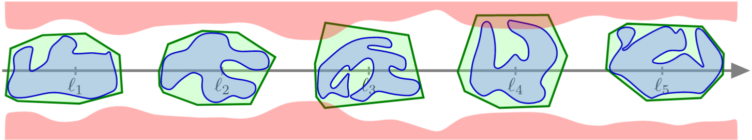


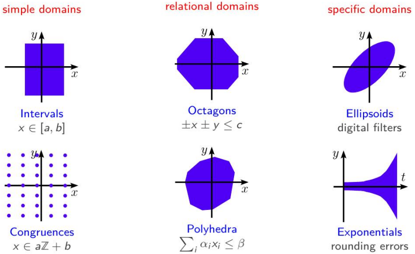
- Precision/Complexity trade-off.
- ReLU layers create trouble.

Abstract Interpretation

Abstract Interpretation

A general theory of sound approximations for program semantics.

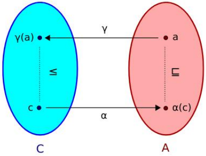




Abstract domains

Abstract Interpretation

A general theory of sound approximations for program semantics.



Galois Connections (Adjunctions)

Cousot-Cousot

```
(S0)
assume X in [0,1000];
(S1)
I := 0;
(S2)
while (S3) I < X do
    (S4)
    I := I + 2;
    (S5)
(S6)
```

program

$$\mathcal{S}_i \in \mathcal{D} \overset{\text{def}}{=} \mathcal{P}(\{I, X\} \to \mathbb{Z})$$

$$\mathcal{S}_0 = \{\, (i, x) \mid i, x \in \mathbb{Z} \,\}$$
$$\mathcal{S}_1 = [\![X \in [0, 1000]]\!] (\mathcal{S}_0)$$
$$\mathcal{S}_2 = [\![I \leftarrow 0]\!] (\mathcal{S}_1)$$
$$\mathcal{S}_3 = \mathcal{S}_2 \cup \mathcal{S}_5$$
$$\mathcal{S}_4 = [\![I < X]\!] (\mathcal{S}_3)$$
$$\mathcal{S}_5 = [\![I \leftarrow I + 2]\!] (\mathcal{S}_4)$$
$$\mathcal{S}_6 = [\![I \geq X]\!] (\mathcal{S}_3)$$

concrete semantics

$$\mathcal{S}_i^\sharp \in \mathcal{D}^\sharp$$

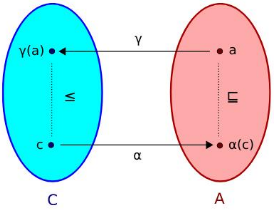
$$\mathcal{S}_0^\sharp = \top^\sharp$$
$$\mathcal{S}_1^\sharp = [\![X \in [0, 1000]]\!]^\sharp (\mathcal{S}_0^\sharp)$$
$$\mathcal{S}_2^\sharp = [\![I \leftarrow 0]\!]^\sharp (\mathcal{S}_1^\sharp)$$
$$\mathcal{S}_3^\sharp = \mathcal{S}_2^\sharp \cup^\sharp \mathcal{S}_5^\sharp$$
$$\mathcal{S}_4^\sharp = [\![I < X]\!]^\sharp (\mathcal{S}_3^\sharp)$$
$$\mathcal{S}_5^\sharp = [\![I \leftarrow I + 2]\!]^\sharp (\mathcal{S}_4^\sharp)$$
$$\mathcal{S}_6^\sharp = [\![I \geq X]\!]^\sharp (\mathcal{S}_3^\sharp)$$

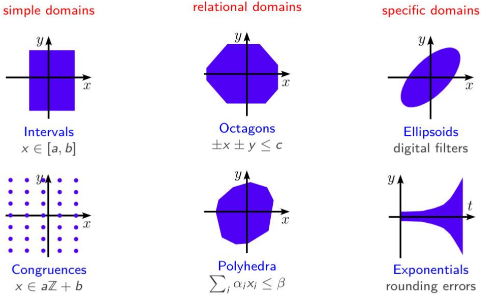
abstract semantics

Verification, compiler optimization, ...

Examples of Abstract Domains

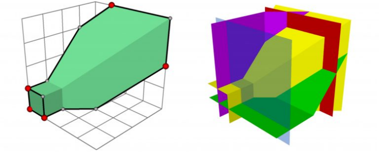
domain	invariants	memory cost	time cost (per operation)
intervals	$V \in [\ell, h]$	$\mathcal{O}(\lvert n \rvert)$	$\mathcal{O}(\lvert n \rvert)$
linear equalities	$\sum_i \alpha_i V_i = \beta_i$	$\mathcal{O}(\lvert n \rvert^2)$	$\mathcal{O}(\lvert n \rvert^3)$
zones	$V_i - V_j \leq c$	$\mathcal{O}(\lvert n \rvert^2)$	$\mathcal{O}(\lvert n \rvert^3)$
polyhedra	$\sum_i \alpha_i V_i \geq \beta_i$	unbounded, exponential in practice	



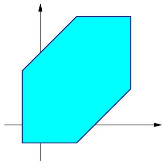


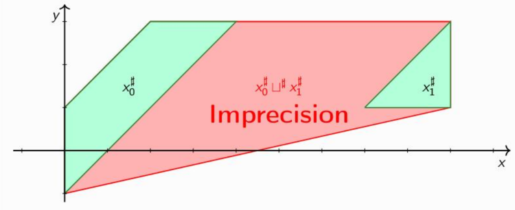
Tropical Abstract Domains

Much work done in Xavier Allamigeon's thesis, for memory models.



A tropical polytope (left) and the associated arrangement of tropical hyperplanes (right). Source: X. Allamigeon, P. Benchimol, S. Gaubert, and M. Joswig, Tropicalizing the Simplex Algorithm *SIAM J. Discrete Math.*, 29(2), 751–795.





Non-disjunctive non-convex abstract domains ?
Back to this zone question...

Our pipeline

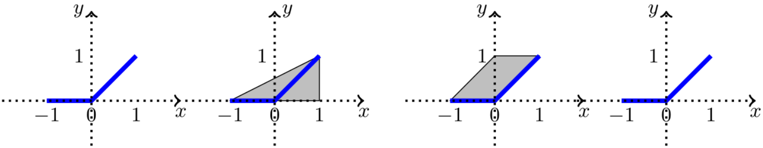
Our problem

$$\mathrm{NN} : \mathbb{R}^n \to \mathbb{R}^m \text{ Fully connected ReLU network}$$

- Φ a linear property between input and output.
- Do we have $\forall x\, \Phi(x, \mathrm{NN}(x))$, i.e.

$$\mathrm{NN} \vDash \Phi\,?$$

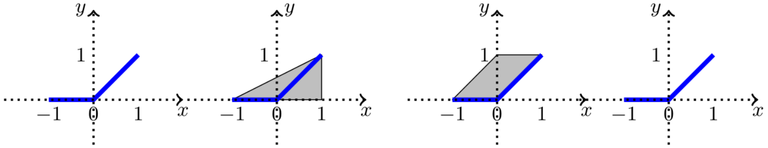
Éric Goubault, Sylvie Putot, Sébastien Palumby, Sriram Sankaranarayanan, Xavier Allamigeon, ...



(a) Exact (b) 1-ReLU (DeepPoly) (c) Zones (d) Tropical polyhedra

Our problem

$$\mathrm{NN} \models \Phi\,?$$



(a) Exact (b) 1-ReLU (DeepPoly) (c) Zones (d) Tropical polyhedra

We will do abstract interpretation, and work the families T of the tropical polyhedra, the extension $T_{\pm}$, hypercubes K, zones Z, and octagons Oct.

Disclaimer

- New ideas to be explored
- Complexity yet unsatisfactory : an exponential operation hidden between cubical ones

Abstraction

Abstract interpretation in $T_{(\pm)}$.
We need to do the following operations on tropical polyhedra.

- ReLU layers. $\mathrm{ReLU} : T \to T$
- Linear layers. $L : T \to T$.
- Verification of linear properties. $\Phi \in (\mathbb{R}^n)^*$, $t \in T_{(\pm)}$. $t \models \Phi$?

Abstraction

Abstract interpretation in $T_{(\pm)}$.
We need to do the following operations on tropical polyhedra.

- ReLU layers. $\mathrm{ReLU} : T \to T$. Exact !

$$T \xrightarrow{\mathbb{1}\oplus\cdot} T$$

- Linear layers. $L : T \to T$.

- Verification of linear properties. $\Phi \in (\mathbb{R}^n)^*$, $t \in T_{(\pm)}$. $t \vDash \Phi$?

Abstraction

Abstract interpretation in $T_{(\pm)}$.
We need to do the following operations on tropical polyhedra.

- ReLU layers. $\mathrm{ReLU} : T \to T$. Exact !

$$T \xrightarrow{\mathbb{1}\oplus\cdot} T$$

- Linear layers. $L : T \to T$.

- Verification of linear properties. $\Phi \in (\mathbb{R}^n)^*$, $t \in T_{(\pm)}$. $t \models \Phi$?

$$T_{\pm} \to Z_{\pm} \to \mathrm{Oct} \Rightarrow \mathsf{LP}$$

or MILP, Tropical Fourier-Motzkin, ... (but we don't earn much)

Abstraction

Abstract interpretation in $T_{(\pm)}$.
We need to do the following operations on tropical polyhedra.

- ReLU layers. $\mathrm{ReLU} : T \to T$. Exact !

$$T \xrightarrow{\mathbb{1}\oplus\cdot} T$$

- Linear layers. $L : T \to T$.

$$T \to K \to Z \to T$$

- Verification of linear properties. $\Phi \in (\mathbb{R}^n)^*$, $t \in T_{(\pm)}$. $t \models \Phi$?

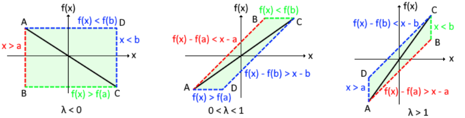
$$T_\pm \to Z_\pm \to \mathrm{Oct} \Rightarrow \mathsf{LP}$$

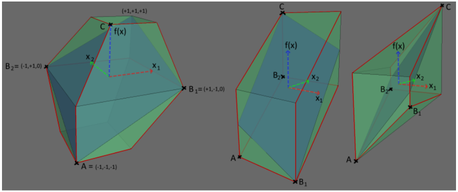
Abstraction of the linear layer

$$L : x \mapsto \left(\sum \lambda_{ij} x_i \right)_j$$

- Linear layers. $L : T \to T$.

$$T \to K \to Z \to T$$



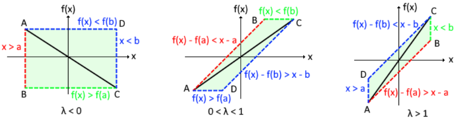


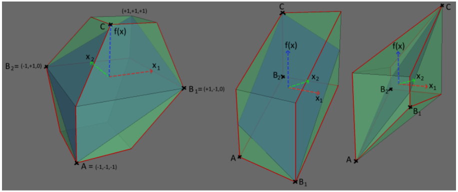
Abstraction of the linear layer ($T_{\pm}$)

$$L : x \mapsto \left(\sum \lambda_{ij} x_i \right)_j$$

- Linear layers. $L : T_{\pm} \to T_{\pm}$.

$$T_{\pm} \to K \to \mathrm{Oct} \to T_{\pm}$$



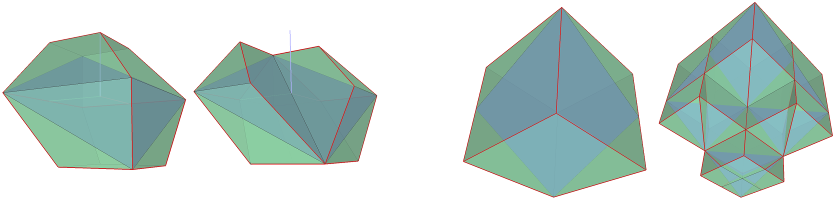


Abstraction of the linear layer (Subdivisions)

$$L : x \mapsto \left(\sum \lambda_{ij} x_i \right)_j$$

- Linear layers. $L : T \to T$.

$$T \to K \to Z \to T$$



Complexity and bottlenecks

- "Mostly" $O(n^3)$
- The double representation problem. Conversion is costly !

	External representation (constraints)	Internal representation (generators)
ReLU layer	Easy	Trivial
Linear layer	OK	OK
Hypercube approx.	?	Trivial
Subdivisions	OK	OK
Multi-layer (Intersection)	OK	?
$\mathrm{NN} \models \Phi\,?$	?	OK, LP

Conclusion

Conclusion

- Conclusion on our pipeline
- Geometrical methods in Machine Learning
- Next steps
  - Full tropical abstraction, without cubes.
  - Zonotopes/cubes relation, tightest zonotope around octagon.
  - T, $T_{\pm}$... Go to higher degrees. Tropical Gröbner bases.
- Open problems
  - **Double description problem.**
    Can we avoid double description in our pipeline ?
    More generally, can we improve the conversion algorithm, maybe for subfamilies of problems ?
  - **Disjunction problem.**
    Which union of zones are tropical polyhedra ?

Thank you !

- **Double description problem.**
  Can we avoid double description in our pipeline ?
  Can the conversion algorithm be improved ?
- **Disjunction problem.**
  Which union of zones are tropical polyhedra ?