Complexity of Deciding Injectivity and Surjectivity of ReLU Networks

Math Machine Learning Seminar

November 8, 2025

Vincent Froese, **Moritz Grillo** and Martin Skutella (TU Berlin)

Introduction

- A neural network is a parameterized function $\phi_\theta \colon \mathbb{R}^d \to \mathbb{R}^m$.

Introduction

- A neural network is a parameterized function $\phi_\theta \colon \mathbb{R}^d \to \mathbb{R}^m$.
- Training of neural networks $=$ optimizing θ with heuristics in order to fit ϕ_θ to some data set

Introduction

- A neural network is a parameterized function $\phi_\theta \colon \mathbb{R}^d \to \mathbb{R}^m$.
- Training of neural networks $=$ optimizing θ with heuristics in order to fit ϕ_θ to some data set
- Given θ, what can we say about the properties of ϕ_θ?

Introduction

- A neural network is a parameterized function $\phi_\theta\colon \mathbb{R}^d \to \mathbb{R}^m$.
- Training of neural networks = optimizing θ with heuristics in order to fit ϕ_θ to some data set
- Given θ, what can we say about the properties of ϕ_θ?
- Injectivity is important whenever invertibility is relevant (inverse problems or likelihood estimation)

Introduction

- A neural network is a parameterized function $\phi_\theta \colon \mathbb{R}^d \to \mathbb{R}^m$.
- Training of neural networks $=$ optimizing θ with heuristics in order to fit ϕ_θ to some data set
- Given θ, what can we say about the properties of ϕ_θ?
- Injectivity is important whenever invertibility is relevant (inverse problems or likelihood estimation)
- Surjectivity is closely related to network verification.

Introduction

- A neural network is a parameterized function $\phi_\theta \colon \mathbb{R}^d \to \mathbb{R}^m$.
- Training of neural networks = optimizing θ with heuristics in order to fit ϕ_θ to some data set
- Given θ, what can we say about the properties of ϕ_θ?
- Injectivity is important whenever invertibility is relevant (inverse problems or likelihood estimation)
- Surjectivity is closely related to network verification.
- Network verification: Given $P \subseteq \mathbb{R}^d$, $Q \subseteq \mathbb{R}^m$, does it hold that $\phi_\theta(P) \subseteq Q$? Important in safety-critical applications.

Overview

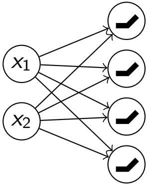
ReLU-Layers and their Geometry

Computational Complexity of Injectivity

Computational Complexity of Surjectivity

ReLU-Layer and its Geometry

ReLU-layer with d input and m output neurons given by weight matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, bias vector $\mathbf{b} \in \mathbb{R}^{m}$

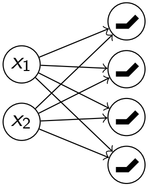


ReLU-Layer and its Geometry

ReLU-layer with d input and m output neurons given by weight matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, bias vector $\mathbf{b} \in \mathbb{R}^m$

Computes map

$$\phi_{\mathbf{W},\mathbf{b}} \colon \mathbb{R}^d \to \mathbb{R}^m, \quad \mathbf{x} \mapsto [\mathbf{W}\mathbf{x} + \mathbf{b}]_+$$

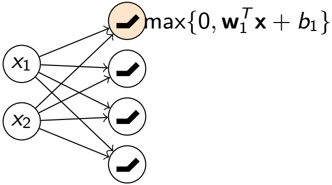


ReLU-Layer and its Geometry

ReLU-layer with d input and m output neurons given by weight matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, bias vector $\mathbf{b} \in \mathbb{R}^m$

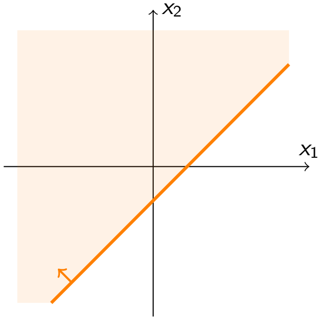
Computes map

$$\phi_{\mathbf{W},\mathbf{b}} : \mathbb{R}^d \to \mathbb{R}^m, \quad \mathbf{x} \mapsto [\mathbf{W}\mathbf{x} + \mathbf{b}]_+$$



Terminology

- output neuron i active at $\mathbf{x} \in \mathbb{R}^d$ if $\mathbf{w}_i^T \mathbf{x} + b_i \geq 0$

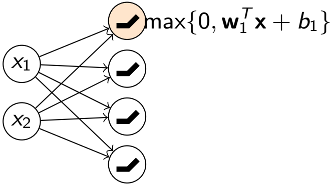


ReLU-Layer and its Geometry

ReLU-layer with d input and m output neurons given by weight matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, bias vector $\mathbf{b} \in \mathbb{R}^m$

Computes map

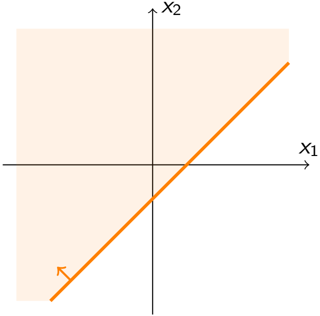
$$\phi_{\mathbf{W},\mathbf{b}} \colon \mathbb{R}^d \to \mathbb{R}^m, \quad \mathbf{x} \mapsto [\mathbf{W}\mathbf{x} + \mathbf{b}]_+$$



Terminology

- output neuron i active at $\mathbf{x} \in \mathbb{R}^d$ if $\mathbf{w}_i^T\mathbf{x} + b_i \geq 0$
- output neurons induce (oriented) hyperplanes

$$H_{\mathbf{w}_i,b_i} := \{\mathbf{w}_i^T\mathbf{x} + b_i = 0\} \subseteq \mathbb{R}^d$$

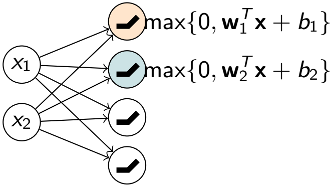


ReLU-Layer and its Geometry

ReLU-layer with d input and m output neurons given by weight matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, bias vector $\mathbf{b} \in \mathbb{R}^m$

Computes map

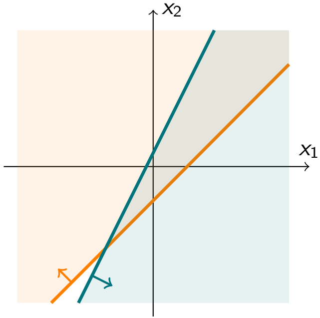
$$\phi_{\mathbf{W},\mathbf{b}} \colon \mathbb{R}^d \to \mathbb{R}^m, \quad \mathbf{x} \mapsto [\mathbf{W}\mathbf{x} + \mathbf{b}]_+$$



Terminology

- output neuron i active at $\mathbf{x} \in \mathbb{R}^d$ if $\mathbf{w}_i^T \mathbf{x} + b_i \geq 0$
- output neurons induce (oriented) hyperplanes

$$H_{\mathbf{w}_i, b_i} := \{\mathbf{w}_i^T \mathbf{x} + b_i = 0\} \subseteq \mathbb{R}^d$$

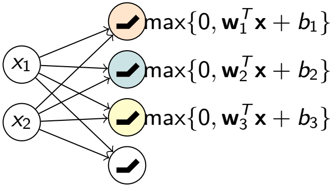


ReLU-Layer and its Geometry

ReLU-layer with d input and m output neurons given by weight matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, bias vector $\mathbf{b} \in \mathbb{R}^m$

Computes map

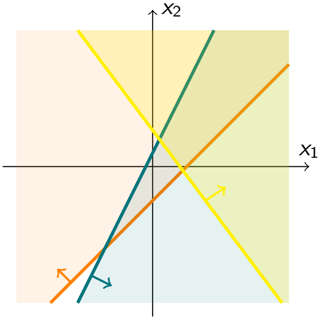
$$\phi_{\mathbf{W},\mathbf{b}} \colon \mathbb{R}^d \to \mathbb{R}^m, \quad \mathbf{x} \mapsto [\mathbf{W}\mathbf{x} + \mathbf{b}]_+$$



Terminology

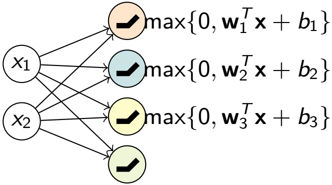
- output neuron i active at $\mathbf{x} \in \mathbb{R}^d$ if $\mathbf{w}_i^T \mathbf{x} + b_i \geq 0$
- output neurons induce (oriented) hyperplanes

$$H_{\mathbf{w}_i, b_i} := \{\mathbf{w}_i^T \mathbf{x} + b_i = 0\} \subseteq \mathbb{R}^d$$



ReLU-Layer and its Geometry

ReLU-layer with d input and m output neurons given by weight matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, bias vector $\mathbf{b} \in \mathbb{R}^m$



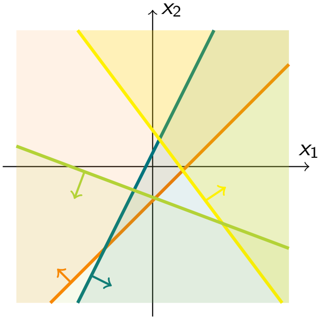
Computes map

$$\phi_{\mathbf{W},\mathbf{b}} \colon \mathbb{R}^d \to \mathbb{R}^m, \quad \mathbf{x} \mapsto [\mathbf{W}\mathbf{x} + \mathbf{b}]_+$$

Terminology

- output neuron i active at $\mathbf{x} \in \mathbb{R}^d$ if $\mathbf{w}_i^T \mathbf{x} + b_i \geq 0$
- output neurons induce (oriented) hyperplanes

$$H_{\mathbf{w}_i, b_i} := \{\mathbf{w}_i^T \mathbf{x} + b_i = 0\} \subseteq \mathbb{R}^d$$

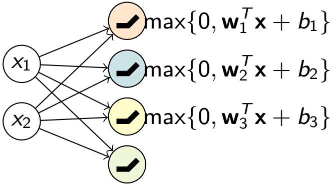


ReLU-Layer and its Geometry

ReLU-layer with d input and m output neurons given by weight matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, bias vector $\mathbf{b} \in \mathbb{R}^m$

Computes map

$$\phi_{\mathbf{W},\mathbf{b}} \colon \mathbb{R}^d \to \mathbb{R}^m, \quad \mathbf{x} \mapsto [\mathbf{W}\mathbf{x} + \mathbf{b}]_+$$

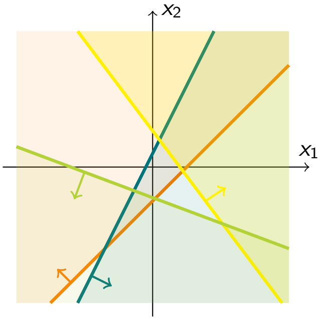


Terminology

- output neuron i active at $\mathbf{x} \in \mathbb{R}^d$ if $\mathbf{w}_i^T \mathbf{x} + b_i \geq 0$
- output neurons induce (oriented) hyperplanes

$$H_{\mathbf{w}_i, b_i} := \{\mathbf{w}_i^T \mathbf{x} + b_i = 0\} \subseteq \mathbb{R}^d$$

- they partition $\mathbb{R}^d$ into polyhedral cells, corresponding to subsets of active neurons

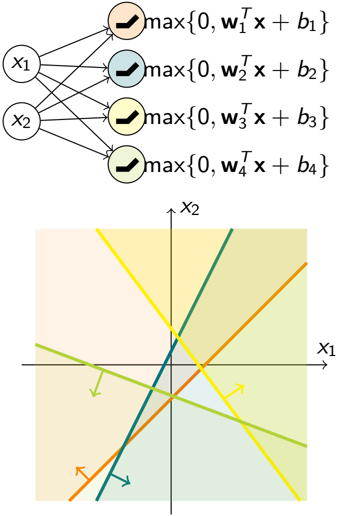


ReLU-Layer: Injectivity

ReLU-Layer Injectivity

Given: matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, vector $\mathbf{b} \in \mathbb{R}^m$

Question: is the map $\phi_{\mathbf{W},\mathbf{b}}$ injective?



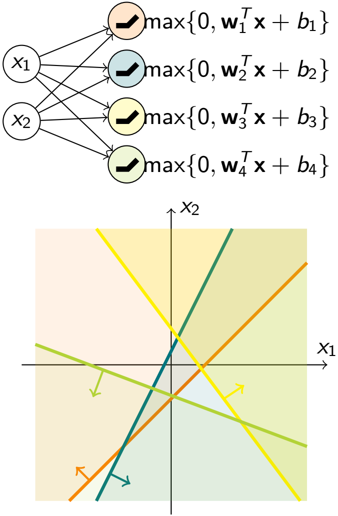
ReLU-Layer: Injectivity

ReLU-Layer Injectivity

Given: matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, vector $\mathbf{b} \in \mathbb{R}^{m}$

Question: is the map $\phi_{\mathbf{W},\mathbf{b}}$ injective?

Theorem [Puthawala et al., 2022] $\phi_{\mathbf{W},\mathbf{b}}$ injective $\iff$ for every cell, active neurons (rows of $\mathbf{W}$) have rank d



ReLU-Layer: Injectivity

ReLU-Layer Injectivity

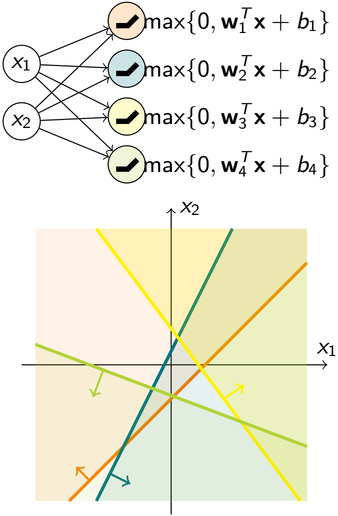
Given: matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, vector $\mathbf{b} \in \mathbb{R}^m$

Question: is the map $\phi_{\mathbf{W},\mathbf{b}}$ injective?

Theorem [Puthawala et al., 2022] $\phi_{\mathbf{W},\mathbf{b}}$ injective $\iff$ for every cell, active neurons (rows of $\mathbf{W}$) have rank d

Computational Complexity

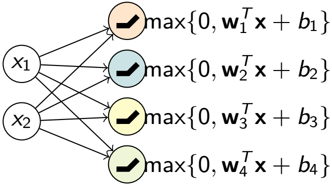
- number of cells in $O(\min\{2^m, m^d\})$ [Zaslavsky, 1975]



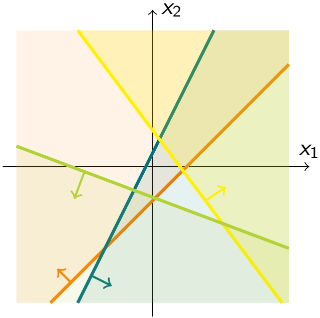
ReLU-Layer: Injectivity

ReLU-Layer Injectivity

Given: matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$, vector $\mathbf{b} \in \mathbb{R}^m$
Question: is the map $\phi_{\mathbf{W},\mathbf{b}}$ injective?



Theorem [Puthawala et al., 2022] $\phi_{\mathbf{W},\mathbf{b}}$ injective $\iff$ for every cell, active neurons (rows of $\mathbf{W}$) have rank d



Computational Complexity

- number of cells in $O(\min\{2^m, m^d\})$ [Zaslavsky, 1975]
- algorithm with runtime $O(\text{poly}(m) \min\{2^m, m^d\})$

Hardness-Result

Theorem [Froese,G.,Skutella]
ReLU-layer Injectivity is coNP-complete.

Hardness-Result

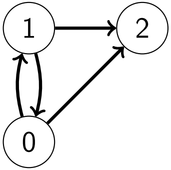
Theorem [Froese,G.,Skutella]
ReLU-layer Injectivity is coNP-complete.

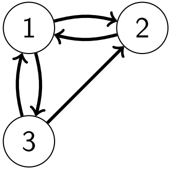
Proof.
Reduction from complement of:

Acyclic 2-disconnection

Input: A digraph $D = (V, A)$.

Question: Is there a subset $A' \subseteq A$ of arcs such that (V, A') is acyclic and $(V, A \setminus A')$ is not weakly connected?





□

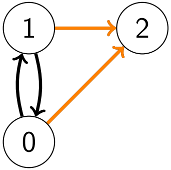
Hardness-Result

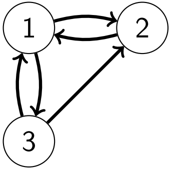
Theorem [Froese,G.,Skutella]
ReLU-layer Injectivity is coNP-complete.

Proof.
Reduction from complement of:

Acyclic 2-disconnection

Input: A digraph $D = (V, A)$.
Question: Is there a subset $A' \subseteq A$ of arcs such that (V, A') is acyclic and $(V, A \setminus A')$ is not weakly connected?





□

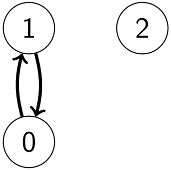
Hardness-Result

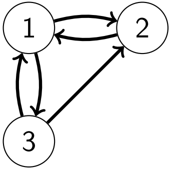
Theorem [Froese,G.,Skutella]
ReLU-layer Injectivity is coNP-complete.

Proof.
Reduction from complement of:

Acyclic 2-disconnection

Input: A digraph $D = (V, A)$.
Question: Is there a subset $A' \subseteq A$ of arcs such that (V, A') is acyclic and $(V, A \setminus A')$ is not weakly connected?





□

Hardness-Result

Theorem [Froese,G.,Skutella]
ReLU-layer Injectivity is coNP-complete.

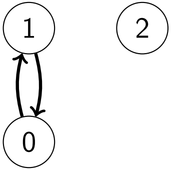
Proof.
Reduction from complement of:

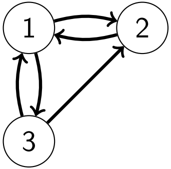
Acyclic 2-disconnection

Input: A digraph $D = (V, A)$.
Question: Is there a subset $A' \subseteq A$ of arcs such that (V, A') is acyclic and $(V, A \setminus A')$ is not weakly connected?

Theorem [Froese,G.,Skutella]
Acyclic 2-disconnection is NP-complete.

□





ReLU Layer from Digraph

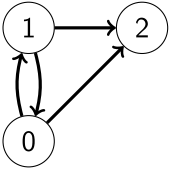
- Encode ACYCLIC 2-DISCONNECTION as RELU-LAYER INJECTIVITY

ReLU Layer from Digraph

- Encode Acyclic 2-disconnection as ReLU-layer Injectivity
- Given a digraph $D = (V, A)$ with $V = \{0, \ldots, d\}$ and $m = |A|$, we construct the following ReLU-layer $\phi \colon \mathbb{R}^d \to \mathbb{R}^m$ by

$$\phi(\mathbf{x}) = (\max\{0, \mathbf{x}_i - \mathbf{x}_j\})_{(i,j) \in A},$$

where $\mathbf{x}_0 := 0$.

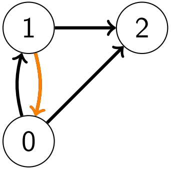


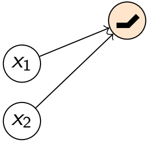
ReLU Layer from Digraph

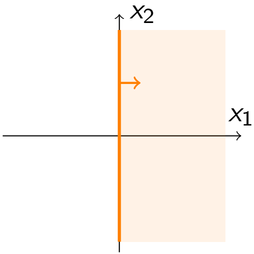
- Encode Acyclic 2-disconnection as ReLU-layer Injectivity
- Given a digraph $D = (V, A)$ with $V = \{0, \ldots, d\}$ and $m = |A|$, we construct the following ReLU-layer $\phi\colon \mathbb{R}^d \to \mathbb{R}^m$ by

$$\phi(\mathbf{x}) = (\max\{0, \mathbf{x}_i - \mathbf{x}_j\})_{(i,j) \in A},$$

where $\mathbf{x}_0 := 0$.





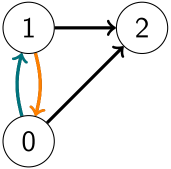


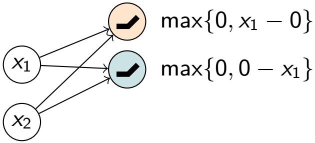
ReLU Layer from Digraph

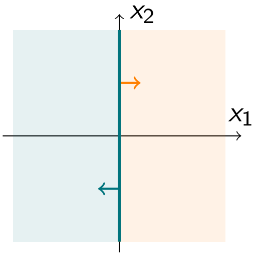
- Encode Acyclic 2-disconnection as ReLU-layer Injectivity
- Given a digraph $D = (V, A)$ with $V = \{0, \dots, d\}$ and $m = |A|$, we construct the following ReLU-layer $\phi\colon \mathbb{R}^d \to \mathbb{R}^m$ by

$$\phi(\mathbf{x}) = (\max\{0, \mathbf{x}_i - \mathbf{x}_j\})_{(i,j)\in A},$$

where $\mathbf{x}_0 := 0$.





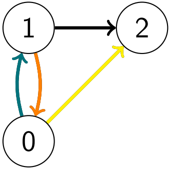


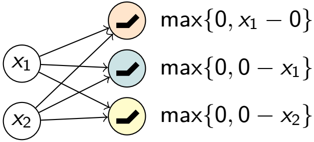
ReLU Layer from Digraph

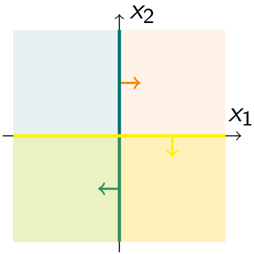
- Encode Acyclic 2-disconnection as ReLU-layer Injectivity
- Given a digraph $D = (V, A)$ with $V = \{0, \dots, d\}$ and $m = |A|$, we construct the following ReLU-layer $\phi\colon \mathbb{R}^d \to \mathbb{R}^m$ by

$$\phi(\mathbf{x}) = (\max\{0, \mathbf{x}_i - \mathbf{x}_j\})_{(i,j)\in A},$$

where $\mathbf{x}_0 := 0$.





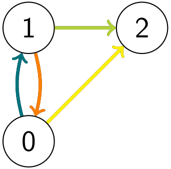


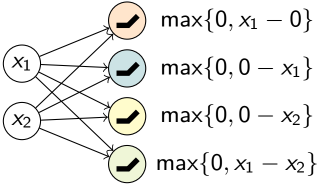
ReLU Layer from Digraph

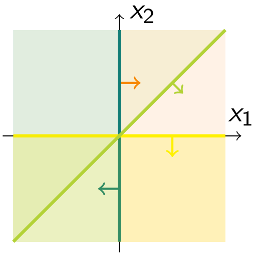
- Encode Acyclic 2-disconnection as ReLU-layer Injectivity
- Given a digraph $D = (V, A)$ with $V = \{0, \dots, d\}$ and $m = |A|$, we construct the following ReLU-layer $\phi \colon \mathbb{R}^d \to \mathbb{R}^m$ by

$$\phi(\mathbf{x}) = (\max\{0, \mathbf{x}_i - \mathbf{x}_j\})_{(i,j) \in A},$$

where $\mathbf{x}_0 := 0$.





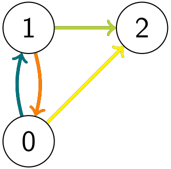


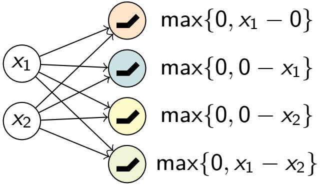
ReLU Layer from Digraph

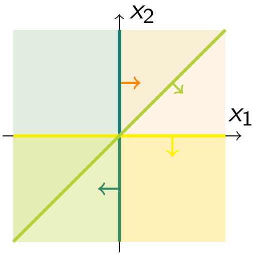
- Encode Acyclic 2-disconnection as ReLU-layer Injectivity
- Given a digraph $D = (V, A)$ with $V = \{0, \ldots, d\}$ and $m = |A|$, we construct the following ReLU-layer $\phi \colon \mathbb{R}^d \to \mathbb{R}^m$ by

$$\phi(\mathbf{x}) = (\max\{0, \mathbf{x}_i - \mathbf{x}_j\})_{(i,j) \in A},$$

  where $\mathbf{x}_0 := 0$.



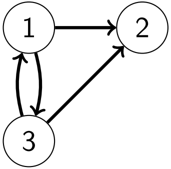


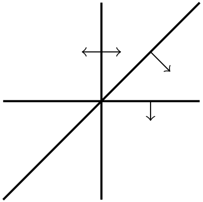


- All hyperplanes are of the form $\{\mathbf{x}_i = \mathbf{x}_j\}$ for $i, j \in [d+1]$

Full Rank = Weakly Connected

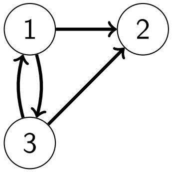
- For all $\pi \in \mathcal{S}_{d+1}$ it holds that ϕ is linear on $C_\pi = \{\mathbf{x}_{\pi(1)} \leq \ldots \leq \mathbf{x}_{\pi(d+1)}\}$ where $\mathbf{x}_{d+1} := 0$.

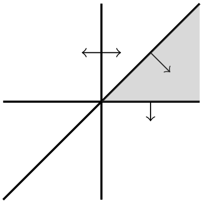




Full Rank = Weakly Connected

- For all $\pi \in \mathcal{S}_{d+1}$ it holds that ϕ is linear on $C_\pi = \{\mathbf{x}_{\pi(1)} \leq \ldots \leq \mathbf{x}_{\pi(d+1)}\}$ where $\mathbf{x}_{d+1} := 0$.
- ϕ is injective if and only if $\mathbf{W}_{C_\pi}$ has full rank for all $\pi \in \mathcal{S}_{d+1}$

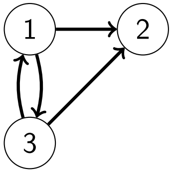


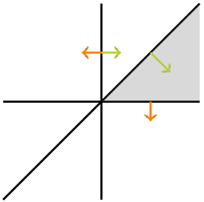


$0 \leq x_2 \leq x_1$

Full Rank = Weakly Connected

- For all $\pi \in \mathcal{S}_{d+1}$ it holds that ϕ is linear on $C_\pi = \{\mathbf{x}_{\pi(1)} \leq \ldots \leq \mathbf{x}_{\pi(d+1)}\}$ where $\mathbf{x}_{d+1} := 0$.
- ϕ is injective if and only if $\mathbf{W}_{C_\pi}$ has full rank for all $\pi \in \mathcal{S}_{d+1}$
- A neuron corresponding to the arc (i, j) is active on C_π if and only if $\pi(i) > \pi(j)$

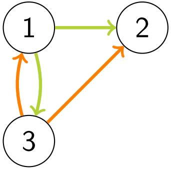


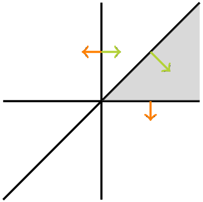


$0 \leq x_2 \leq x_1$

Full Rank = Weakly Connected

- For all $\pi \in \mathcal{S}_{d+1}$ it holds that ϕ is linear on $C_\pi = \{\mathbf{x}_{\pi(1)} \leq \ldots \leq \mathbf{x}_{\pi(d+1)}\}$ where $\mathbf{x}_{d+1} := 0$.
- ϕ is injective if and only if $\mathbf{W}_{C_\pi}$ has full rank for all $\pi \in \mathcal{S}_{d+1}$
- A neuron corresponding to the arc (i, j) is active on C_π if and only if $\pi(i) > \pi(j)$
- Let $A_\pi = \{(i,j) \in A \mid \pi(i) < \pi(j)\} \subseteq A$ be the (acyclic) set of arcs corresponding to the inactive neurons on C_π and $A \setminus A_\pi$ the set of arcs corresponding to the active neurons on C_π

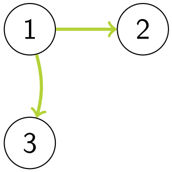


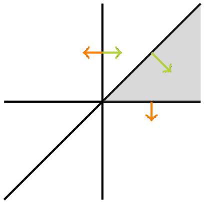


$0 \leq x_2 \leq x_1$

Full Rank = Weakly Connected

- For all $\pi \in \mathcal{S}_{d+1}$ it holds that ϕ is linear on $C_\pi = \{\mathbf{x}_{\pi(1)} \leq \ldots \leq \mathbf{x}_{\pi(d+1)}\}$ where $\mathbf{x}_{d+1} := 0$.
- ϕ is injective if and only if $\mathbf{W}_{C_\pi}$ has full rank for all $\pi \in \mathcal{S}_{d+1}$
- A neuron corresponding to the arc (i,j) is active on C_π if and only if $\pi(i) > \pi(j)$
- Let $A_\pi = \{(i,j) \in A \mid \pi(i) < \pi(j)\} \subseteq A$ be the (acyclic) set of arcs corresponding to the inactive neurons on C_π and $A \setminus A_\pi$ the set of arcs corresponding to the active neurons on C_π

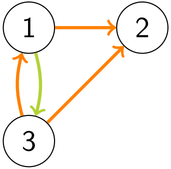


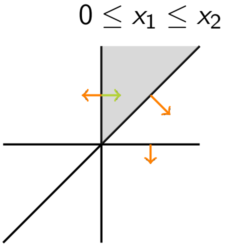


$0 \leq x_2 \leq x_1$

Full Rank = Weakly Connected

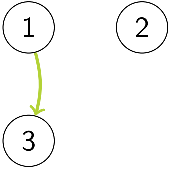
- For all $\pi \in \mathcal{S}_{d+1}$ it holds that ϕ is linear on $C_\pi = \{\mathbf{x}_{\pi(1)} \leq \ldots \leq \mathbf{x}_{\pi(d+1)}\}$ where $\mathbf{x}_{d+1} := 0$.
- ϕ is injective if and only if $\mathbf{W}_{C_\pi}$ has full rank for all $\pi \in \mathcal{S}_{d+1}$
- A neuron corresponding to the arc (i,j) is active on C_π if and only if $\pi(i) > \pi(j)$
- Let $A_\pi = \{(i,j) \in A \mid \pi(i) < \pi(j)\} \subseteq A$ be the (acyclic) set of arcs corresponding to the inactive neurons on C_π and $A \setminus A_\pi$ the set of arcs corresponding to the active neurons on C_π



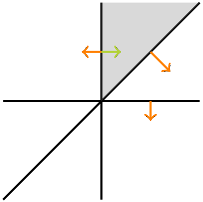


Full Rank = Weakly Connected

- For all $\pi \in \mathcal{S}_{d+1}$ it holds that ϕ is linear on $C_\pi = \{\mathbf{x}_{\pi(1)} \leq \ldots \leq \mathbf{x}_{\pi(d+1)}\}$ where $\mathbf{x}_{d+1} := 0$.
- ϕ is injective if and only if $\mathbf{W}_{C_\pi}$ has full rank for all $\pi \in \mathcal{S}_{d+1}$
- A neuron corresponding to the arc (i, j) is active on C_π if and only if $\pi(i) > \pi(j)$
- Let $A_\pi = \{(i, j) \in A \mid \pi(i) < \pi(j)\} \subseteq A$ be the (acyclic) set of arcs corresponding to the inactive neurons on C_π and $A \setminus A_\pi$ the set of arcs corresponding to the active neurons on C_π

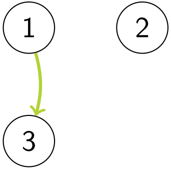


$0 \leq x_1 \leq x_2$

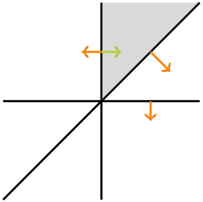


Full Rank = Weakly Connected

- For all $\pi \in \mathcal{S}_{d+1}$ it holds that ϕ is linear on $C_\pi = \{\mathbf{x}_{\pi(1)} \leq \ldots \leq \mathbf{x}_{\pi(d+1)}\}$ where $\mathbf{x}_{d+1} := 0$.
- ϕ is injective if and only if $\mathbf{W}_{C_\pi}$ has full rank for all $\pi \in \mathcal{S}_{d+1}$
- A neuron corresponding to the arc (i, j) is active on C_π if and only if $\pi(i) > \pi(j)$
- Let $A_\pi = \{(i, j) \in A \mid \pi(i) < \pi(j)\} \subseteq A$ be the (acyclic) set of arcs corresponding to the inactive neurons on C_π and $A \setminus A_\pi$ the set of arcs corresponding to the active neurons on C_π
- $D_\pi = (V, A \setminus A_\pi)$ is weakly connected $\iff$ $\mathbf{W}_{C_\pi}$ has full rank.



$0 \leq x_1 \leq x_2$



Correctness of Reduction

Not Injective $\iff$ $\exists$ Acyclic Disconnection

- ϕ is not injective $\iff$

Correctness of Reduction

Not Injective $\iff$ $\exists$ Acyclic Disconnection

- ϕ is not injective $\iff$
- there is $\pi \in \mathcal{S}_{d+1}$ such that $\mathbf{W}_{C_\pi}$ does not have full rank $\iff$

Correctness of Reduction

Not Injective $\iff \exists$ Acyclic Disconnection

- ϕ is not injective $\iff$
- there is $\pi \in \mathcal{S}_{d+1}$ such that $\mathbf{W}_{C_\pi}$ does not have full rank $\iff$
- there is A_π such that (V, A_π) is acyclic and $(V, A \setminus A_\pi)$ is not weakly connected.

Correctness of Reduction

Not Injective $\iff$ $\exists$ Acyclic Disconnection

- ϕ is not injective $\iff$
- there is $\pi \in \mathcal{S}_{d+1}$ such that $\mathbf{W}_{C_\pi}$ does not have full rank $\iff$
- there is A_π such that (V, A_π) is acyclic and $(V, A \setminus A_\pi)$ is not weakly connected.
- there is $A' \subseteq A$ such that (V, A') is acyclic and $(V, A \setminus A')$ is not weakly connected.

An FPT-Algorithm

- Can we do better than $O(m^d)$?

An FPT-Algorithm

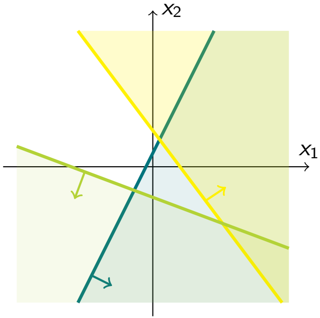
- Can we do better than $O(m^d)$?
- Yes!

Theorem [Froese, G., Skutella. 2024]

ReLU-Layer Injectivity can be solved in $O\big(\text{poly}(m)(d+1)^d\big)$ time (i.e., FPT for dimension d).

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

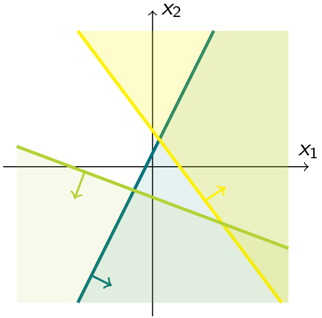
Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

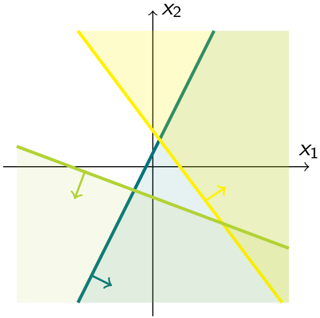
Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists

Simpler: does every cell have at least one active neuron?



FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists

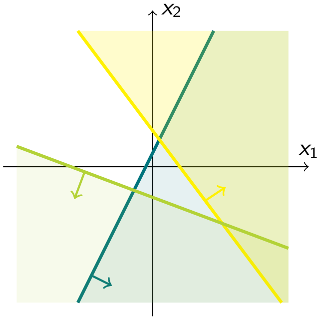


Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



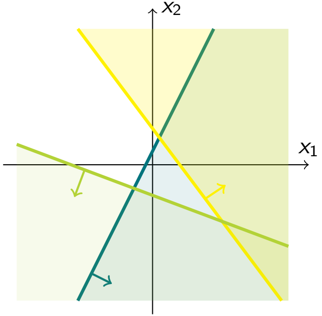
Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

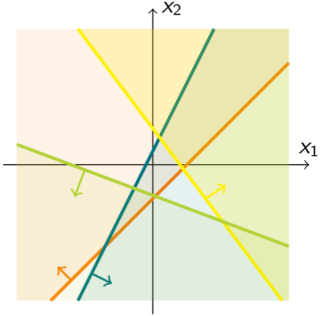
– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

$\rightarrow$ efficiently solvable via LP

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

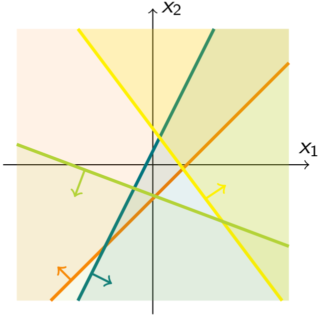
– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

$\rightarrow$ efficiently solvable via LP

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

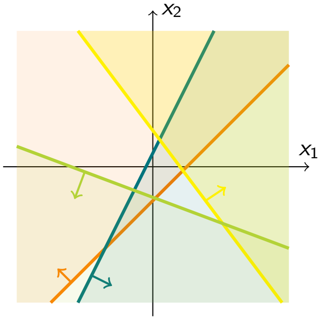
$\rightarrow$ efficiently solvable via LP

Theorem [Helly 1913; Radon 1921; Kőnig 1922]

Let $X_1, \ldots, X_m \subseteq \mathbb{R}^d$ convex with $\bigcap_{i=1}^m X_i = \emptyset$.

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

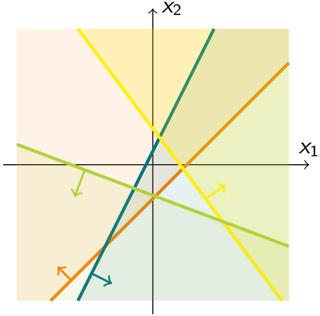
$\rightarrow$ efficiently solvable via LP

Theorem [Helly 1913; Radon 1921; Kőnig 1922]

Let $X_1, \ldots, X_m \subseteq \mathbb{R}^d$ convex with $\bigcap_{i=1}^m X_i = \emptyset$. Then: there is $I \subseteq \{1, \ldots, m\}$ with $|I| \leq d+1$ and $\bigcap_{i \in I} X_i = \emptyset$.

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

$\rightarrow$ efficiently solvable via LP

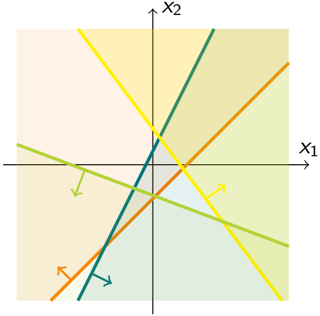
Theorem [Helly 1913; Radon 1921; Kőnig 1922]

Let $X_1, \ldots, X_m \subseteq \mathbb{R}^d$ convex with $\bigcap_{i=1}^m X_i = \emptyset$. Then: there is $I \subseteq \{1, \ldots, m\}$ with $|I| \leq d+1$ and $\bigcap_{i \in I} X_i = \emptyset$.

Algorithmic Idea:

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

$\rightarrow$ efficiently solvable via LP

Theorem [Helly 1913; Radon 1921; Kőnig 1922]

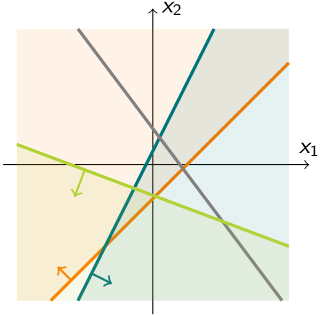
Let $X_1, \ldots, X_m \subseteq \mathbb{R}^d$ convex with $\bigcap_{i=1}^m X_i = \emptyset$. Then: there is $I \subseteq \{1, \ldots, m\}$ with $|I| \leq d+1$ and $\bigcap_{i \in I} X_i = \emptyset$.

Algorithmic Idea:

- find (at most) $d+1$ neurons whose halfspaces cover $\mathbb{R}^d$ (if not, $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective)

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

$\rightarrow$ efficiently solvable via LP

Theorem [Helly 1913; Radon 1921; Kőnig 1922]

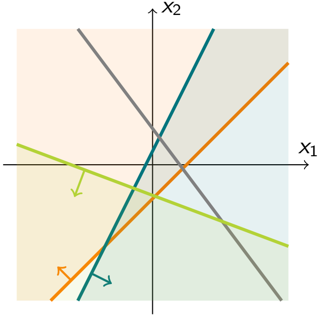
Let $X_1, \ldots, X_m \subseteq \mathbb{R}^d$ convex with $\bigcap_{i=1}^m X_i = \emptyset$. Then: there is $I \subseteq \{1, \ldots, m\}$ with $|I| \leq d+1$ and $\bigcap_{i \in I} X_i = \emptyset$.

Algorithmic Idea:

- find (at most) $d+1$ neurons whose halfspaces cover $\mathbb{R}^d$ (if not, $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective)

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

$\rightarrow$ efficiently solvable via LP

Theorem [Helly 1913; Radon 1921; Kőnig 1922]

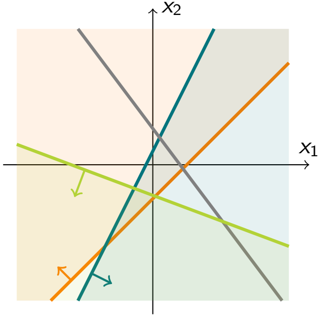
Let $X_1, \ldots, X_m \subseteq \mathbb{R}^d$ convex with $\bigcap_{i=1}^m X_i = \emptyset$. Then: there is $I \subseteq \{1, \ldots, m\}$ with $|I| \leq d+1$ and $\bigcap_{i \in I} X_i = \emptyset$.

Algorithmic Idea:

- find (at most) $d+1$ neurons whose halfspaces cover $\mathbb{R}^d$ (if not, $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective)
- branch over the $d+1$ halfspaces; always discard linearly dependent neurons

FPT $O(\text{poly}(m)(d+1)^d)$ Algorithm for ReLU-Layer Injectivity

Task: Find cell whose active neurons have rank $< d$ or decide that no such cell exists



Simpler: does every cell have at least one active neuron?

– if not, then $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective

Equivalent: is intersection of halfspace-complements empty?

$\rightarrow$ efficiently solvable via LP

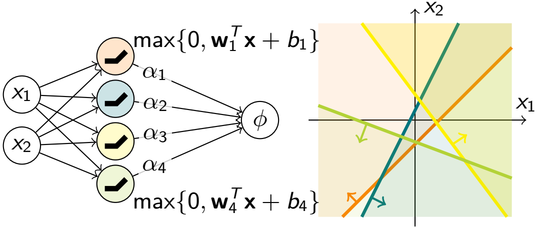
Theorem [Helly 1913; Radon 1921; Kőnig 1922]

Let $X_1, \ldots, X_m \subseteq \mathbb{R}^d$ convex with $\bigcap_{i=1}^m X_i = \emptyset$. Then: there is $I \subseteq \{1, \ldots, m\}$ with $|I| \leq d+1$ and $\bigcap_{i \in I} X_i = \emptyset$.

Algorithmic Idea:

- find (at most) $d+1$ neurons whose halfspaces cover $\mathbb{R}^d$ (if not, $\phi_{\mathbf{W},\mathbf{b}}$ is *not* injective)
- branch over the $d+1$ halfspaces; always discard linearly dependent neurons
- branching strictly increases rank of active neurons $\rightarrow$ depth of branching at most d

ReLU-Layer: Surjectivity and Positivity

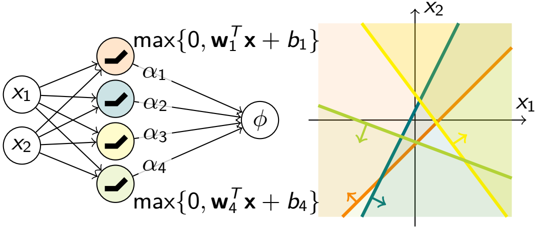


ReLU-Layer: Surjectivity and Positivity

Question: is map $\phi : \mathbb{R}^d \to \mathbb{R}$ with

$$\phi(\mathbf{x}) := \sum_{i=1}^{m} \alpha_i \max\{0, \mathbf{w}_i^T \mathbf{x} + b_i\}$$

surjective?



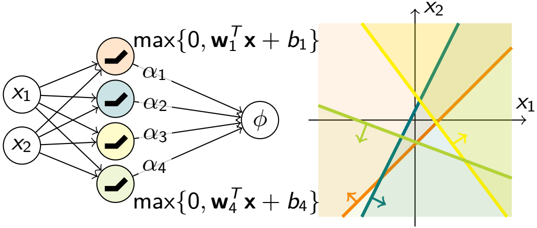
ReLU-Layer: Surjectivity and Positivity

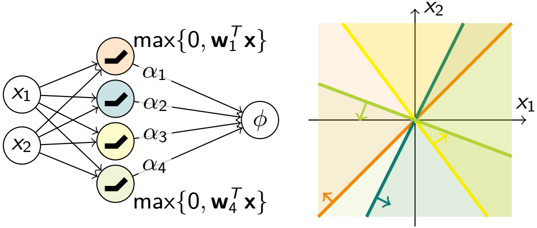
Question: is map $\phi : \mathbb{R}^d \to \mathbb{R}$ with

$$\phi(\mathbf{x}) := \sum_{i=1}^{m} \alpha_i \max\{0, \mathbf{w}_i^T \mathbf{x} + b_i\}$$

surjective?

Observation 1. W.l.o.g. $\mathbf{b} = \mathbf{0}$
$\to \phi$ is positively homogeneous,
i.e., $\phi(\lambda \mathbf{x}) = \lambda \phi(\mathbf{x})$ for all $\lambda > 0$.





ReLU-Layer: Surjectivity and Positivity

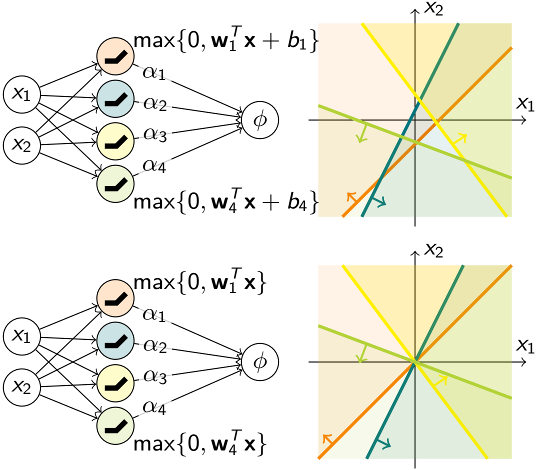
Question: is map $\phi : \mathbb{R}^d \to \mathbb{R}$ with

$$\phi(\mathbf{x}) := \sum_{i=1}^{m} \alpha_i \max\{0, \mathbf{w}_i^T \mathbf{x} + b_i\}$$

surjective?

Observation 1. W.l.o.g. $\mathbf{b} = \mathbf{0}$
$\to \phi$ is positively homogeneous,
i.e., $\phi(\lambda \mathbf{x}) = \lambda \phi(\mathbf{x})$ for all $\lambda > 0$.

Observation 2. The problem is equivalent to:
$\exists \mathbf{x} \in \mathbb{R}^d : \phi(\mathbf{x}) > 0\,?$



NP-Completeness of 2-LAYER RELU-POSITIVITY

Theorem [Froese,G.,Skutella]
Deciding if there exists $\mathbf{x} \in \mathbb{R}^d$ such that $\phi(\mathbf{x}) > 0$ is NP-complete.

NP-Completeness of 2-Layer ReLU-Positivity

Theorem [Froese,G.,Skutella]
Deciding if there exists $\mathbf{x} \in \mathbb{R}^d$ such that $\phi(\mathbf{x}) > 0$ is NP-complete.

Proof.
Reduction from NP-complete problem:

Positive Cut

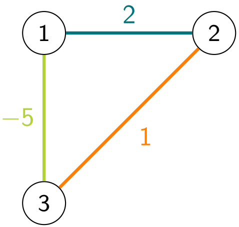
Input: A graph $G = (V = \{1, \dots n\}, E)$ and a weight function $c\colon E \to \mathbb{Z}$.
Question: Is there a subset $S \subseteq V$ such that $\delta(S) := \sum_{\substack{\{u,v\} \in E \\ v \in S, u \notin S}} c(u, v) > 0$?

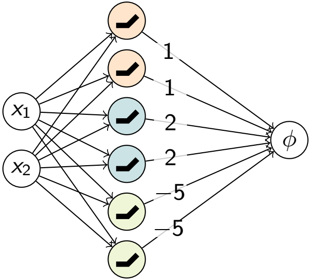
□

Proof idea

- Define the 2-layer ReLU neural network $\phi\colon \mathbb{R}^n \to \mathbb{R}$ with

$$\phi(\mathbf{x}) = \sum_{\{i,j\}\in E} c(\{i,j\}) \cdot ([\mathbf{x}_i - \mathbf{x}_j]_+ + [\mathbf{x}_j - \mathbf{x}_i]_+)$$

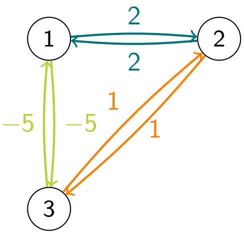


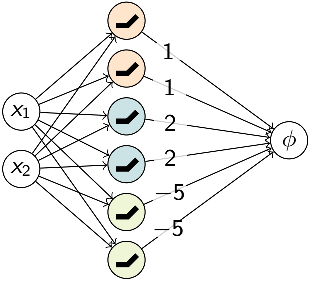


Proof idea

- Define the 2-layer ReLU neural network $\phi\colon \mathbb{R}^n \to \mathbb{R}$ with

$$\phi(\mathbf{x}) = \sum_{\{i,j\}\in E} c(\{i,j\}) \cdot ([\mathbf{x}_i - \mathbf{x}_j]_+ + [\mathbf{x}_j - \mathbf{x}_i]_+)$$





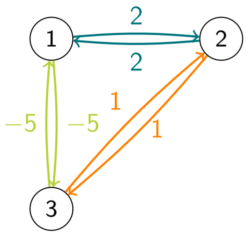
Proof idea

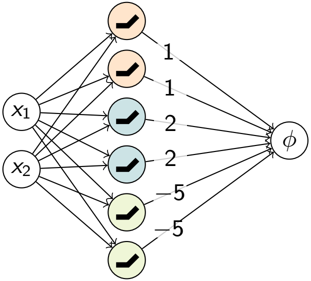
- Define the 2-layer ReLU neural network $\phi\colon \mathbb{R}^n \to \mathbb{R}$ with

$$\phi(\mathbf{x}) = \sum_{\{i,j\}\in E} c(\{i,j\}) \cdot ([\mathbf{x}_i - \mathbf{x}_j]_+ + [\mathbf{x}_j - \mathbf{x}_i]_+)$$

- For a subset $S \subseteq V$, let $\mathbb{1}_S := \sum_{i\in S} \mathbf{e}_i \in \mathbb{R}^n$, it holds that

$$\phi(\mathbb{1}_S) = \delta(S) = \sum_{u\in S, v\notin S} c(\{u,v\})$$





Proof idea

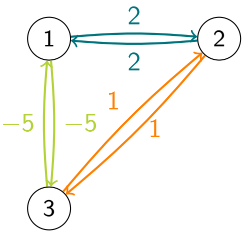
- Define the 2-layer ReLU neural network $\phi\colon \mathbb{R}^n \to \mathbb{R}$ with

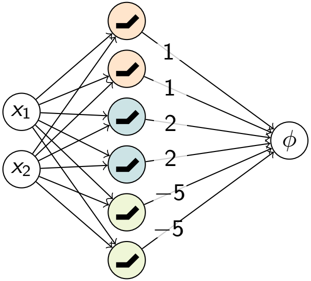
$$\phi(\mathbf{x}) = \sum_{\{i,j\}\in E} c(\{i,j\}) \cdot ([\mathbf{x}_i - \mathbf{x}_j]_+ + [\mathbf{x}_j - \mathbf{x}_i]_+)$$

- For a subset $S \subseteq V$, let $\mathbb{1}_S := \sum_{i\in S} \mathbf{e}_i \in \mathbb{R}^n$, it holds that

$$\phi(\mathbb{1}_S) = \delta(S) = \sum_{u\in S, v\notin S} c(\{u,v\})$$

- ϕ attains a positive value if and only if δ does





Proof idea

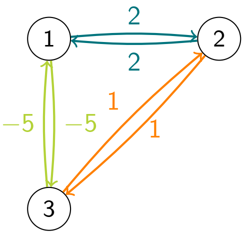
- Define the 2-layer ReLU neural network $\phi \colon \mathbb{R}^n \to \mathbb{R}$ with

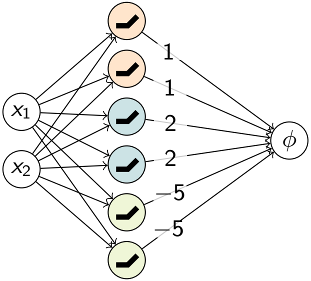
$$\phi(\mathbf{x}) = \sum_{\{i,j\} \in E} c(\{i,j\}) \cdot ([\mathbf{x}_i - \mathbf{x}_j]_+ + [\mathbf{x}_j - \mathbf{x}_i]_+)$$

- For a subset $S \subseteq V$, let $\mathbb{1}_S := \sum_{i \in S} \mathbf{e}_i \in \mathbb{R}^n$, it holds that

$$\phi(\mathbb{1}_S) = \delta(S) = \sum_{u \in S, v \notin S} c(\{u,v\})$$

- ϕ attains a positive value if and only if δ does
- Remark: ϕ is Lovász extension of $\delta \colon 2^V \to \mathbb{Z}$.





Translation to Polytopal World

Question: $\exists\, \mathbf{x} \in \mathbb{R}^d : \ \phi(\mathbf{x}) \ := \ \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} > 0\,?$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d : \ \phi(\mathbf{x}) := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} > 0\,?$

$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\}$

$\mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d : \ \phi(\mathbf{x}) := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} > 0\,?$

$$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\} \quad \xrightarrow{\psi} \quad \mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$$

$$\psi\left(\max_{i \in I}\left\{\mathbf{a}_i^T \mathbf{x}\right\}\right) := \operatorname{conv}\left\{\mathbf{a}_i \mid i \in I\right\}$$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d : \ \phi(\mathbf{x}) \ := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} > 0\,?$

$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\} \xrightarrow{\psi} \mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$

$\to$ semi-ring $(\mathcal{F}_d, \max, +)$

$\to$ semi-ring $(\mathcal{P}_d, \text{conv}, +)$

semi-ring isomorphism: $\psi\left(\max_{i \in I}\{\mathbf{a}_i^T \mathbf{x}\}\right) := \text{conv}\,\{\mathbf{a}_i \mid i \in I\}$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d : \ \phi(\mathbf{x}) := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} > 0\,?$

$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\}$ $\xrightarrow{\psi}$ $\mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$

$\to$ semi-ring $(\mathcal{F}_d, \max, +)$

$\to$ semi-ring $(\mathcal{P}_d, \mathrm{conv}, +)$

semi-ring isomorphism: $\psi\left(\max_{i \in I}\{\mathbf{a}_i^T \mathbf{x}\}\right) := \mathrm{conv}\,\{\mathbf{a}_i \mid i \in I\}$

$\psi\left(\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}\right)$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d : \ \phi(\mathbf{x}) \ := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} > 0\,?$

$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\} \quad \xrightarrow{\psi} \quad \mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$

$\to$ semi-ring $(\mathcal{F}_d, \max, +)$

$\to$ semi-ring $(\mathcal{P}_d, \operatorname{conv}, +)$

semi-ring isomorphism: $\psi\left(\max_{i \in I}\{\mathbf{a}_i^T \mathbf{x}\}\right) := \operatorname{conv}\{\mathbf{a}_i \mid i \in I\}$

$$\psi\Big(\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}\Big) = \sum_{i \in I^+} \psi\left(\max\{0, \mathbf{w}_i^T \mathbf{x}\}\right)$$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d: \ \phi(\mathbf{x}) := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} > 0\,?$

$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\} \quad \xrightarrow{\psi} \quad \mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$

$\to$ semi-ring $(\mathcal{F}_d, \max, +)$

$\to$ semi-ring $(\mathcal{P}_d, \mathrm{conv}, +)$

semi-ring isomorphism: $\psi\left(\max_{i \in I}\{\mathbf{a}_i^T \mathbf{x}\}\right) := \mathrm{conv}\,\{\mathbf{a}_i \mid i \in I\}$

$$\psi\Big(\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}\Big) = \sum_{i \in I^+} \psi\left(\max\{0, \mathbf{w}_i^T \mathbf{x}\}\right) = \sum_{i \in I^+} \mathrm{conv}\{\mathbf{0}, \mathbf{w}_i\}$$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d: \ \phi(\mathbf{x}) := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\text{convex, pos. homogeneous}} > 0\,?$

$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\} \xrightarrow{\psi} \mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$

$\to$ semi-ring $(\mathcal{F}_d, \max, +)$

$\to$ semi-ring $(\mathcal{P}_d, \text{conv}, +)$

semi-ring isomorphism: $\psi\left(\max_{i \in I}\{\mathbf{a}_i^T \mathbf{x}\}\right) := \text{conv}\{\mathbf{a}_i \mid i \in I\}$

$$\psi\Big(\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}\Big) = \sum_{i \in I^+} \psi\left(\max\{0, \mathbf{w}_i^T \mathbf{x}\}\right) = \sum_{i \in I^+} \text{conv}\{\mathbf{0}, \mathbf{w}_i\} =: Z^+ \text{(zonotope)}$$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d : \ \phi(\mathbf{x}) := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\mapsto Z^+} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\mapsto Z^-} > 0\,?$

$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\} \quad \xrightarrow{\psi} \quad \mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$

$\to$ semi-ring $(\mathcal{F}_d, \max, +)$

$\to$ semi-ring $(\mathcal{P}_d, \text{conv}, +)$

semi-ring isomorphism: $\psi\left(\max_{i \in I}\{\mathbf{a}_i^T \mathbf{x}\}\right) := \text{conv}\{\mathbf{a}_i \mid i \in I\}$

$$\psi\Big(\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}\Big) = \sum_{i \in I^+} \psi\left(\max\{0, \mathbf{w}_i^T \mathbf{x}\}\right) = \sum_{i \in I^+} \text{conv}\{\mathbf{0}, \mathbf{w}_i\} =: Z^+ (\text{zonotope})$$

Translation to Polytopal World

Question: $\exists \mathbf{x} \in \mathbb{R}^d : \ \phi(\mathbf{x}) := \underbrace{\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\mapsto Z^+} - \underbrace{\sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\}}_{\mapsto Z^-} > 0\,?$

$\mathcal{F}_d := \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ CPWL, convex, p.h.}\} \quad \xrightarrow{\psi} \quad \mathcal{P}_d := \{P \subseteq \mathbb{R}^d \mid P \text{ polytope}\}$

$\to$ semi-ring $(\mathcal{F}_d, \max, +)$

$\to$ semi-ring $(\mathcal{P}_d, \text{conv}, +)$

semi-ring isomorphism: $\psi\left(\max_{i \in I}\{\mathbf{a}_i^T \mathbf{x}\}\right) := \text{conv}\{\mathbf{a}_i \mid i \in I\}$

$$\psi\Big(\sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\}\Big) = \sum_{i \in I^+} \psi\left(\max\{0, \mathbf{w}_i^T \mathbf{x}\}\right) = \sum_{i \in I^+} \text{conv}\{\mathbf{0}, \mathbf{w}_i\} =: Z^+ (\text{zonotope})$$

Observation

$$\exists \mathbf{x} \in \mathbb{R}^d : \ \sum_{i \in I^+} \max\{0, \mathbf{w}_i^T \mathbf{x}\} - \sum_{i \in I^-} \max\{0, \mathbf{w}_i^T \mathbf{x}\} > 0 \quad \Longleftrightarrow \quad Z^+ \not\subseteq Z^-$$

ReLU-Layer Positivity equivalent to complement of Zonotope Containment

Conclusion

Summary

- Relation between cuts in (di-)graphs and elementary properties of ReLU layers via graphical hyperplane arrangements.

Conclusion

Summary

- Relation between cuts in (di-)graphs and elementary properties of ReLU layers via graphical hyperplane arrangements.
- Injectivity is coNP-complete.

Conclusion

Summary

- Relation between cuts in (di-)graphs and elementary properties of ReLU layers via graphical hyperplane arrangements.
- Injectivity is coNP-complete.
- Injectivity is fixed parameter tractable for d.

Conclusion

Summary

- Relation between cuts in (di-)graphs and elementary properties of ReLU layers via graphical hyperplane arrangements.
- Injectivity is coNP-complete.
- Injectivity is fixed parameter tractable for d.
- Surjectivity is polynomial equivalent to positivity and both are NP-complete.

Conclusion

Summary

- Relation between cuts in (di-)graphs and elementary properties of ReLU layers via graphical hyperplane arrangements.
- Injectivity is coNP-complete.
- Injectivity is fixed parameter tractable for d.
- Surjectivity is polynomial equivalent to positivity and both are NP-complete.
- Positivity is equivalent to (complement of) zonotope containment

Conclusion

Summary

- Relation between cuts in (di-)graphs and elementary properties of ReLU layers via graphical hyperplane arrangements.
- Injectivity is coNP-complete.
- Injectivity is fixed parameter tractable for d.
- Surjectivity is polynomial equivalent to positivity and both are NP-complete.
- Positivity is equivalent to (complement of) zonotope containment

Open Questions

- Is injectivity FPT for deep neural networks?

Conclusion

Summary

- Relation between cuts in (di-)graphs and elementary properties of ReLU layers via graphical hyperplane arrangements.
- Injectivity is coNP-complete.
- Injectivity is fixed parameter tractable for d.
- Surjectivity is polynomial equivalent to positivity and both are NP-complete.
- Positivity is equivalent to (complement of) zonotope containment

Open Questions

- Is injectivity FPT for deep neural networks?
- How can we characterize surjectivity for more than one output neuron?

Conclusion

Summary

- Relation between cuts in (di-)graphs and elementary properties of ReLU layers via graphical hyperplane arrangements.
- Injectivity is coNP-complete.
- Injectivity is fixed parameter tractable for d.
- Surjectivity is polynomial equivalent to positivity and both are NP-complete.
- Positivity is equivalent to (complement of) zonotope containment

Open Questions

- Is injectivity FPT for deep neural networks?
- How can we characterize surjectivity for more than one output neuron?
- Is zonotope containment FPT for parameter d??

Thank you for your attention!