Intelligent Robotics and Automation Laboratory
Computer Vision, Speech Communication & Signal Processing Group,
National Technical University of Athens, Greece (NTUA)
https://robotics.ntua.gr, http://cvsp.cs.ntua.gr

Tropical Geometry for Machine Learning and Optimization

Petros Maragos

slides: https://robotics.ntua.gr/icassp-2024-tutorial/

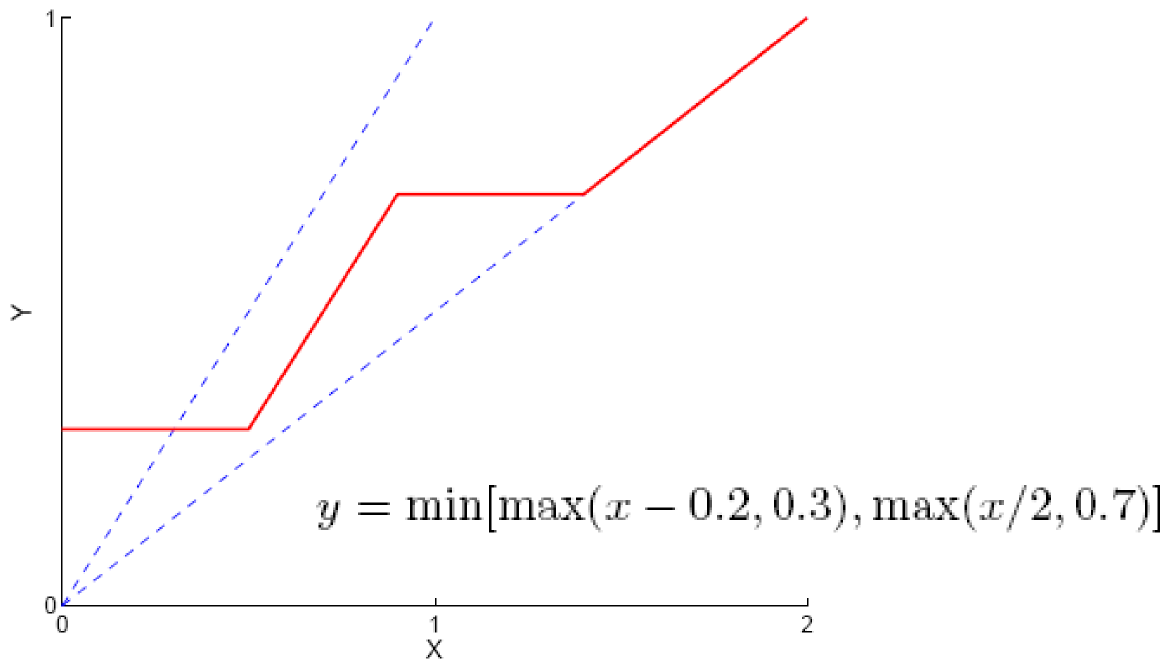
Tutorial at IEEE International Conference on Acoustics, Speech and Signal Processing 2024,
Seoul, Korea, 15 April 2024

What does TROPICAL mean?

- The adjective “tropical” was coined by French mathematicians Dominique Perrin and Jean-Eric Pin, to honor their Brazilian colleague Imre Simon, a pioneer of min-plus algebra as applied to finite automata in computer science.

- Tropical (**Τροπικός** in Greek) comes from the greek word «**Τροπή**» which means “turning” or “changing the way/direction”.

Polygonal lines



ICASSP 2024 KOREA

Tutorial Outline

- 1. Elements from Tropical Geometry and Max-Plus Algebra
- 2. Neural Networks with Piecewise-linear (PWL) Activations
- 3. Morphological (Max-plus) Neural Networks
- 4. Approximation Using Tropical Mappings
- 5. Piecewise-linear (PWL) Regression

This research work was supported by the Hellenic Foundation for Research & Innovation (H.F.R.I.) under the "2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers" (Project Number:2656, Acronym: TROGEMAL).

H.F.R.I. Hellenic Foundation for Research & Innovation

ICASSP 2024 KOREA

Collaborators

TG&ML: Petros Maragos, Vasilis Charisopoulos, Manos Theodosis



Neural Net Minimization:
Georgios Smyrnis, Panos Misiakos, George Retsinas, Nikos Dimitriadis




Tropical Approximation: Ioannis Kordonis

Tropical Sparsity: Anastasios Tsiamis, Nikos Tsilivis


Elements of Tropical Geometry

"*TG is a marriage between algebraic geometry and polyhedral geometry. A piecewise-linear version of algebraic geometry.*" [Maclagan & Sturmfels 2015]

Our view: TG is a "dequantized" version of Euclidean (coordinate-free) geometry and analytic geometry.

References on TG and its Applications to Machine Learning & Optimization

Books & Math Articles on Tropical Geometry (TG):

- D. Maclagan & B. Sturmfels, *Introduction to Tropical Geometry*, AMS 2015.
- I. Itenberg, G. Mikhalkin, and E. I. Shustin, *Tropical Algebraic Geometry,* Springer 2009.
- M. Joswig, *Essentials of Tropical Combinatorics*, AMS 2021.
- *Max-plus Convex Sets/Cones*: [Cuninghame-Green 1979; Butkovic 2007], [Litvinov, Maslov & Sphiz 2001], [Cohen, Gaubert & Quadrat 2004; Gaubert & Katz 2007; Allamigeon et al 2010]
- *Tropical Convexity, Tropical Halfspaces/Polyhedra*: [Maslov 1987], [Develin & Sturmfels 2004], [Joswig 2005], [Gaubert & Katz 2011]. *TG and Mean Payoff Games*: [Akian et al 2012; Akian et al 2021]
- O. Viro, *Dequantization of Real Algebraic Geometry on Logarithmic Paper*, ArXiv 2000.

Some Applications of TG to Machine Learning:

- L. Pachter & B. Sturmfels, *Tropical geometry of statistical models*, PNAS 2004.
- V.Charisopoulos & P.M., *Tropical Approach to Neural Nets with Piecewise Linear Activations*, ISMM2017, ArXiv2018.
- L. Zhang, G. Naitzat, L.-H. Lim, *Tropical Geometry of Deep Neural Networks*, ICML 2018.
- P.M., V. Charisopoulos & E. Theodosis, *Tropical Geometry and Machine Learning*, Proc. IEEE 2021.
- NTUA Group: P.M., Charisopoulos, Dimitriadis, Kordonis, Misiakos, Retsinas, Smyrnis, Theodosis, Tsiamis, Tsilivis
- + Other References in this tutorial.

Tropical Semirings

Scalar Arithmetic Rings

Integer/Real Addition & Multiplication Ring: $(\mathbb{R},+,\times)$, $(\mathbb{Z},+,\times)$

Tropical Semirings

$\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$, $\mathbb{R}_{\min} = \mathbb{R} \cup \{+\infty\}$

$\vee = \max$, $\wedge = \min$

Max-plus semiring: $(\mathbb{R}_{\max}, \vee, +)$

Min-plus semiring: $(\mathbb{R}_{\min}, \wedge, +)$

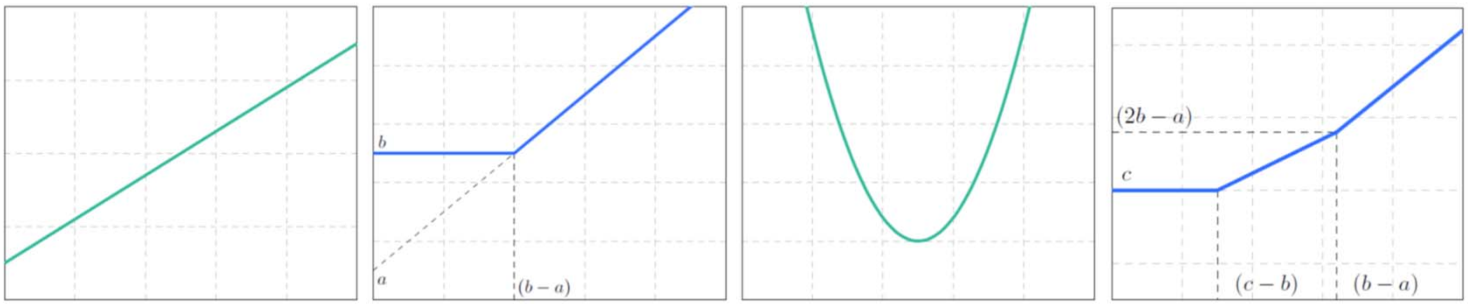
Correspondences between linear and (max, +) arithmetic

Linear arithmetic	(max, +) arithmetic
$+$	max
$\times$	$+$
0	$-\infty$
1	0
$x^{-1} = 1/x$	$x^{-1} = -x$

ICASSP 2024 KOREA

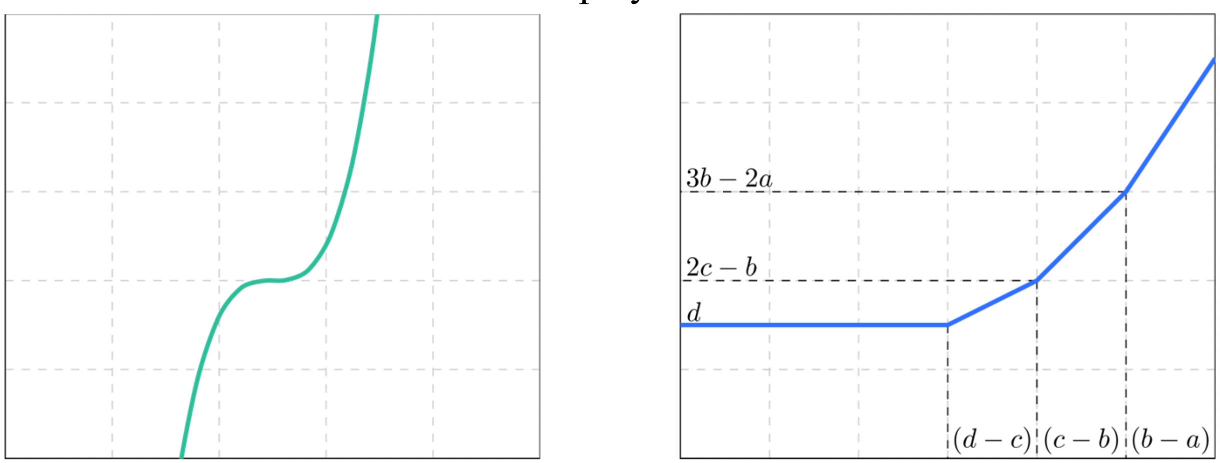
Graphs of Max-plus Tropical 1D Polynomials

$$y_{\text{t-line}} = \max(a + x, b), \quad y_{\text{t-parab}} = \max(a + 2x, b + x, c)$$



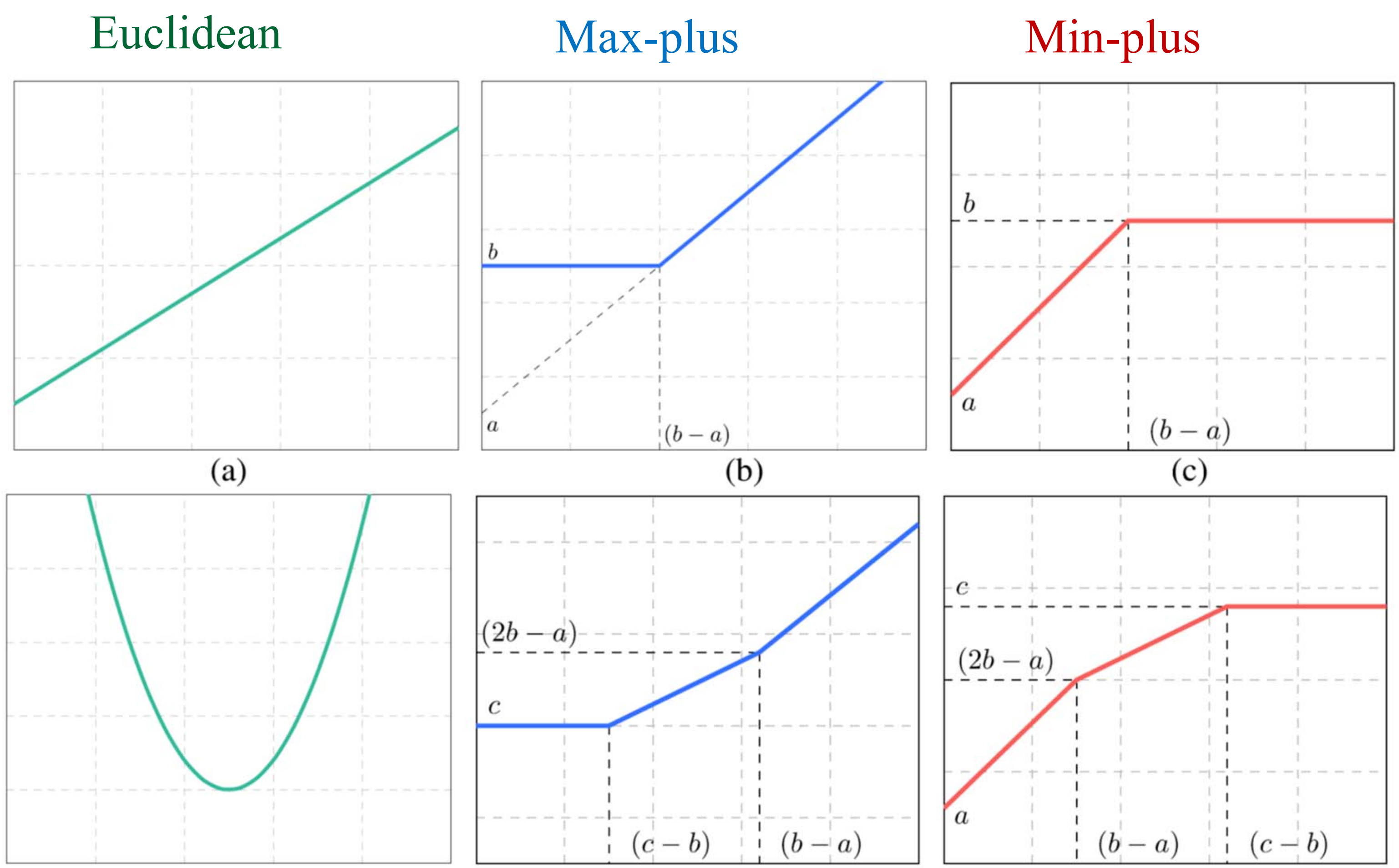
(a) Euclidean line (b) Tropical line (c) Euclid parabola (d) Tropic parabola

Cubic polynomial



ICASSP 2024 KOREA

Max-plus and Min-Plus Tropical 1D Polynomials



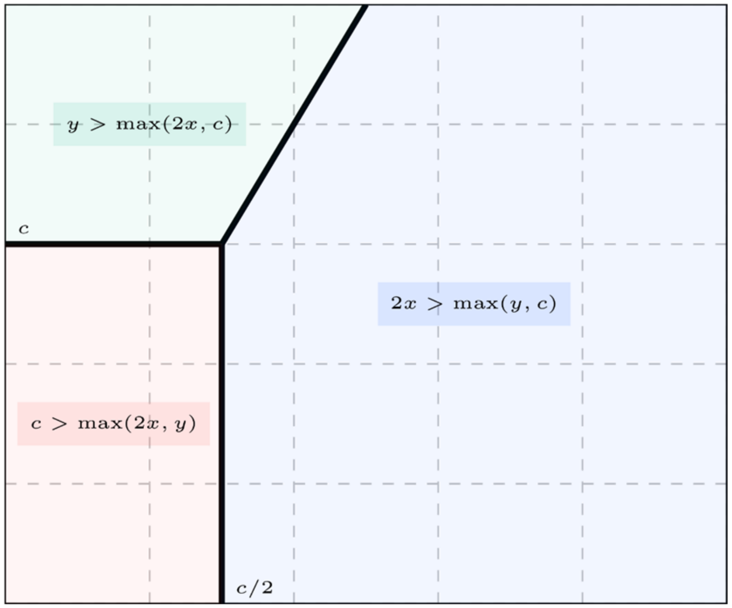
(a) (b) (c)

(d) (e) (f)

Tropical Curve of Max/Min-Polynomials

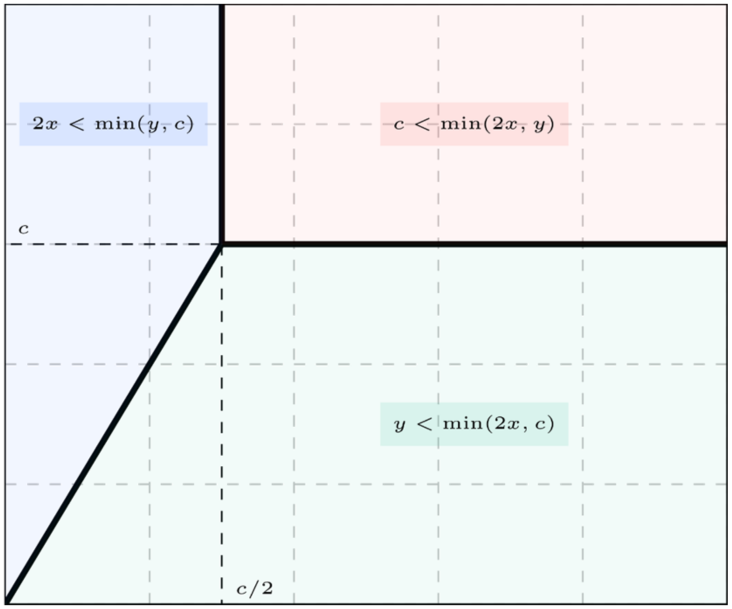
Tropical curve of $p(x,y)$ =

"Zero locus" of a max/min polynomial is the set of points where the max/min is attained by more than one of the "monomial" terms of the polynomial.



Tropical curve of the max-polynomial

$$p(x,y) = \max(2x, y, c)$$



Tropical curve of the min-polynomial

$$p'(x,y) = \min(2x, y, c)$$

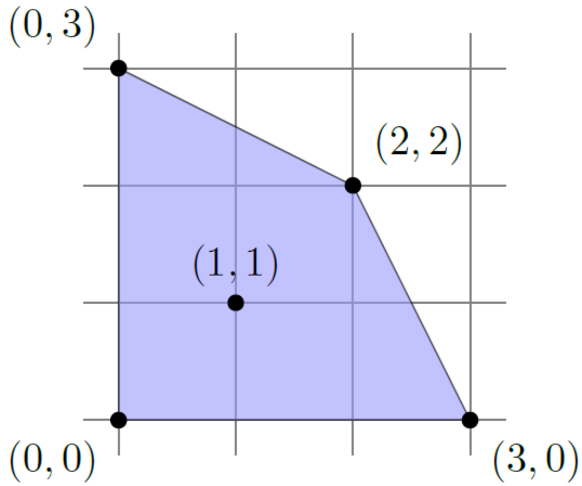
Newton Polytope of Tropical Polynomial

Max polynomial

$$p(\boldsymbol{x}) = \max_{i \in 1,2,\ldots,k} \{c_{i1}x_1 + c_{i2}x_2 + \cdots + c_{in}x_n\} = \bigvee_{i=1}^{k} \boldsymbol{c}_i^T \boldsymbol{x}$$

Newton polytope $N(p)$ of max polynomial p is the convex hull of its coefficients' vectors.

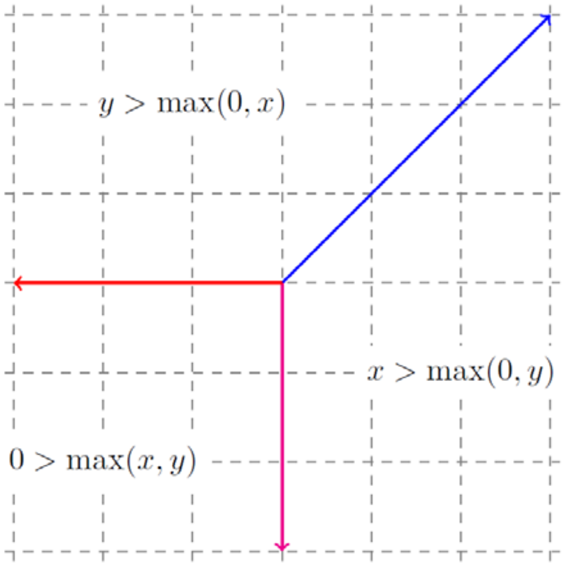
$$p(\boldsymbol{x}) = \max(0, x_1 + x_2, 2x_1 + 2x_2, 3x_1, 3x_2)$$



Tropical Curve vs Newton Polytope

Max polynomial: $p(x,y) = \max(x,y,0)$

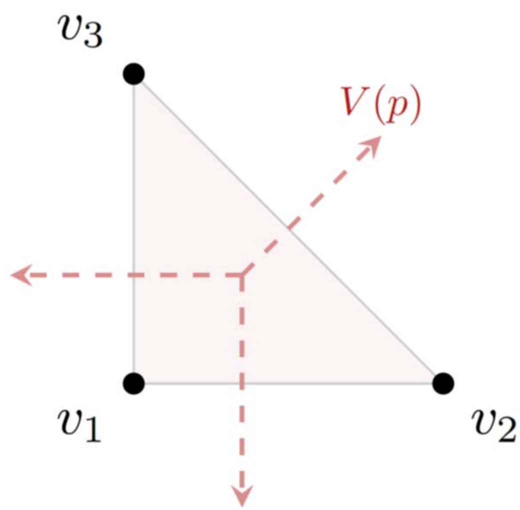
Tropical curve ("Zero locus") $V(p)$ of a max polynomial p is the set of points where the max is attained by more than two polynomial terms.



Tropical curve $V(p)$
of $p(x,y) = \max(x,y,0)$

Newton polytope $N(p)$ of max polynomial p is the convex hull of its coefficients' vectors.

$$\mathcal{N}(p) = \text{conv}\{v_1, v_2, v_3\}$$



Duality between Newton polytope $N(p)$ and tropical curve $V(p)$

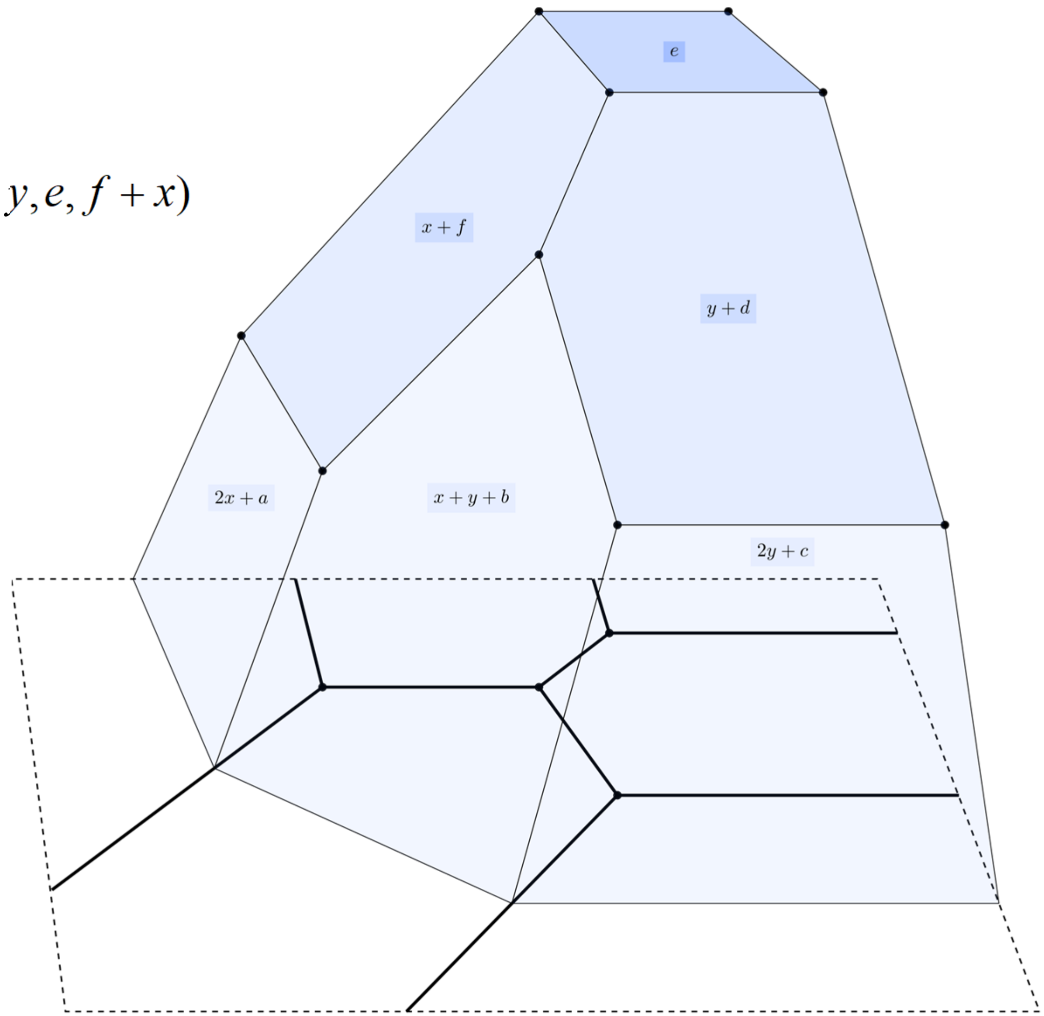
Graph and Tropic Curve of a tropical "Conic" polynomial

Tropical Polynomial of degree 2 in two variables

classical: $"ax^2 + bxy + cy^2 + dy + e + fx"$

tropical: $p(x, y) = \min(a + 2x, b + x + y, c + 2y, d + y, e, f + x)$

Graph ("tent") of $p(x,y)$
and
its **Tropical Curve** = set of (x,y) points where the min is attained by more than one terms.

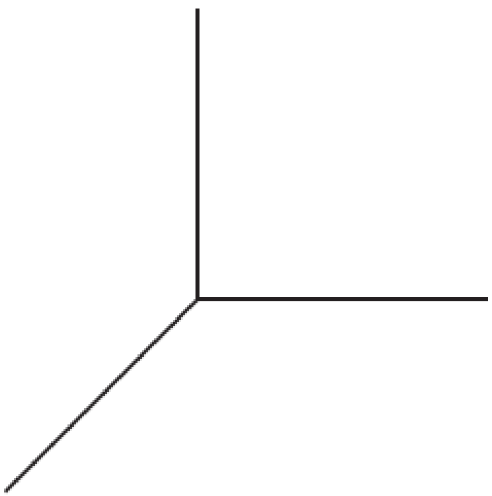


Tropical Curves of Min-plus Polynomials on the plane

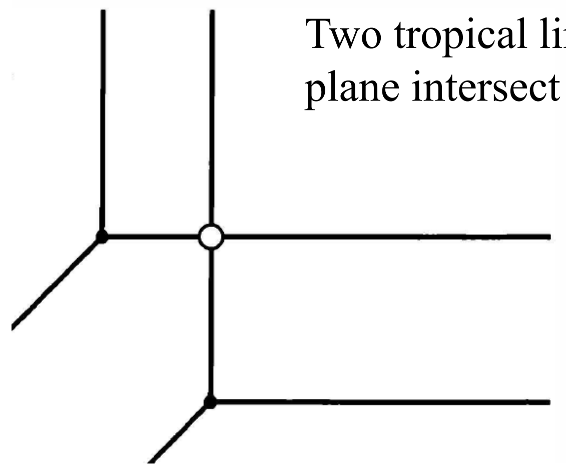
Min Polynomial of degree 1 in two variables

$$p(x,y) = \min(a+x, b+y, c)$$
$$= (a+x) \wedge (b+y) \wedge c$$

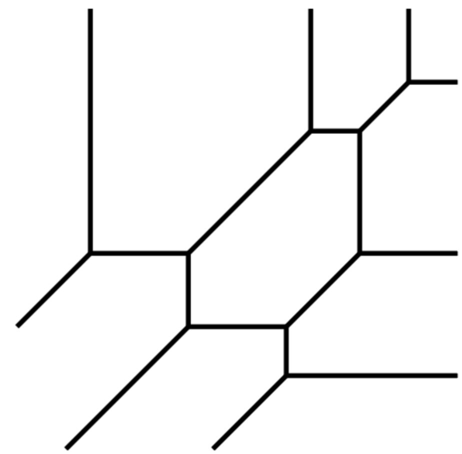
Tropical curve of $p(x,y)$



Two tropical lines on the plane intersect at one point



Cubic tropical curve



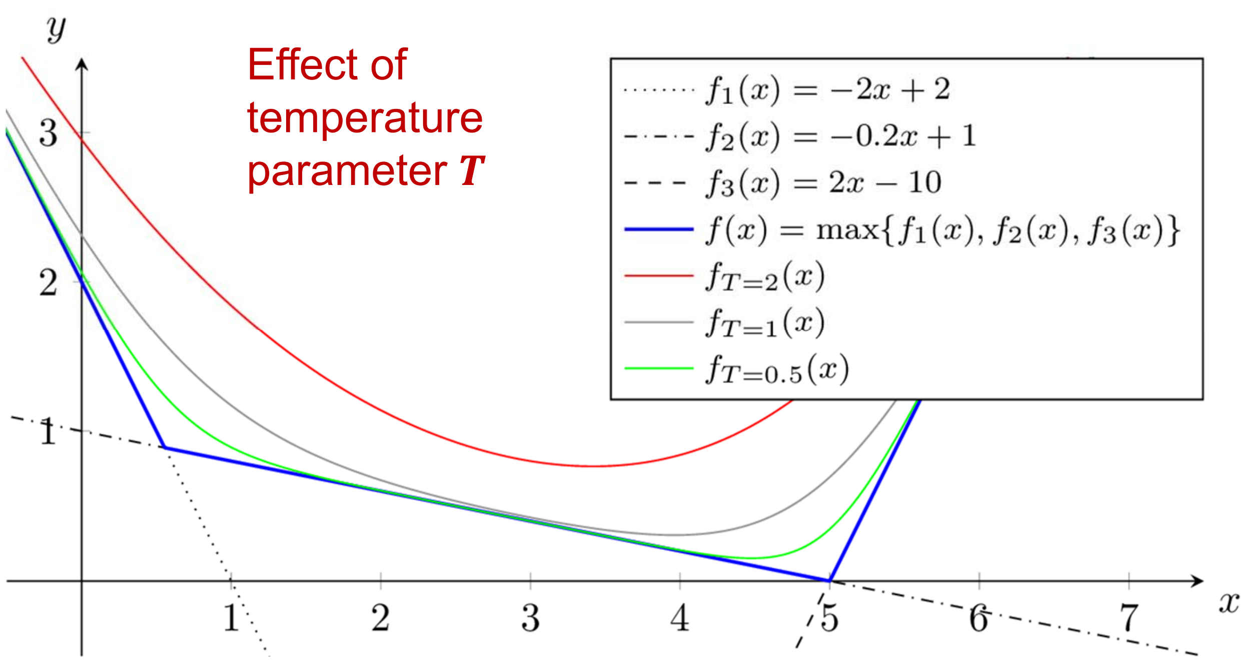
Maslov Dequantization (Tropicalization) → Log - Sum - Exp approximation

Log-Sum-Exp (LSE) approximation

(Maslov "Dequantization" in idempotent mathematics [Maslov 1987, Litvinov 2007])

$$\lim_{T\downarrow 0} T\cdot \log(e^{a/T}+e^{b/T}) = \max(a,b)$$

$$\lim_{T\downarrow 0} (-T)\log(e^{-a/T}+e^{-b/T}) = \min(a,b)$$



Obtain Tropical Polynomials via Dequantization

Classic polynomial: $f(\mathbf{u}) = \sum_{k=1}^{K} c_k u_1^{a_{k1}} u_2^{a_{k2}} \cdots u_n^{a_{kn}}, \quad \mathbf{u} = (u_1, u_2, \ldots, u_n)$

Posynomial if $c_k > 0, \ \mathbf{a}_k = (a_{k1}, \ldots, a_{kn}) \in \mathbb{R}^n, \ \mathbf{u} > 0;$

Log-Sum-Exp (Viro's "logarithmic paper" [Viro 2001]):

$\mathbf{x} = \log(\mathbf{u}), \quad b_k = \log(c_k)$

$$\lim_{T \downarrow 0} T \cdot \log f(e^{\mathbf{x}/T}) = \lim_{T \downarrow 0} T \cdot \log \sum_{k=1}^{K} \exp(\langle \mathbf{a}_k, \mathbf{x}/T \rangle + b_k / T) \ \rightarrow$$

Tropical (max-plus) **Polynomial** = Piecewise-Linear Function

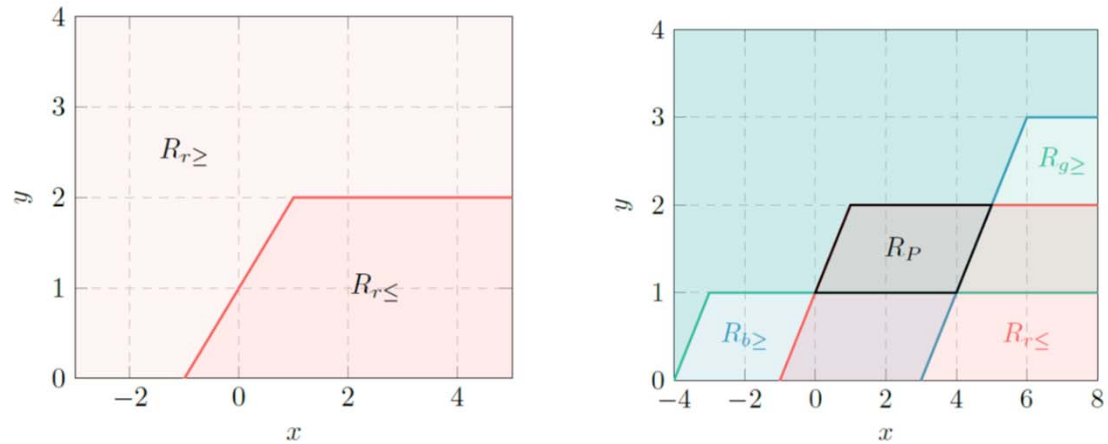
$$p(\mathbf{x}) = \operatorname*{MAX}_{k=1}^{K} \{\langle \mathbf{a}_k, \mathbf{x} \rangle + b_k\} = \operatorname*{MAX}_{k=1}^{K} \{a_{k1}x_1 + \cdots + a_{kn}x_n + b_k\}$$

Tropical Half-spaces and Polytopes in 2D

Tropical (affine) Half-space of $\mathbb{R}^n_{\max}$

[Gaubert & Katz 2011]

$$\mathscr{T}(\mathbf{a},\mathbf{b}) \triangleq \{\mathbf{x} \in \mathbb{R}^n_{\max} : \max(a_{n+1}, \bigvee_{i=1}^{n} a_i + x_i) \leq \max(b_{n+1}, \bigvee_{i=1}^{n} b_i + x_i)\}$$



(a) Single region (b) Multiple regions

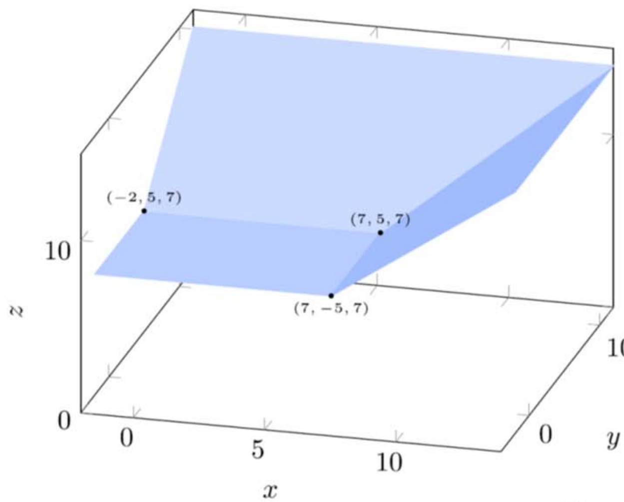
The region separating boundaries are tropical lines (or hyper-planes).

Tropical **Polyhedra** are formed from finite intersections of tropical half-spaces. **Polytopes** are compact polyhedra.

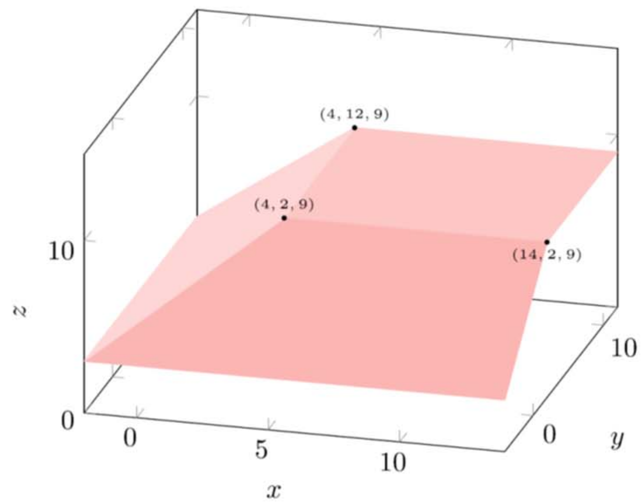
ICASSP 2024 KOREA

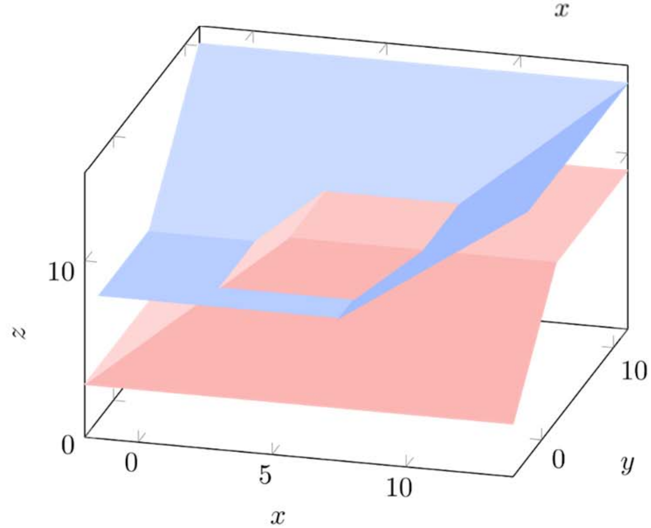
Tropical Halfspaces and Polyhedra in 3D

$$f(x, y) = \max(x, 2 + y, 7)$$



$$g(x, y) = \min(5 + x, 7 + y, 9)$$





(Extended) Newton Polytope

Let $p(\boldsymbol{x}) = \max_{i=1,\ldots,k} (\boldsymbol{a}_i^T \boldsymbol{x} + b_i)$ be a max-polynomial.

Definition ((Extended) Newton Polytope): We define as the (Extended) Newton Polytope of p the following:

$$\mathrm{Newt}(p) = \mathrm{conv}\{\boldsymbol{a}_i, i = 1, \ldots, k\}$$

$$\mathrm{ENewt}(p) = \mathrm{conv}\{(\boldsymbol{a}_i, b_i), i = 1, \ldots, k\}$$

where conv denotes the convex hull of the given set.

Theorem [Charisopoulos & Maragos, 2018; Zhang et al., 2018]:

Max-polynomials with the same vertices in the upper hull of their Extended Newton Polytope correspond to the same function.

ICASSP 2024 KOREA

Examples of (Ext) Newton Polytopes

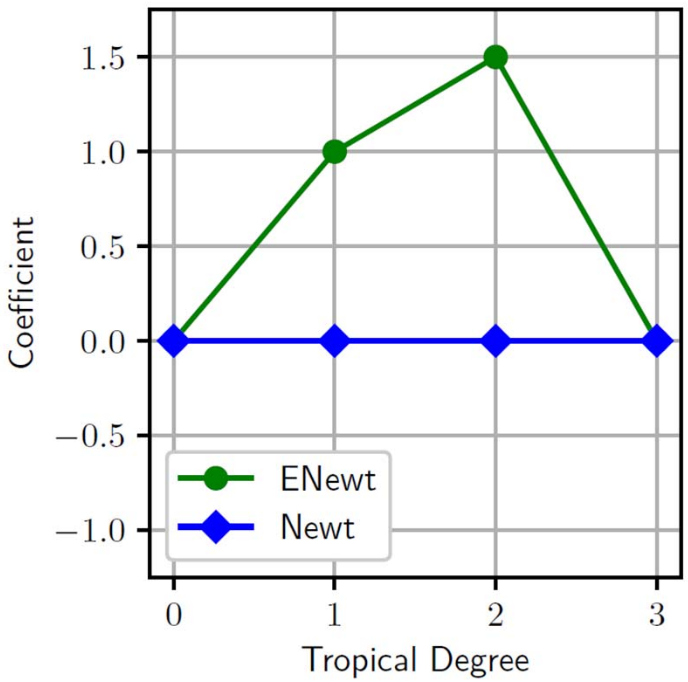


Figure: Polytopes of $\max(3x, 2x + 1.5, x + 1, 0)$.

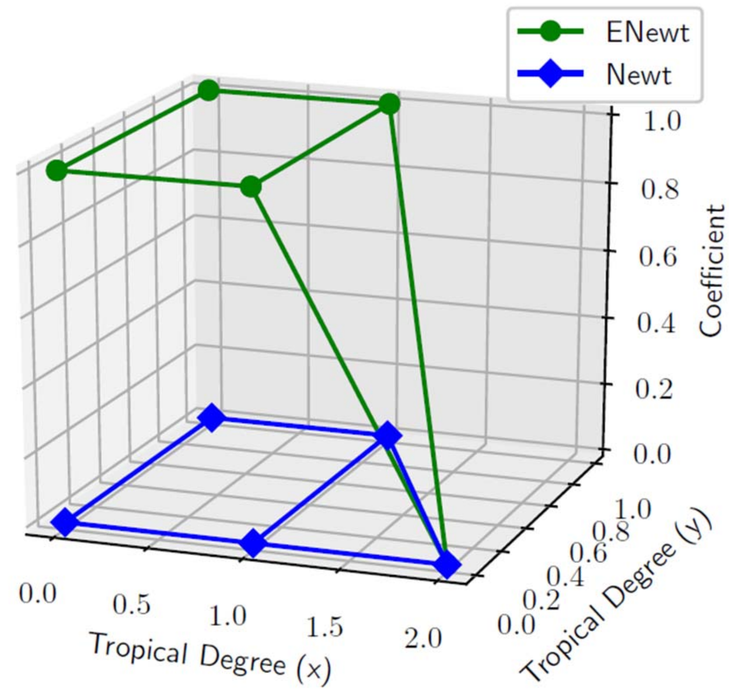


Figure: Polytopes of $\max(2x, x + y + 1, x + 1, y + 1, 1)$.

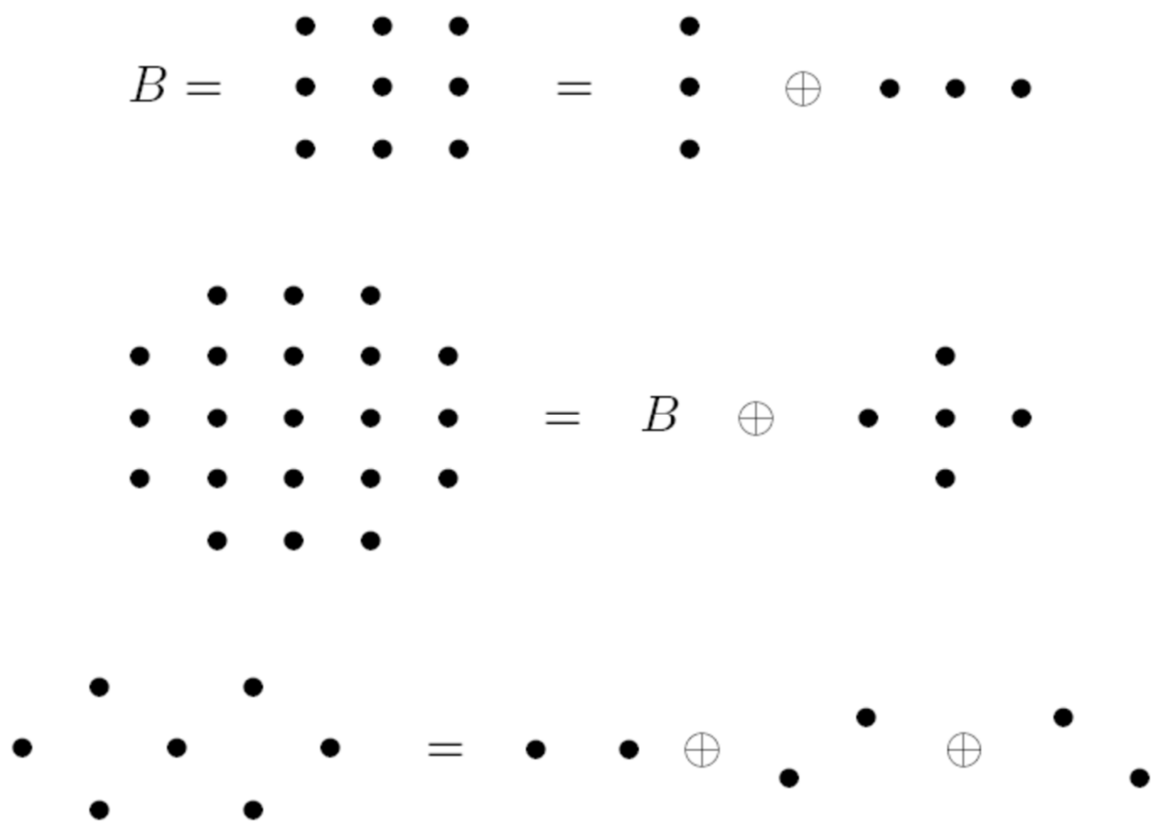
ICASSP 2024 KOREA

Minkowski Set Addition (Shape Dilation)

Definition

$$X \oplus Y = \{a + b : a \in X,\ b \in Y\}, \qquad X, Y \subseteq \mathbb{R}^d$$

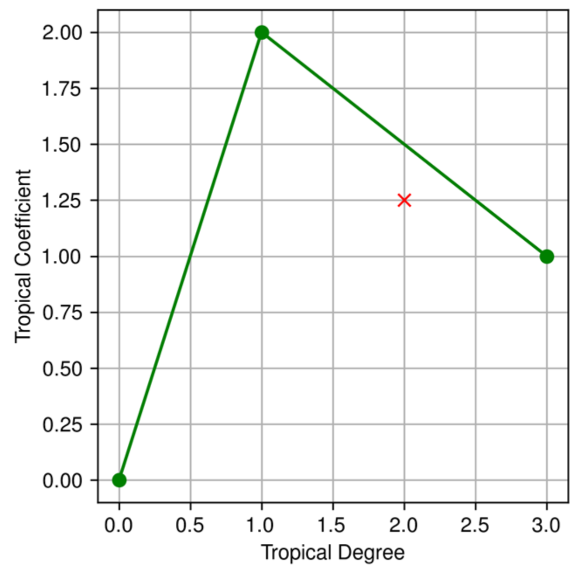
Examples



Newton Polytope and Max-polynomial Function

- “Upper” vertices of $\mathrm{ENewt}(p)$ define $p(x)$ as a function.
- Geometrically:
  $\max(3x + 1, 2x + 1.25, x + 2, 0)$
  $= \max(3x + 1, x + 2, 0)$

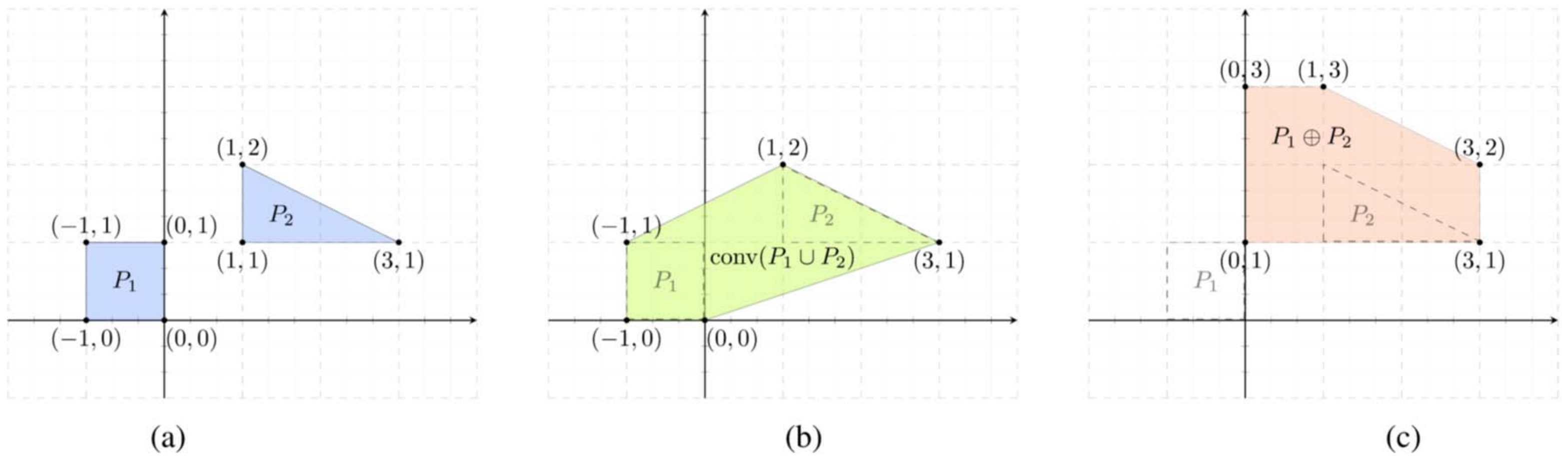
(extra point is not on the upper hull).



$\mathrm{ENewt}(p)$, $p(x) = \max(3x + 1, x + 2,0)$

Tropical Algebra of Max-plus Polynomials ←→ Tropical Geometry of their Newton Polytopes

$$\mathrm{Newt}\,(p_1 \vee p_2) = \mathrm{conv}\,(\mathrm{Newt}\,(p_1) \cup \mathrm{Newt}\,(p_2))$$
$$\mathrm{Newt}\,(p_1 + p_2) = \mathrm{Newt}\,(p_1) \oplus \mathrm{Newt}\,(p_2)$$



Newton polytopes of (a) two max-polynomials

$p_1(x,y) = max(0, -x, y, y-x)$ and $p_2(x,y) = max(x+y, 3x+y, x+2y)$,

(b) their $max(p_1 , p_2)$, and (c) their sum $p_1 + p_2$

ICASSP 2024 KOREA

Elements of Max-plus Algebra

(“*Linear algebra of Dynamic Programming & Combinatorics*”: [Butkovic 2010])

Some Earlier Special Cases and Applications

ICASSP 2024 KOREA

Research Areas using Max/Min(+) Algebra

- **Scheduling & Operations Research, Graphs**: **Minimax Algebra** [Cuninghame-Green 1979]: mainly Max-Plus.
- **Tropical Arithmetic:** **Min-plus/Max-plus Semirings** [I. Simon 1994; J.-E. Pin 1998]
- **Image & Vision, Nonlinear SP:** Image Algebra [Ritter et al, 1980s-90s], Math. Morphology [Serra 88; Heijmans & Ronse 1990s]. Morphological & Rank Filters, [Maragos & Schafer 1987]. Nonlinear Scale-Space PDEs [Brockett & Maragos 1992; Alvarez et al 1993]. Distance Transforms [Borgefors 1984; Felzenszwalb et al 2004].
- **Control**: Discrete-Event Dynamical Systems [Cohen et al 1985; Kamen 1993; Cassandras et al 2013; Heidergot et al 2006]. Dioid algebra [Cohen et al 1989; Baccelli et al 1992-2001; Gaubert & Max-plus Group 1997; Lahaye & Hardouin et al 2004; Gondran & Minoux 2008], Max-Linear Systems [Butkovic 2010, van den Boom & de Shutter 2012]. Optimization/Approximation on Semimodules [Cohen et al 2004, Akian et al 2011].
- **Speech & Language** Processing: Weighted Finite-StateAutomata/Transducers: Tropical Semiring Algorithms on Graphs [Mohri, Pereira et al, 1990s; Hori & Nakamura 2013].
- **Probabilistic Graphical Models**: Max-Sum and Max-Product algorithms in Belief Propagation [Pearl 1988; Bishop 2006; Felzenszwalb 2011].
- **Math-Physics**: Convex analysis & Optimization [Bellman & Karush 1960's; Rockafellar 1970; Lucet 2010]. **Lattices** [Birkhoff 1967]. Residuation and Ordered Algebraic Structures [Blyth 2005]. **Idempotent Mathematics** [Maslov 1987; Litvinov, Maslov et al 2000s].

Linear vs. Max-Plus Algebra: Scalar Operations

$+ \longrightarrow \max$

$\times \longrightarrow +$

Max-plus has properties similar to linear algebra:

- Commutativity: $a \vee b = b \vee a$
- Associativity: $a \vee (b \vee c) = (a \vee b) \vee c$
- Distributivity: $a + (b \vee c) = (a + b) \vee (a + c)$
- Idempotency: $3 \vee 3 = 3$
- Inverse?: $3 \vee x = 6 \Rightarrow x = 6$

  $3 \vee x = 3 \Rightarrow x = ?$

Max-plus Matrix Algebra

(Finite-dimensional Weighted Lattices)

- vector/matrix **'addition'** = pointwise max

$$\begin{aligned} \mathbf{x} \vee \mathbf{y} &= [x_1 \vee y_1, \ldots, x_n \vee y_n]^T \\ \mathbf{A} \vee \mathbf{B} &= [a_{ij} \vee b_{ij}] \end{aligned}$$

- vector/matrix **'dual addition'** = pointwise min

$$\begin{aligned} \mathbf{x} \wedge \mathbf{y} &= [x_1 \wedge y_1, \ldots, x_n \wedge y_n]^T \\ \mathbf{A} \wedge \mathbf{B} &= [a_{ij} \wedge b_{ij}] \end{aligned}$$

- vector/matrix **'multiplication by scalar'**

$$\begin{aligned} c + \mathbf{x} &= [c + x_1, \ldots, c + x_n]^T \\ c + \mathbf{A} &= [c + a_{ij}] \end{aligned}$$

- $(\max, +)$ **'matrix multiplication'**

$$[\mathbf{A} \boxplus \mathbf{B}]_{ij} = \bigvee_{k=1}^{n} a_{ik} + b_{kj}$$

- $(\min, +)$ **'matrix dual multiplication'**

$$[\mathbf{A} \boxplus' \mathbf{B}]_{ij} = \bigwedge_{k=1}^{n} a_{ik} + b_{kj}$$

ICASSP 2024 KOREA

Tropical (Max-plus) Matrix Multiplication

Matrix-Vector Multiplication

$$[\boldsymbol{A} \boxplus \boldsymbol{x}]_i = \max_j (A_{ij} + x_j)$$

Matrix Multiplication

$$[\boldsymbol{A} \boxplus \boldsymbol{B}]_{ij} = \max_l (A_{il} + B_{lj})$$

Example

$$\begin{bmatrix} 1 & -1 & 3 \\ 3 & -2 & -1 \end{bmatrix} \boxplus \begin{bmatrix} 2 & 1 & 2 \\ -4 & 3 & -1 \\ 0 & 2 & -2 \end{bmatrix} = \begin{bmatrix} \cdot & \max(1+1, -1+3, 3+2) & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$$

Morphological Operators on Lattices

($\leq$ = partial ordering, V = supremum, Λ = infimum)

- ψ is **increasing** iff $f \leq g \Rightarrow \psi(f) \leq \psi(g)$.
- δ is **dilation** iff $\delta(\vee_i f_i) = \vee_i \delta(f_i)$.
- ε is **erosion** iff $\varepsilon(\wedge_i f_i) = \wedge_i \varepsilon(f_i)$.
- α is **opening** iff increasing and antiextensive $(\alpha(f) \leq f)$,

  and idempotent $(\alpha = \alpha^2)$: lattice projection
- β is **closing** iff increasing and extensive $(\beta(f) \geq f)$,

  and idempotent $(\beta = \beta^2)$: lattice projection
- (δ, ε) is **adjunction** iff $\delta(f) \leq g \Leftrightarrow f \leq \varepsilon(g)$ (Galois connection)

  Then: δ is dilation, ε is erosion,

  $\delta\varepsilon$ is opening, $\varepsilon\delta$ is closing.

[Serra 1988; Heijmans & Ronse 1990]

ICASSP 2024 KOREA

Tropical Semirings versus Weighted Lattices

Weighted Lattice = Tropical Space

	Flat Lattice $(\mathbb{R} \cup \{-\infty, +\infty\}, \vee, \wedge)$	
Max-plus Semiring $(\mathbb{R} \cup \{-\infty\}, \vee, +)$	$(\mathbb{R} \cup \{-\infty\}, \max)$ is Idempotent Semigroup	$(\mathbb{R}, +)$ is Group Addition $(+)$ distributes over $\vee$
Min-plus Semiring $(\mathbb{R} \cup \{+\infty\}, \wedge, +')$	$(\mathbb{R} \cup \{+\infty\}, \min)$ is Idempotent Semigroup	$(\mathbb{R}, +')$ is Group Dual Addition $(+')$ distributes over $\wedge$
	Duality between $\vee$ and $\wedge$	

[P. Maragos, "*Dynamical Systems on Weighted Lattices: General Theory*", Math. Control, Signals and Systems, 2017.]

Linear and Nonlinear Spaces

Linear spaces (Vector Spaces):

Signal Superposition (+): $f(t) + g(t)$

Scaling (x): $c \cdot f(t)$

$$\sum_i c_i f_i(t)] \longrightarrow \boxed{\text{Linear system } \Gamma} \longrightarrow \sum_i c_i \Gamma[f_i(t)]$$

Nonlinear spaces (Tropical spaces = Weighted Lattices):

Signal Superposition : max: $f(t) \vee g(t)$ **min:** $f(t) \wedge g(t)$

Scaling (+): $c + f(t)$

$$\bigvee_i c_i + f_i(t) \longrightarrow \boxed{\text{Tropical system } \Delta} \longrightarrow \bigvee_i c_i + \Delta[f_i(t)]$$

ICASSP 2024 KOREA

Operation	Meaning
$\bigvee$	Maximum/Supremum: applies for scalars, vectors and matrices
$\bigwedge$	Minimum/Infimum: applies for scalars, vectors and matrices
$\boxed{\star}$ ($\boxed{\star}'$)	General max-$\star$ (min-$\star'$) matrix multiplication
$\boxplus$ ($\boxplus'$)	Max-sum (min-sum) matrix multiplication
$\boxtimes$ ($\boxtimes'$)	Max-product (min-product) matrix multiplication
$\circledast$ ($\circledast'$)	General max-$\star$ (min-$\star'$) signal convolution
$\oplus$ ($\oplus'$)	Max-sum (min-sum) signal convolution
$\otimes$ ($\otimes'$)	Max-product (min-product) signal convolution

max-sum and min-sum

matrix multiplications

$$\boldsymbol{C} = \boldsymbol{A} \boxplus \boldsymbol{B} = [c_{ij}] \quad , \quad c_{ij} = \bigvee_{k=1}^{n} a_{ik} + b_{kj}$$

$$\boldsymbol{C} = \boldsymbol{A} \boxplus' \boldsymbol{B} = [c_{ij}] \quad , \quad c_{ij} = \bigwedge_{k=1}^{n} a_{ik} + b_{kj}$$

signal convolutions

$$(f \oplus h)(t) = \bigvee_{k=-\infty}^{+\infty} f(t-k) + h(k)$$

$$(f \oplus' h)(t) = \bigwedge_{k=-\infty}^{+\infty} f(t-k) + h(k)$$

ICASSP 2024 KOREA

Examples of Adjunctions

- **Set Operator Adjunction**: Minkowski set addition $\oplus$ and subtraction $\ominus$: for $X, B \subseteq \mathbb{R}^d$

$$\begin{aligned} \delta_B(X) = X \oplus B &:= \{\mathbf{x}+\mathbf{b} \in \mathbb{R}^d : \mathbf{x} \in X, \mathbf{b} \in B\} \\ \varepsilon_B(X) = X \ominus B &:= \{\mathbf{x}-\mathbf{b} \in \mathbb{R}^d : \mathbf{x} \in X, \mathbf{b} \in B\} \end{aligned}$$

- **Vector Operator Adjunction**: max-plus vector multiplication by matrix $\mathbf{A} \in \overline{\mathbb{R}}^{m\times n}$ and min-plus vector multiplication by matrix $\mathbf{A}^* = -\mathbf{A}^T$:

$$\begin{aligned} \delta_{\mathbf{A}}(\mathbf{x}) &= \mathbf{A} \boxplus \mathbf{x}, \quad [\delta_{\mathbf{A}}(\mathbf{x})]_i = \bigvee_{j=1}^{n} a_{ij} + x_j \\ \varepsilon_{\mathbf{A}}(\mathbf{y}) &= \mathbf{A}^* \boxplus' \mathbf{y}, \quad [\varepsilon_A(\mathbf{y})]_j = \bigwedge_{i=1}^{m} y_i - a_{ij} \end{aligned}$$

- **Signal Operator Adjunction**: max-plus convolution of $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ with k and min-plus convolution of $g(\mathbf{x})$ with $-k(-\mathbf{x})$:

$$\begin{aligned} \delta_k(f)(\mathbf{x}) = f \oplus k(\mathbf{x}) &:= \bigvee_{\mathbf{y}} \{f(\mathbf{y}-\mathbf{x}) + k(\mathbf{y})\} \\ \varepsilon_k(g)(\mathbf{x}) = g \ominus k(\mathbf{x}) &:= \bigwedge_{\mathbf{y}} \{g(\mathbf{x}+\mathbf{y}) - k(\mathbf{y})\} \end{aligned}$$

ICASSP 2024 KOREA

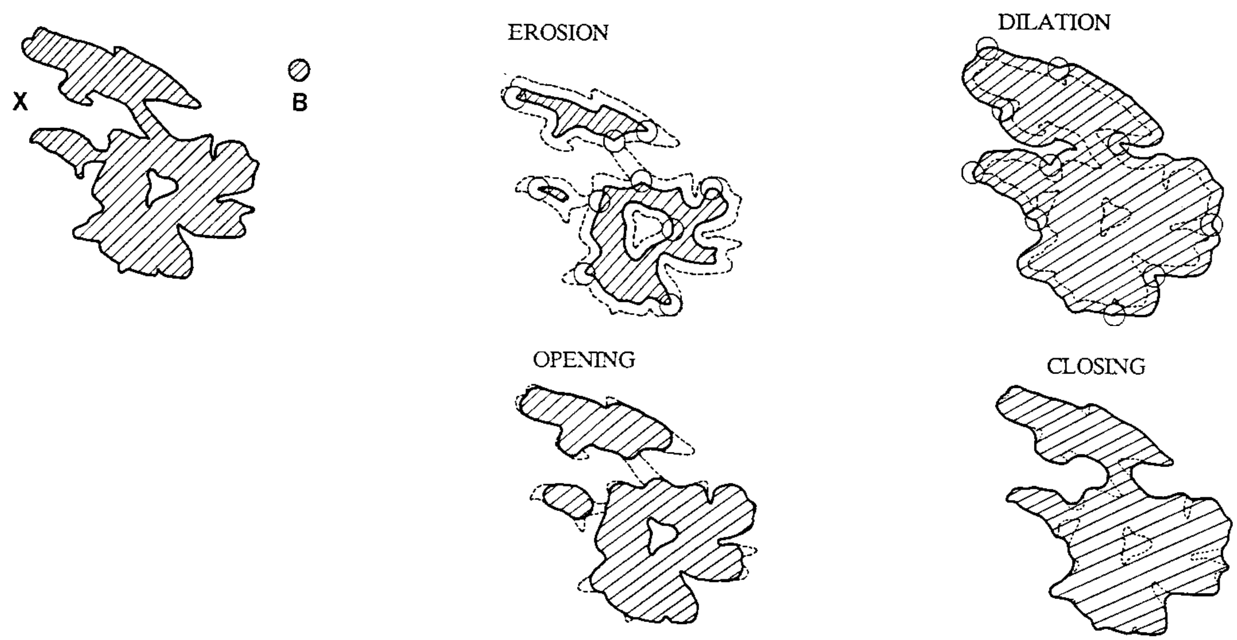
Minkowski-Hadwiger Morphological Set Operators

Translation: $B_{+z} = \{b + z : b \in B\}$ **Symmetric:** $B^s = \{-b : b \in B\}$

Dilation (Minkowski addition): $X \oplus B = \{z : (B^s)_{+z} \cap X \neq \varnothing\} = \bigcup_{b \in B} X_{+b}$

Erosion (Minkowski subtraction): $X \ominus B = \{z : B_z \subseteq X\} = \bigcap_{b \in B} X_{-b}$

Hadwiger Opening: $X \circ B = (X \ominus B) \oplus B$ **Closing:** $X \bullet B = (X \oplus B) \ominus B$



Max/Min-plus Convolutions and Filters-Projections

Max-plus Convolution (**Dilation**) by a square (flat *g*)
(= Max Pooling in CNNs)

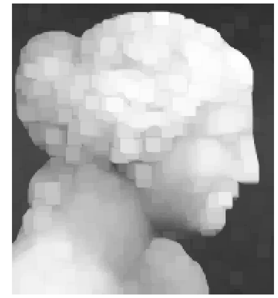
$$(f \oplus g)(x) = \bigvee_y f(y) + g(x-y)$$

Adjoint Min-plus Correlation (**Erosion**)

$$(f \ominus g)(x) = \bigwedge_y f(y) - g(y-x)$$

IMAGE

DILATION 9x9



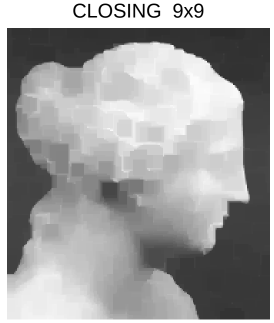
EROSION 9x9

Serial compositions of max-convolution and adjoint min-plus correlation: **Opening**, **Closing**

$$f \circ g \triangleq (f \ominus g) \oplus g \qquad f \bullet g \triangleq (f \oplus g) \ominus g$$

OPENING 9x9

CLOSING 9x9



Idempotent Operators = **Projections**
on Nonlinear Spaces (Weighted Lattices)

$$(f \circ g) \circ g = f \circ g$$
$$(f \bullet g) \bullet g = f \bullet g$$

Solve Max-plus Equations via Adjunctions

- **Problems**:

  (1) Exact problem: Solve $\delta_{\mathbf{A}}(\mathbf{x}) = \overbrace{\mathbf{A} \boxplus \mathbf{x}}^{\text{max-plus}} = \mathbf{b}, \quad \mathbf{A} \in \overline{\mathbb{R}}^{m \times n}, \ \ \mathbf{b} \in \overline{\mathbb{R}}^{m}$

  (2) Approximate Constrained: Min $\|\mathbf{A} \boxplus \mathbf{x} - \mathbf{b}\|_{p=1\ldots\infty}$ s.t. $\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}$

- **Theorem**: The **greatest (sub)solution** of (1) and unique solution of (2) is

$$\hat{\mathbf{x}} = \varepsilon_{\mathbf{A}}(\mathbf{b}) = \mathbf{A}^* \boxplus' \mathbf{b}, \quad [\hat{\mathbf{x}}]_j = \bigwedge_{i=1}^{m} b_i - a_{ij}, \quad \mathbf{A}^* \triangleq -\mathbf{A}^T$$

  and yields the **Greatest Lower Estimate (GLE)** of data $\mathbf{b}$:

$$\delta_{\mathbf{A}}(\varepsilon_{\mathbf{A}}(\mathbf{b})) = \underbrace{\mathbf{A} \boxplus \left(\underbrace{\mathbf{A}^* \boxplus' \mathbf{b}}_{\text{min-plus}} \right)}_{\text{max-plus matrix product}} \leq \mathbf{b}$$

- **Geometry**: Operators δ, ε are vector dilation and erosion, and
  the GLE $\mathbf{b} \mapsto \delta(\varepsilon(\mathbf{b}))$ is an opening (lattice projection).

- **Complexity**: $O(mn)$

Adjunction versus Residuation pairs

- An increasing operator $\psi : \mathcal{L} \to \mathcal{M}$ between complete lattices is called **residuated** if there exists an increasing operator $\psi^\sharp : \mathcal{M} \to \mathcal{L}$ such that

$$\psi\psi^\sharp \leq \mathbf{id} \leq \psi^\sharp\psi$$

  $\psi^\sharp$ is called the **residual** of ψ, is unique, and closest to being an inverse of ψ.

- A residuation pair $(\psi, \psi^\sharp)$ can solve **inverse problems** $\psi(X) = Y$ either *exactly* since $\hat{X} = \psi^\sharp(Y)$ is the greatest solution of $\psi(X) = Y$ if a solution exists, or *approximately* since $\hat{X}$ is the **greatest subsolution**:

$$\hat{X} = \psi^\sharp(Y) = \bigvee\{X : \psi(X) \leq Y\}$$

- A pair (δ, ε) of operators $\delta : \mathcal{L} \to \mathcal{M}$ and $\varepsilon : \mathcal{M} \to \mathcal{L}$ is called **adjunction** if

$$\delta(X) \leq Y \Longleftrightarrow X \leq \varepsilon(Y) \quad \forall X \in \mathcal{L}, Y \in \mathcal{M}$$

  δ is a **dilation** and ε is an **erosion**.
  Each dilation δ corresponds to a unique *adjoint erosion*

$$\varepsilon(Y) = \delta^\sharp(Y) = \bigvee\{X : \delta(X) \leq Y\}$$

- Adjunction $\Longleftrightarrow$ Residuation *iff* $\psi = \delta$ and $\psi^\sharp = \varepsilon$.
- Viewing (δ, ε) as adjunction instead of residuation offers *geometric intuition.*

Some Earlier Special Cases of Max-plus Algebra and Applications

Linear versus Max-Plus Systems

- **State space representation**: linear vs. max-plus

$$x(k) = Ax(k-1) + Bu(k)$$
$$y(k) = Cx(k) + Du(k)$$

$$x(k) = A \boxplus x(k-1) \vee B \boxplus u(k)$$
$$y(k) = C \boxplus x(k) \vee D \boxplus u(k)$$

- **Matrix products**
  - Linear: $$[AB]_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$
  - Max-plus: $$[A \boxplus B]_{ij} = \bigvee_{k=1}^{n} a_{ik} + b_{kj}$$
- **Example**

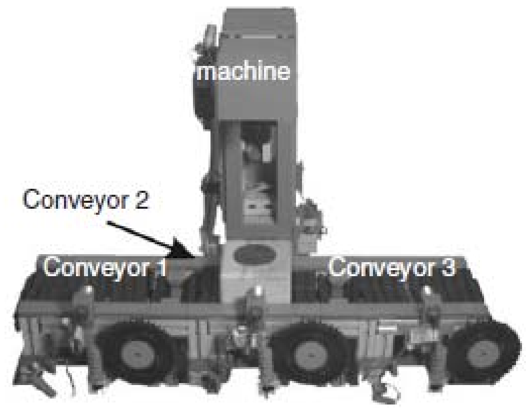
$$\begin{bmatrix} 4 & -1 \\ 2 & -\infty \end{bmatrix} \boxplus \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \left\{ \begin{aligned} \max(x+4, y-1) = 3 \\ x+2 = 1 \end{aligned} \right\} \Longrightarrow \begin{aligned} x = -1 \\ y \leq 4 \end{aligned}$$

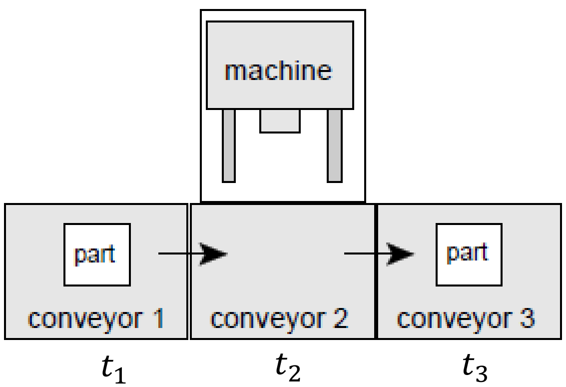
- What can we model with max-plus systems?

ICASSP 2024 KOREA

Automated Manufacturing as Max-plus System

Discrete event systems (*)





$x_i(k)$: time product k enters conveyor i
$u(k)$: time we put product k in conveyor 1
t_i: conveyor i waiting time
Only one product in a conveyor during each cycle

$$x_1(k) = \max(x_1(k-1) + t_1, u(k))$$
$$x_2(k) = \max(x_1(k) + t_1, x_2(k-1) + t_2)$$
$$x_3(k) = \max(x_2(k) + t_2, x_3(k-1) + t_3)$$

$$A = \begin{bmatrix} t_1 & -\infty & -\infty \\ 2t_1 & t_2 & -\infty \\ 2t_1 + t_2 & 2t_2 & t_3 \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ t_1 \\ t_1 + t_2 \end{bmatrix}$$

$$x(k) = A \boxplus x(k-1) \vee B \boxplus u(k)$$

(*) Example from: [G. Schullerus, V. Krebs, B. De Schutter & T. van den Boom, "*Input signal design for identification of max-plus-linear systems*", Automatica 2006.]

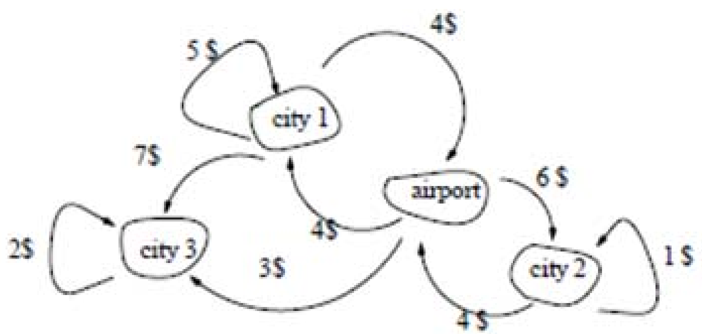
ICASSP 2024 KOREA

Longest/Shortest Paths as Max/Min-plus Systems

Dynamic Programming

- **Taxi drivers (*)**

$$\boldsymbol{x}(k+1) = \boldsymbol{A}^T \boxplus \boldsymbol{x}(k)$$



$$A^T = \begin{bmatrix} 5 & 4 & -\infty & 7 \\ 4 & -\infty & 6 & 3 \\ -\infty & 4 & 1 & -\infty \\ -\infty & -\infty & -\infty & 2 \end{bmatrix}$$

$$money_i(k) = \max_j \left(money_j(k-1) + a_{ji} \right)$$

x_1, x_2, x_3, x_4 correspond to city 1, airport, city 2 and city 3

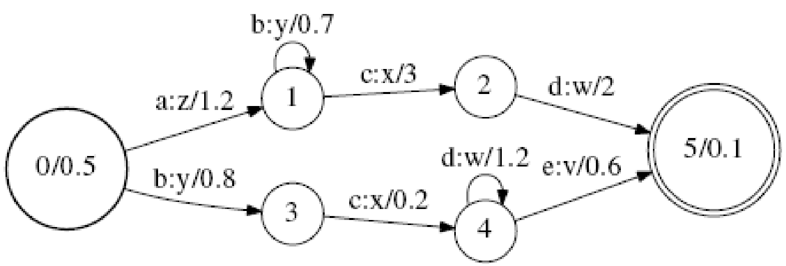
(*) Example from:
[S. Gaubert and Max-Plus group, "*Methods and applications of (max,+) linear algebra*“, STACS 1997.]

WFSTs for Speech Recognition: Tropical (Min-Plus) Algebra

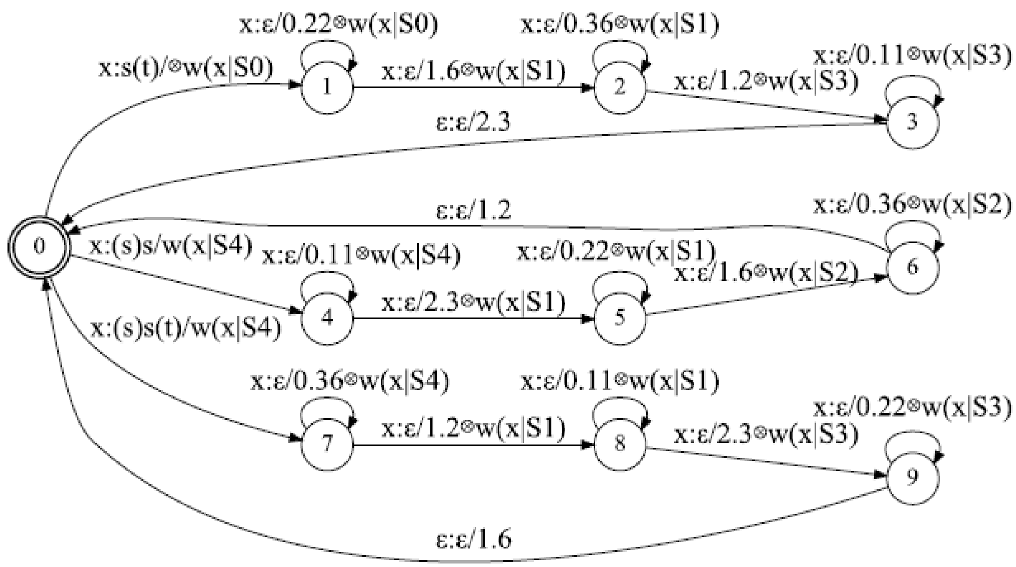
Weighted Finite State Transducer (WFST)

[Mohri, Pereira & Ripley, CSL 2002]

[Hori and Nakamura, 2013]



HMM Transducer: converts an input speech signal into a seq of context-dependent phone units

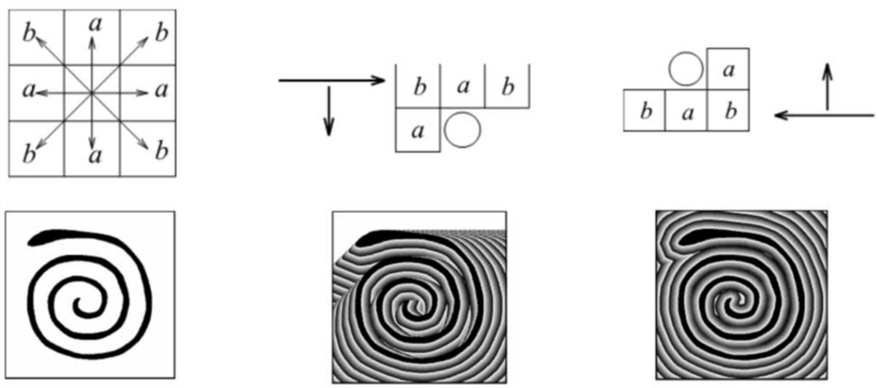


ICASSP 2024 KOREA

Distance Computation with Min-plus Difference Eqns

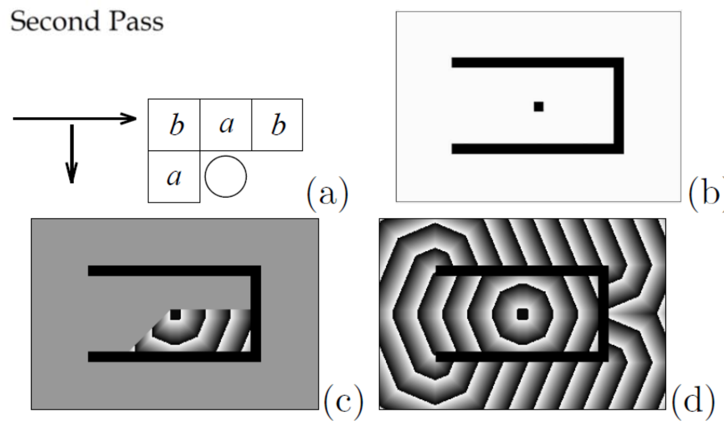
Two – Pass Algorithm

$$\begin{aligned} u_1[i,j] &= \min(u_1[i,j-1]+a, u_1[i-1,j]+a, \\ &\quad u_1[i-1,j-1]+b, u_1[i-1,j+1]+b, u_0[i,j]) \\ u_2[i,j] &= \min(u_2[i,j+1]+a, u_2[i+1,j]+a, \\ &\quad u_2[i+1,j+1]+b, u_2[i+1,j-1]+b, u_1[i,j]) \end{aligned}$$



Initial Image First Pass Second Pass

Sequential Distance Computation with Obstacles



Gaussian Scale-Space → Maslov Dequantiz → Dilation/Erosion Scale-Space

Heat PDE: $\dfrac{\partial u}{\partial t} = \dfrac{h}{2}\dfrac{\partial^2 u}{\partial x^2}$

Substitution (LSE - LogSumExpon): $u = e^{-W/h}$

F

$\Rightarrow$

Multiscale Gaussian Blurring

Hopf's eqn: $\dfrac{\partial W}{\partial t} + \dfrac{1}{2}\left(\dfrac{\partial W}{\partial x}\right)^2 - \dfrac{h}{2}\dfrac{\partial^2 W}{\partial x^2} = 0$

Dequantization: $\lim_{h\to 0} h\cdot \log(e^{-a/h} + e^{-b/h}) = \min(a,b)$

HJE: $\dfrac{\partial S}{\partial t} + \dfrac{1}{2}\left(\dfrac{\partial S}{\partial x}\right)^2 = 0$

$\Rightarrow S(x,t) = \text{Multiscale Erosion by Parabola } (-x^2/2t)$

Multiscale Max/Min Pooling

$\Rightarrow$

- Erosion (-F)

Tropical Geometry of Neural Nets with Piecewise-Linear Activations

Main References:

1. Charisopoulos, V., & Maragos, P. (2017, May). *Morphological perceptrons: geometry and training algorithms*, ISMM '17.
2. Charisopoulos, V., & Maragos, P. (2018). A Tropical Approach to Neural Networks with Piecewise Linear Activations. `arXiv:1805.08749`.
3. Zhang, Liwen and Naitzat, Gregory and Lim, Lek-Heng. *Tropical geometry of deep neural networks*, Proc. ICML(35) 2018.

Related:

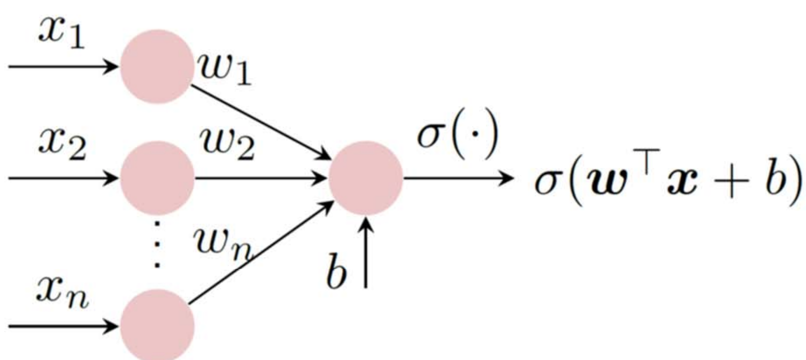
- M. Alfarra et al, *On the decision boundaries of neural networks: A tropical geometry perspective*, arXiv 2020.
- A. Humayun et al., *SplineCam: Visualization of Deep Network Geometry and Decision Boundaries*, CVPR 2023.

ICASSP 2024 KOREA

NNs with PWL functions

Piecewise-linear functions used as *activation* functions σ:

1. **ReLU**: $\max(0, v)$ or $\max(\alpha v, v),\ \alpha \ll 1$ with $v := \boldsymbol{w}^\top \boldsymbol{x} + b$
2. **Maxout**: $\max_{k\in[K]} v_k$ with $v_k := \boldsymbol{W}_k^\top \boldsymbol{x} + b_k$



> **Linear regions**: maximally connected regions of input space on which the NN's output is linear [Montufar et al., 2014].

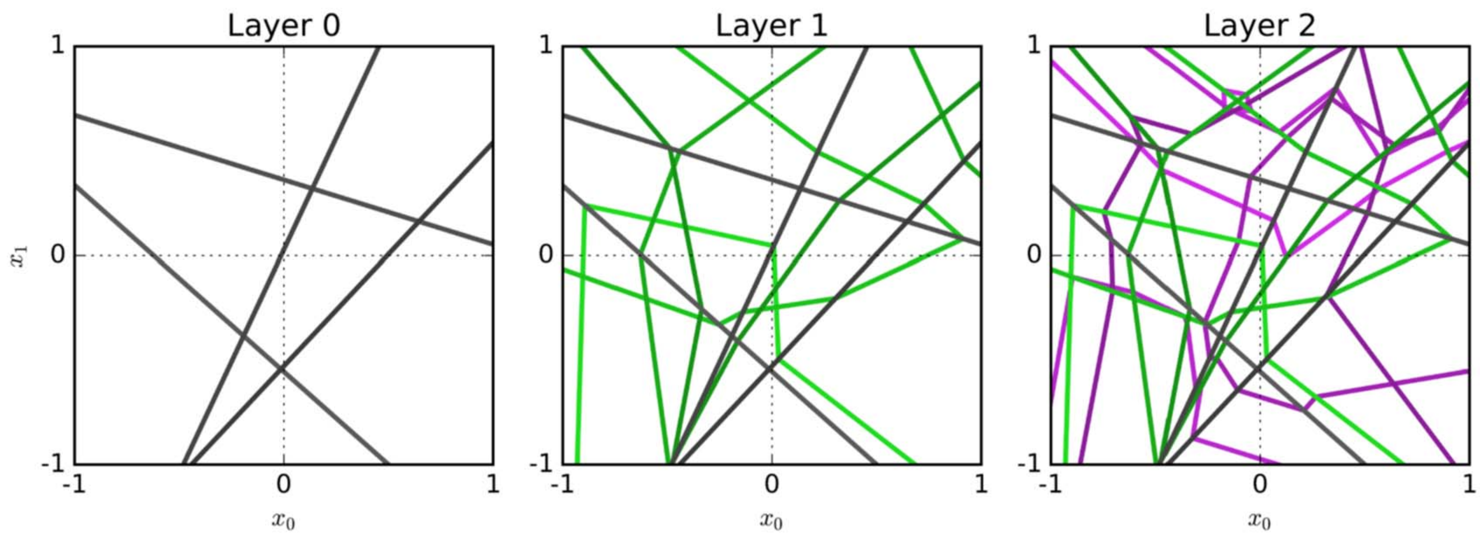


Figure: Input space is subdivided into convex polytopes, each of which is a "linear region" for the NN. Reproduced from [Raghu et al., 2016]

Claim: more linear regions ≡ more expressive power

PWL functions and tropical geometry

Convex + PWL: ideal to study under lens of **tropical geometry**

Formally: *tropical semiring* $(\mathbb{R} \cup \{-\infty\}, \vee, +)$

- binary "addition" $x \vee y := \max(x, y)$, "multiplication" $x + y$
- operations on vectors $\boldsymbol{x}, \boldsymbol{y}$:

$$\boldsymbol{x} \vee \boldsymbol{y} := \begin{pmatrix} \max(x_1, y_1) \\ \vdots \\ \max(x_n, y_n) \end{pmatrix}, \quad \boldsymbol{x}^\top \boxplus \boldsymbol{y} := \bigvee_{i=1}^{n} x_i + y_i$$

Key object: tropical {poly, posy, sig}nomials

ICASSP 2024 KOREA

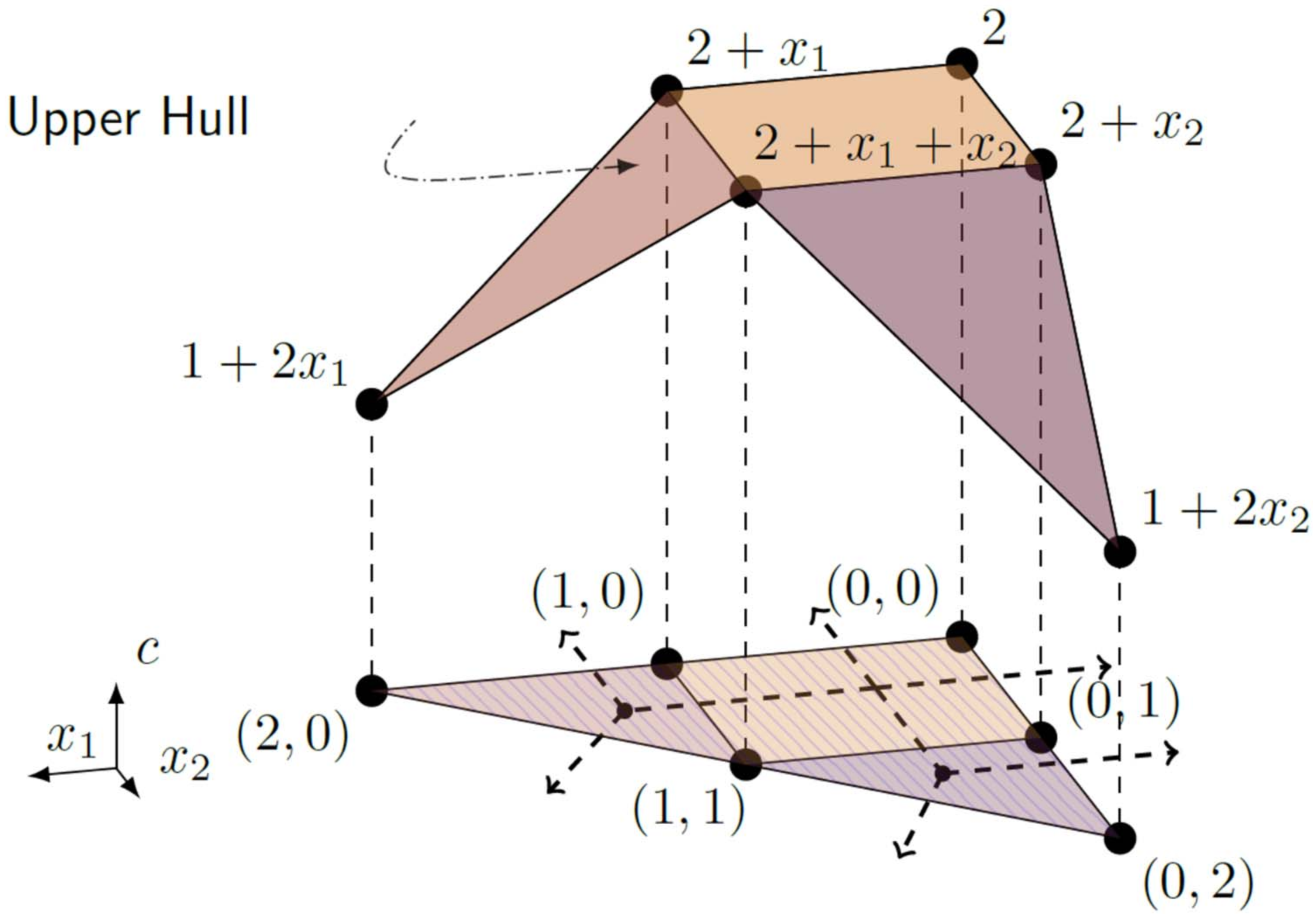
Single neuron result

An application of the fundamental theorem of LP yields:

Proposition [Charisopoulos & Maragos, 2017]

The number of linear regions for a single maxout unit $p(\boldsymbol{x}) = \max_{j\in[k]} \boldsymbol{w}_j^\top \boldsymbol{x} + b_i$ are equal to the number of vertices on the upper hull of $\mathcal{N}(p)$

- subsumes **relu**
- all terms corresponding to interior vertices can be *removed* without affecting $p(\boldsymbol{x})$ *as a function*.



$$p(x_1,x_2)=\max(1+2x_1, 2+x_1, 2, 2+x_2, 1+2x_2, 2+x_1+x_2)$$

For a collection of tropical polynomials, suffices to work with Minkowski sums:

Proposition [Charisopoulos & Maragos, 2018] [Zhang et al., 2018]

The number of linear regions of a layer with n inputs and m neurons is upper bounded by the number of vertices in the upper convex hull of

$$\mathcal{N}(p_1) \oplus \cdots \oplus \mathcal{N}(p_m),$$

where $\oplus$ denotes Minkowski sum.

ICASSP 2024 KOREA

Main Result

Immediate application of a bound from [Gritzmann and Sturmfels, 1993] on faces of Minkowski sums gives

Proposition [Charisopoulos & Maragos, 2018]

The number of linear regions of n input, m output layer consisting of convex PWL activations of rank k is bounded above by

$$\min\left\{k^m, 2\sum_{j=0}^{n}\binom{m^{\frac{k(k-1)}{2}}}{j}\right\}.$$

In case of ReLU, use symmetry of zonotopes to refine to

$$\min\left\{2^m, \sum_{j=0}^{n}\binom{m}{j}\right\}$$

ICASSP 2024 KOREA

Counting in practice

Goal: given a network, count # of linear regions (exactly or approximately)

Exact counting using insight from Newton polytopes:

- ▷ vertex enumeration algorithm for Mink. sums [Fukuda, 2004] $\Rightarrow$ requires solving $\Omega(|\mathrm{vert}(P)|)$ LPs.
- ▷ impractical unless problem is small

MIP representability of NNs [Serra et al., 2018]:

- ▷ Assumes bounded range of input space
- ▷ Requires enumerating solutions of MILPs

Geometric Algorithm: Randomized method for Sampling the Extreme Points of the Upper Hull of a Polytope [Charisopoulos & Maragos 2019, arXiv:1805.08749v2], [Maragos, Charisopoulos & Theodosis, Proc. IEEE 2021]

Computational Geometry: [Karavelas & Tzanaki, ISCG 2015]: A Geometric Approach for the Upper Bound Theorem for Minkowski Sums of Convex Polytopes

Geometry & Algebra of NNs with PWL Activations

Theorem (Wang 2004): A continuous piecewise linear function is equal to the difference of two max-polynomials.

Theorem (Charisopoulos & Maragos 2018): The essential terms of a tropical polynomial are in bijection $1-1$ with the vertices on the upper convex hull of its extended Newton polytope.

Theorem (Zhang et al. 2018): A neural network with ReLU-type activations can be represented as the difference of two max-polynomials(*), i.e. with a tropical rational function.

[(*) Zhang et al. only call “max polynomials" those polynomials with integer slopes]

[Calafiore et al., 2019] use the Maslov dequantization to design universal approximators for convex (+loglog-convex) data

$$f \text{ convex} \Rightarrow f \simeq f_{\mathrm{PWL}} \Leftrightarrow f \simeq f_T,$$

where $f_{\mathrm{PWL}} \leq f_T \leq T \log K + f_{\mathrm{PWL}}$ and are given by

$$\begin{cases} f_{\mathrm{PWL}} := \max_{k \in [K]} \langle \boldsymbol{a}_k, \boldsymbol{x} \rangle + b_k, \\ f_T := T \log \left(\sum_{k=1}^{K} \exp \left\{ b_k + \langle \boldsymbol{a}_k, \boldsymbol{x} \rangle \right\}^{1/T} \right) \end{cases}$$

In particular, fixing $\varepsilon > 0$ and compact $\mathcal{C}$, a small enough T will satisfy

$$\sup_{\boldsymbol{x} \in \mathcal{C}} |f_T(\boldsymbol{x}) - f(\boldsymbol{x})| \leq \varepsilon.$$

ICASSP 2024 KOREA

Morphological Networks: Geometry, Training, and Pruning

References:

- V. Charisopoulos and P. Maragos, "Morphological Perceptrons: Geometry and Training Algorithms", Proc. ISMM 2017, LNCS 10225, Springer.
- N. Dimitriadis and P. Maragos, "Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints", Proc. ICASSP, 2021.

ICASSP 2024 KOREA

Motivation

- Explosion of ML research in the last decade (now models with near-human or even human performance)
- Recent advances indicate shift towards nonlinearity, but…
- ...the “multiply-accumulate” (= linear) operations of the perceptron are still ubiquitous

Our Questions:

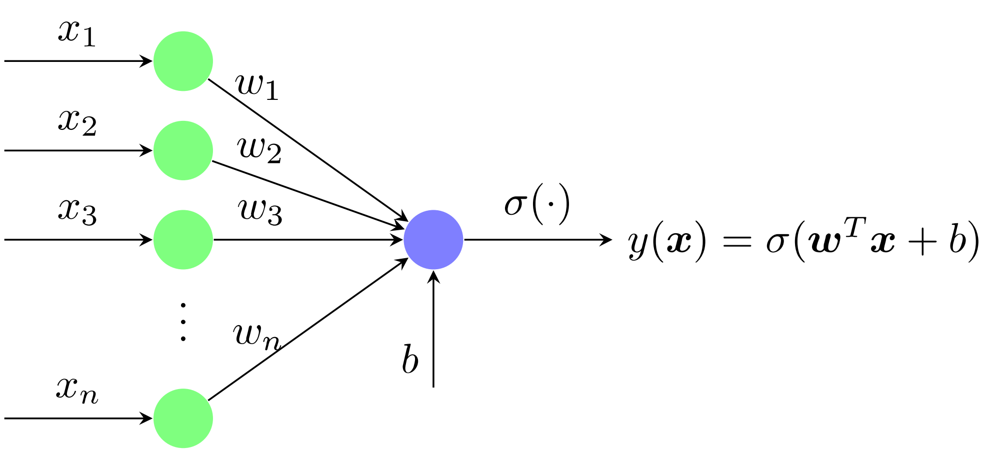
- Are dot products and convolutions the only biologically plausible models of neuronal computation?
- Can we use results and tools from “nonlinear” mathematics to reason about complexity and dimension of learning models in current literature?

Rosenblatt's perceptron

- Introduced in 1943, still prevalent neural model
- Activation: $\phi(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + b$
- Nonlinearity at the output (e.g logistic sigmoid, ReLU):

$$y(\boldsymbol{x}) = \sigma(\phi(\boldsymbol{x}))$$

- Multiply-accumulate architecture → archetypal building block of all architectures (e.g. fully-connected, convolutional etc.)



ICASSP 2024 KOREA

Max-plus Matrix Algebra

- vector/matrix **'addition'** = pointwise max

$$\begin{aligned} \mathbf{x} \vee \mathbf{y} &= [x_1 \vee y_1, \ldots, x_n \vee y_n]^T \\ \mathbf{A} \vee \mathbf{B} &= [a_{ij} \vee b_{ij}] \end{aligned}$$

- vector/matrix **'dual addition'** = pointwise min

$$\begin{aligned} \mathbf{x} \wedge \mathbf{y} &= [x_1 \wedge y_1, \ldots, x_n \wedge y_n]^T \\ \mathbf{A} \wedge \mathbf{B} &= [a_{ij} \wedge b_{ij}] \end{aligned}$$

- vector/matrix **'multiplication by scalar'**

$$\begin{aligned} c + \mathbf{x} &= [c + x_1, \ldots, c + x_n]^T \\ c + \mathbf{A} &= [c + a_{ij}] \end{aligned}$$

- (max, +) **'matrix multiplication'**

$$[\mathbf{A} \boxplus \mathbf{B}]_{ij} = \bigvee_{k=1}^{n} a_{ik} + b_{kj}$$

Matrix Adjoint

$$\mathbf{A}^* \triangleq -\mathbf{A}^T$$

- (min, +) '**matrix dual multiplication**'

$$[\mathbf{A} \boxplus' \mathbf{B}]_{ij} = \bigwedge_{k=1}^{n} a_{ik} + b_{kj}$$

ICASSP 2024 KOREA

Morphological (Max-Plus) Perceptron

- Introduced in the 1990's. Instead of multiply-accumulate, computes a dilation (max-of-sums):

$$\tau(\boldsymbol{x}) = \boldsymbol{w}^T \boxplus \boldsymbol{x} \triangleq \bigvee_{i=1}^{n} w_i + x_i$$

or an erosion:

$$\tau'(\boldsymbol{x}) = \boldsymbol{w}^T \boxplus' \boldsymbol{x} \triangleq \bigwedge_{i=1}^{n} w_i + x_i$$

- Ritter & Urcid (2003): argued about biological plausibility and proved that every compact region in *n*-dim Euclidean space can be approximated by morphological perceptrons to arbitrary accuracy.

- Related to a Maxout unit.

Solve Max-plus Equations via Adjunctions

- **Problems**:

  (1) Exact problem: Solve $\delta_{\mathbf{A}}(\mathbf{x}) = \overbrace{\mathbf{A} \boxplus \mathbf{x}}^{\text{max-plus}} = \mathbf{b}, \quad \mathbf{A} \in \overline{\mathbb{R}}^{m \times n}, \ \mathbf{b} \in \overline{\mathbb{R}}^{m}$

  (2) Approximate Constrained: Min $\|\mathbf{A} \boxplus \mathbf{x} - \mathbf{b}\|_{p=1\ldots\infty}$ s.t. $\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}$

- **Theorem**: The **greatest (sub)solution** of (1) and unique solution of (2) is

$$\hat{\mathbf{x}} = \varepsilon_{\mathbf{A}}(\mathbf{b}) = \mathbf{A}^* \boxplus' \mathbf{b}, \quad [\hat{\mathbf{x}}]_j = \bigwedge_{i=1}^{m} b_i - a_{ij}, \quad \mathbf{A}^* \triangleq -\mathbf{A}^T$$

  and yields the **Greatest Lower Estimate (GLE)** of data $\mathbf{b}$:

$$\delta_{\mathbf{A}}(\varepsilon_{\mathbf{A}}(\mathbf{b})) = \underbrace{\mathbf{A} \boxplus (\underbrace{\mathbf{A}^* \boxplus' \mathbf{b}}_{\text{min-plus}})}_{\text{max-plus matrix product}} \leq \mathbf{b}$$

- **Geometry**: Operators δ, ε are vector dilation and erosion, and the GLE $\mathbf{b} \mapsto \delta(\varepsilon(\mathbf{b}))$ is an opening (lattice projection).

- **Complexity**: $O(mn)$

Feasible Regions & Separability Condition for Max-plus Percepton

Let $\boldsymbol{X} \in \mathbb{R}_{\max}^{k \times n}$ be a matrix containing the patterns to be classified as its rows, let $\boldsymbol{x}^{(k)}$ denote the k-th pattern (row) and let $\mathcal{C}_1, \mathcal{C}_0$ be the two classes

Max-plus perceptron $\tau(\boldsymbol{x}) = \boldsymbol{w}^T \boxplus \boldsymbol{x}$

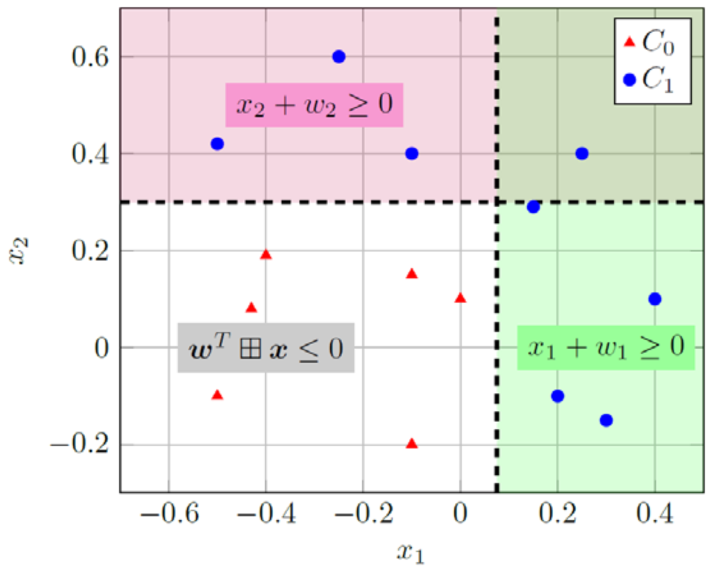
$$\tau(\boldsymbol{x}) = w_0 \vee (w_1 + x_1) \vee \cdots \vee (w_n + x_n) = w_0 \vee \left(\bigvee_{i=1}^{n} w_i + x_i \right)$$

Feasible Region = Tropical Polyhedron

$$\mathcal{T}(\boldsymbol{X}_{\text{pos}}, \boldsymbol{X}_{\text{neg}}) = \{\boldsymbol{w} \in \mathbb{R}_{\max}^{n} : \boldsymbol{X}_{\text{pos}} \boxplus \boldsymbol{w} \geq 0,\ \ \boldsymbol{X}_{\text{neg}} \boxplus \boldsymbol{w} \leq 0\}$$

Separability Condition, equivalent to Nonempty Trop. Polyhedron

$$\mathbf{X}_{\text{pos}} \boxplus (\mathbf{X}_{\text{neg}}^{*} \boxplus' \mathbf{0}) \geq \mathbf{0}$$



[Charisopoulos & Maragos, ISMM 2017]

ICASSP 2024 KOREA

Morphological Neural Nets (MNNs) and Training Approaches

- **Constructive Algorithms**

Dendrite Learning [Ritter & Urcid, 2003], Iterative Partitioning / Competitive Learning [Sussner & Esmi, 2011]: combine (max, +) and (min, +) classifiers, build “bounding boxes” around patterns

- "perfect" fit to data, no concept of outlier

- **Morphological Associative Memories**

Introduce a Hopfield-type network, computing (noniteratively) a morphological/fuzzy response (e.g. Sussner & Valle, 2006):

- **Gradient Descent Variants**

Min-max classifiers [Yang & Maragos, 1995], MRL nodes [Pessoa & Maragos, 2000], Dilation-Erosion Linear Perceptron [Araujo et al. 2012].

- **Recent Approaches:**

Convex-Concave Programming (CCP) for Max-plus Perceptron and DEP (Binary Classification) [Charisopoulos & Maragos 2017]

Reduced Dilation-Erosion Perceptron (r-DEP) trained via CCP for Binary Classification [Valle 2020]

Dense Morphological Networks [Mondal et al. 2019]

Deep Morphological Networks [Franchi et al. 2020]

r-DEP for Multiclass Classification via CCP, L1 Pruning on Dense MNNs [Dimitriadis & Maragos 2021]

A CCP Approach for Training MP on Non-separable Data

Training a (max, +) perceptron can be stated as a difference-of-convex (DC) optimization problem. Solved iteratively (but global optimum not guaranteed) by the Convex-Concave Procedure (**CCP**) [Yuille & Rangarajan 2003], implemented via Disciplined CCP (DCCP - CvxPy) [Shen et al. 2016]

Given a sequence of training data $\{\boldsymbol{x}^k\}_{k=1}^K$:

$$\text{Minimize } J(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{\nu}) = \sum_{k=1}^{K} \nu_k \cdot \max(\xi_k, 0)$$

$$\text{s. t.} \begin{cases} \bigvee_{i=1}^{n} w_i + x_i^{(k)} \leq \xi_k & \text{if } \boldsymbol{x}^{(k)} \in \mathcal{C}_0 \\ \bigvee_{i=1}^{n} w_i + x_i^{(k)} \geq -\xi_k & \text{if } \boldsymbol{x}^{(k)} \in \mathcal{C}_1 \end{cases}$$

Weighted DCCP

Negative target

Positive target

ν_k Some measure of "being outlier“ (e.g. proportional to 1/distance of the k-th pattern from its class centroid)

ξ_k (slack variables) Positive only if misclassification occurs at *k*-th pattern

ICASSP 2024 KOREA

Gradient Descent vs. CCP for Training (max,+) Perceptron

Two Binary Classification Experiments with small datasets,

Ripley (GMM-2) and WBCD (~1k):

Gradient descent with fixed N = 100 epochs vs. CCP using the DCCP toolkit for CvxPy (default parameters).

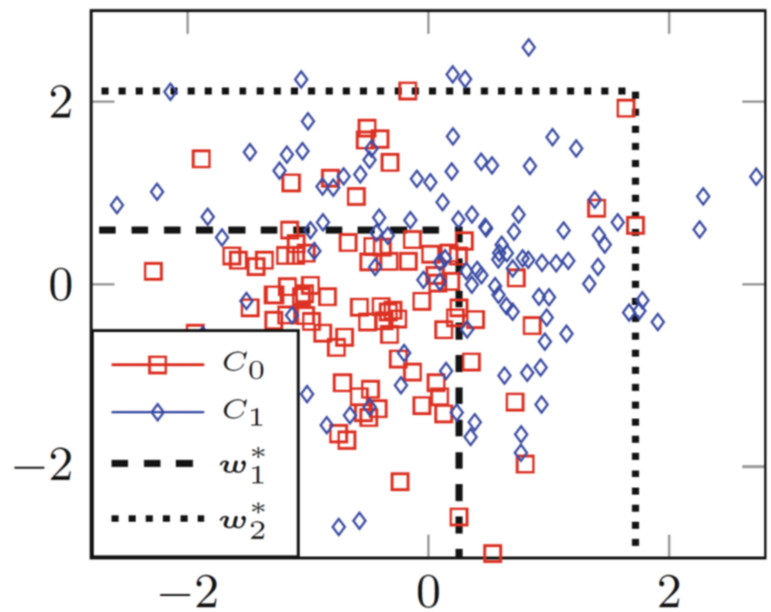
η	Ripleys		WDBC	
	SGD	WDCCP	SGD	WDCCP
0.01	0.838 ± 0.011		0.726 ± 0.002	
0.02	0.739 ± 0.012		0.763 ± 0.006	
0.03	0.827 ± 0.008		0.726 ± 0.004	
0.04	0.834 ± 0.008		0.751 ± 0.007	
0.05	0.800 ± 0.009	**0.902** ± 0.001	0.783 ± 0.012	**0.908** ± 0.001
0.06	0.785 ± 0.008		0.768 ± 0.01	
0.07	0.776 ± 0.009		0.729 ± 0.009	
0.08	0.769 ± 0.01		0.732 ± 0.01	
0.09	0.799 ± 0.009		0.730 ± 0.015	
0.1	0.749 ± 0.011		0.729 ± 0.009	

CCP: more robust results

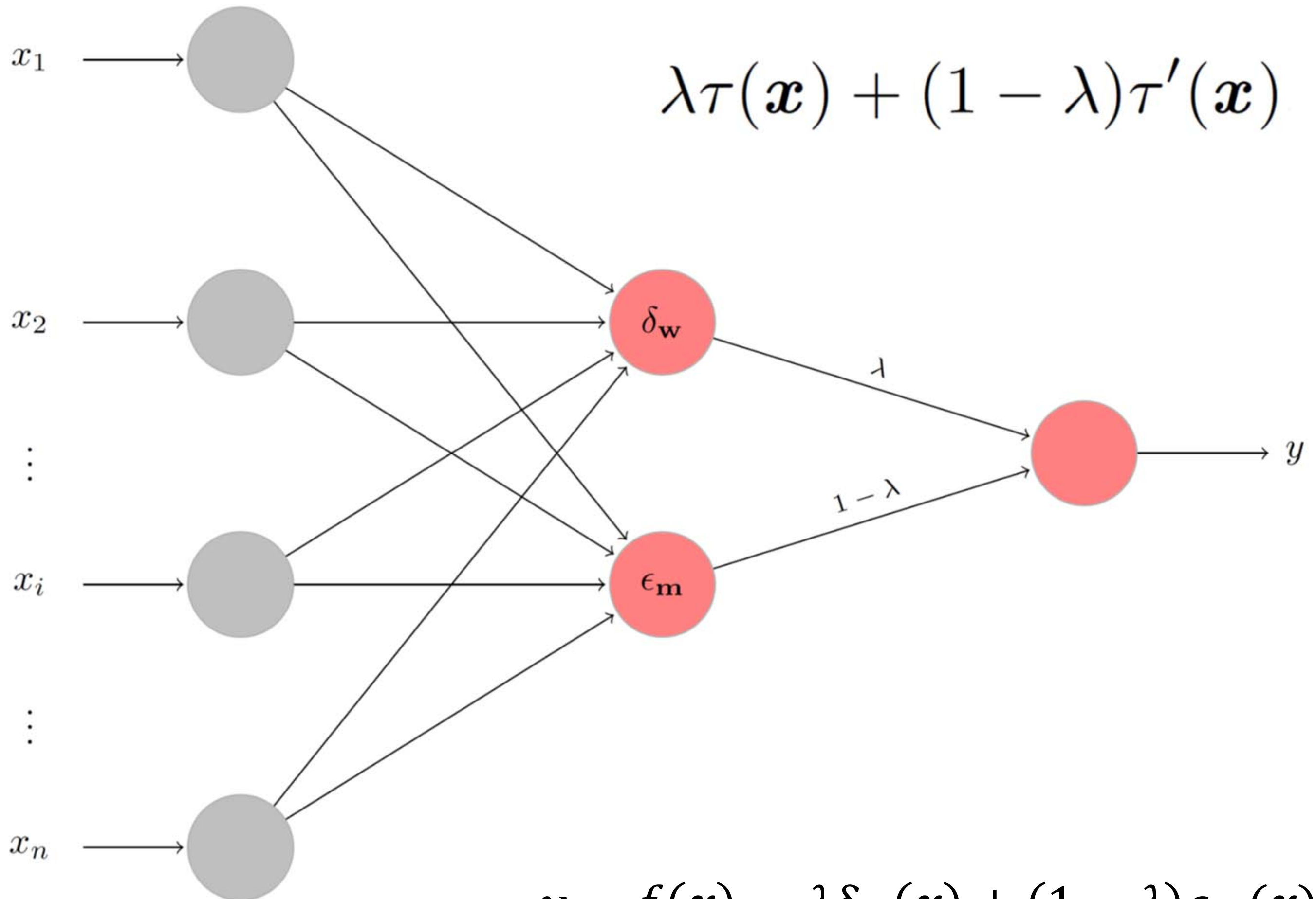
Classification of initially separable Gaussian data with randomly flipped labels 20%:

…… : No regularization (DCCP)

---- : Regularization (Weighted DCCP)



Dilation-Erosion Perceptron (DEP)



$$\lambda\tau(\boldsymbol{x}) + (1-\lambda)\tau'(\boldsymbol{x})$$

$$y = f(\boldsymbol{x}) = \lambda\delta_w(\boldsymbol{x}) + (1-\lambda)\epsilon_m(\boldsymbol{x})$$

ICASSP 2024 KOREA

Dilation-Erosion Perceptron Training

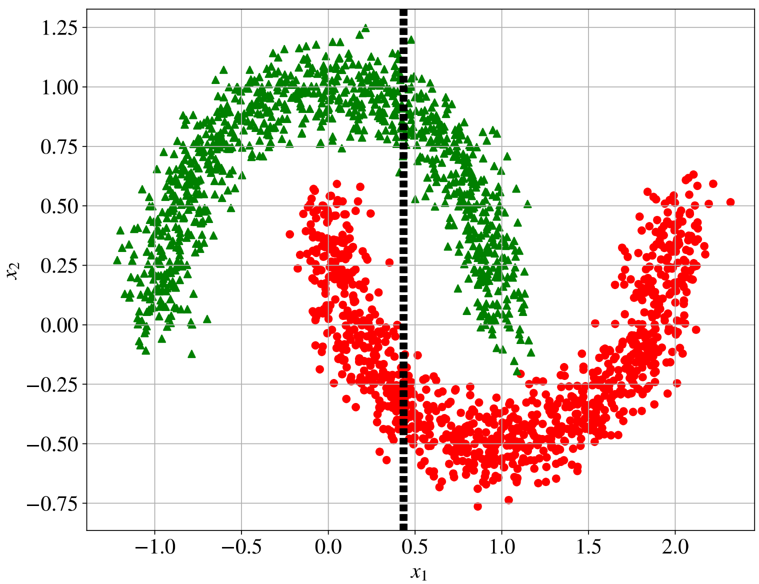
$$y = f(\boldsymbol{x}) = \lambda\delta_w(\boldsymbol{x}) + (1-\lambda)\epsilon_m(\boldsymbol{x}) = \lambda\delta_w(\boldsymbol{x}) - (1-\lambda)[-\epsilon_m(\boldsymbol{x})]$$
$$= convex - (-concave)$$
$$= convex - convex$$

Training as Difference-of-Convex Optimization via Convex-Concave Programming

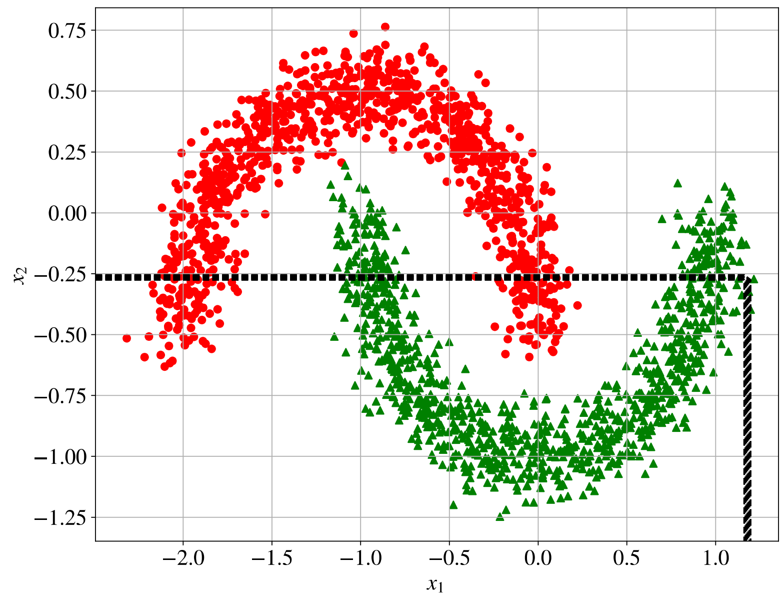
$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{N} v_i \max\{0, \xi_i\} \\ \text{subject to} \quad & \lambda\delta_{\mathrm{w}}(\mathbf{x}_i) + (1-\lambda)\varepsilon_{\mathrm{m}}(\mathbf{x}_i) \geq -\xi_i \quad \forall \mathbf{x}_i \in \mathcal{P}, \\ & \lambda\delta_{\mathrm{w}}(\mathbf{x}_i) + (1-\lambda)\varepsilon_{\mathrm{m}}(\mathbf{x}_i) \leq +\xi_i \quad \forall \mathbf{x}_i \in \mathcal{N} \end{aligned}$$

Effect of $\mathcal{N} \leftrightarrows \mathcal{P}$ and Ordering Vector Data

Double Moons example



Reversed labels



Correct labels

Reduced ordering [Valle 2020] for better ordering feature patterns:

Let V be a nonempty set, $\mathcal{L}$ be a complete lattice and $\rho: V \to \mathcal{L}$ be a surjective mapping.

A reduced ordering is defined as: $\boldsymbol{x} \leq_\rho \boldsymbol{y} \Leftrightarrow \rho(\boldsymbol{x}) \leq \rho(\boldsymbol{y})\ \forall \boldsymbol{x}, \boldsymbol{y} \in V$.

Can be obtained via a supervised training on a set of positive and negative examples.

Experiments: Multiclass r-DEP, CCP training

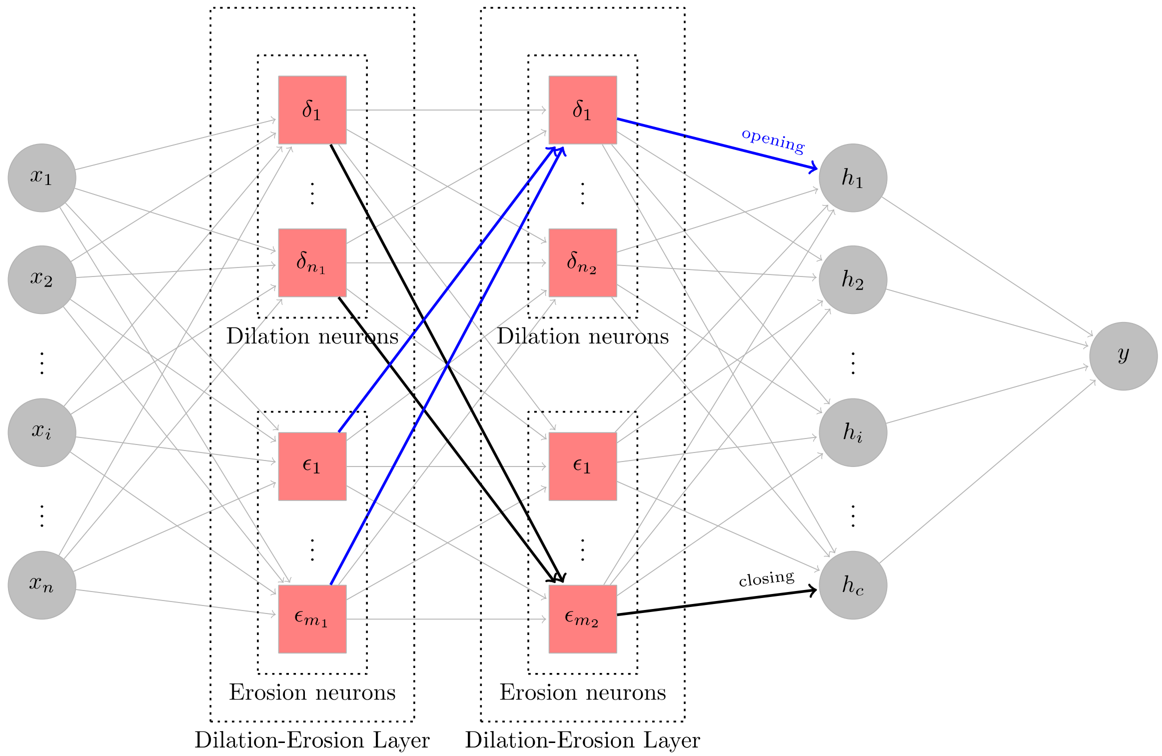
	MNIST	FashionMNIST
$n = 5$	$\mathbf{97.72 \pm 0.01}$	$\mathbf{88.21 \pm 0.01}$
$n = 10$	$\mathbf{97.72 \pm 0.01}$	88.07 ± 0.01
$n = 15$	97.67 ± 0.01	88.11 ± 0.01
$n = 20$	97.64 ± 0.01	88.12 ± 0.01

Table: Results of Bagging *multiclass* r-DEP with n RBF kernels.

- Performance similar to MLP-ReLU architectures trained via SGD
- CCP training is more robust

[Dimitriadis & Maragos 2021]

Dense Morphological Networks



Dense Morphological Network with 2 hidden layers [similar to Mondal et al. 2019]

Focus on Sparsity [Dimitriadis & Maragos 2021] → Apply ℓ_1 Pruning

ICASSP 2024 KOREA

Experiments: Pruning Dense MNN vs MLP-ReLU

		Adaptive Momentum Estimation				Stochastic Gradient Descent			
	p	δ	ε	(δ, ε)	FF-ReLU	δ	ε	(δ, ε)	FF-ReLU
MNIST	100%	97.62	96.17	97.95	98.13	94.86	93.36	96.07	98.16
	75%	97.62	96.18	97.93	98.15	94.86	93.36	96.07	98.12
	50%	97.62	96.22	97.90	98.17	94.86	93.37	96.07	98.08
	25%	97.62	96.09	97.87	97.51	94.86	93.40	96.06	98.01
	10%	97.62	95.78	97.74	93.38	94.86	93.38	96.09	96.67
	7.5%	97.62	95.42	97.76	90.17	94.86	93.38	96.10	95.56
	5%	97.62	94.51	97.66	83.39	94.86	93.40	96.10	92.96
	2.5%	97.62	93.43	97.37	68.93	94.86	93.39	96.09	80.48
	1%	97.62	91.17	97.08	44.22	94.86	93.38	96.08	58.07
FashionMNIST	100%	86.31	86.82	88.32	88.82	82.06	85.23	86.21	87.79
	75%	86.30	86.81	88.30	88.88	82.00	85.23	86.21	87.75
	50%	86.22	86.80	88.33	88.18	82.05	85.25	86.20	87.19
	25%	85.95	86.85	88.31	82.15	81.90	85.26	86.28	84.35
	10%	85.58	86.27	88.05	65.89	81.67	85.27	86.23	73.22
	7.5%	85.47	86.15	87.99	57.93	81.63	85.27	86.21	63.95
	5%	85.37	85.81	87.76	49.12	81.52	85.24	86.22	47.73
	2.5%	84.91	85.47	87.56	42.48	81.14	85.26	86.22	38.84
	1%	81.14	84.86	86.85	28.13	80.68	85.27	86.18	35.46

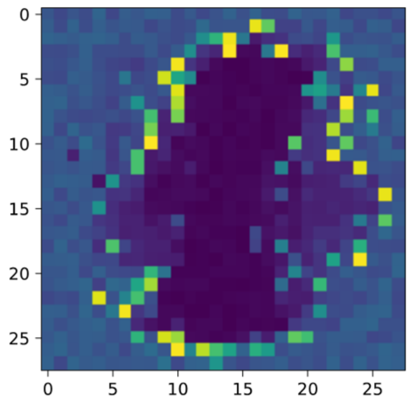
Table: Accuracy of pruned networks on the MNIST and FashionMNIST datasets.

Models: $\delta \rightarrow$ only dilation neurons, $\varepsilon \rightarrow$ only erosion, $(\delta, \varepsilon) \rightarrow$ split equally, FF-ReLU $\rightarrow$ FeedForward NN with ReLU.

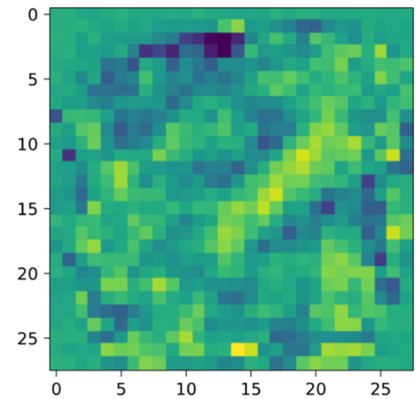
shades of red showcase the degree of (severe) deterioration in accuracy green indicates the absence of performance loss

ICASSP 2024 KOREA

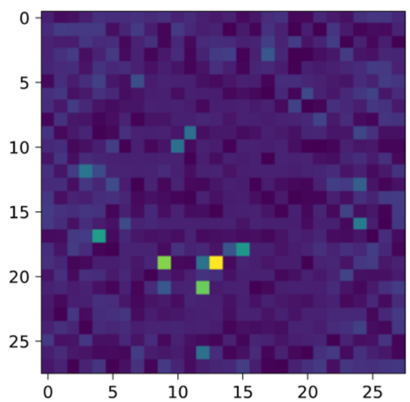
Qualitative Perspectives on Sparsity



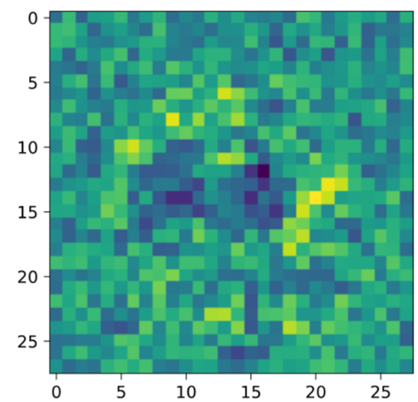
$(\delta, \epsilon) - Adam$



FF−ReLU $- Adam$



$(\delta, \epsilon) - SGD$



FF-ReLU $- SGD$

Examples of hidden layer activations for various NN models (MNIST dataset)

Minimization of Neural Nets via Tropical Division

References:

- G. Smyrnis, P. Maragos and G. Retsinas, "*MaxPolynomial Division With Application to Neural Network Simplification*", Proc. ICASSP 2020.
- G. Smyrnis and P. Maragos, "*Multiclass Neural Network Minimization Via Tropical Newton Polytope Approximation*", Proc. ICML 2020.

ICASSP 2024 KOREA

Tropical Polynomials

Tropical Semiring $(\mathbb{R}_{\max}, \vee, +)$

$$\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$$

$$a \vee b = \max(a, b)$$

$$a + b = a + b$$

Tropical Polynomials

Real coefficients

$$f(\boldsymbol{x}) = \max_{i \in [n]} \{\boldsymbol{a}_i^T \boldsymbol{x} + b_i\}$$

ICASSP 2024 KOREA

Newton Polytopes

Newton Polytopes

$$\text{Newt}\,(f) = \text{conv}\,\{\boldsymbol{a}_i : i \in [n]\}$$

$$\text{ENewt}\,(f) = \text{conv}\,\{(\boldsymbol{a}_i, b_i) : i \in [n]\}$$

Polytope computation

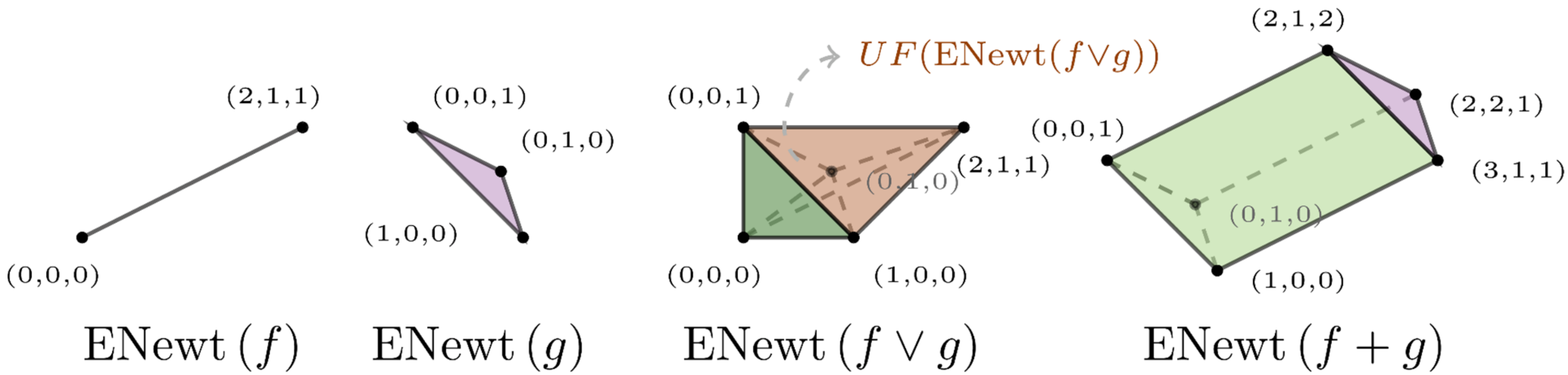
$$\text{ENewt}\,(f \vee g) = \text{conv}\{\text{ENewt}\,(f) \cup \text{ENewt}\,(g)\}$$

$$\text{ENewt}\,(f + g) = \text{ENewt}\,(f) \oplus \text{ENewt}\,(g)$$

Example: Polytope Computation

$$f(x, y) = \max(2x + y + 1, 0)$$

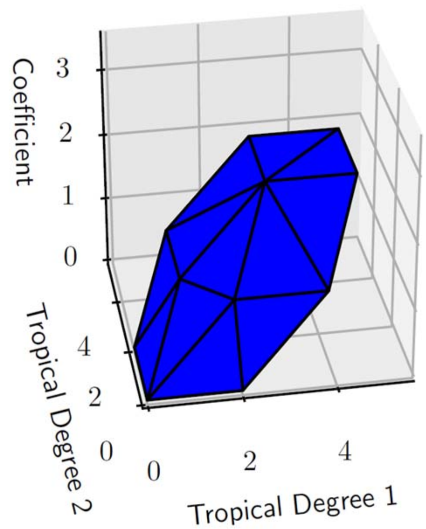
$$g(x.y) = \max(x, y, 1)$$



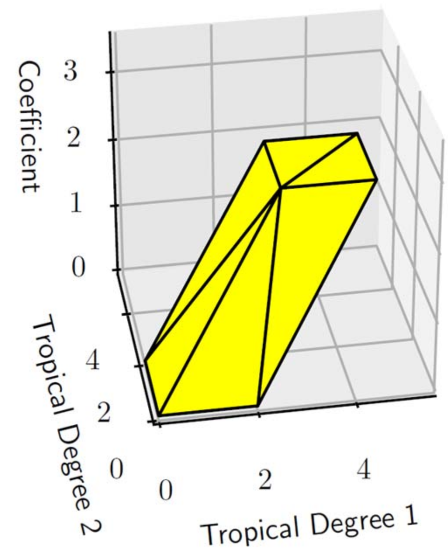
$$f \vee g = \max(2x + y + 1, 0, x, y, 1)$$

$$f + g = \max(x, y, 1, 3x + y + 1, 2x + 2y + 1, 2x + y + 2)$$

General idea for Geometric NN Minimization



Original Network Polytope



Approximate Network Polytope

Max-polynomial Division

Problem: Assume we have two max-polynomials $p(\boldsymbol{x}), d(\boldsymbol{x})$ (dividend and divisor). We want to find two max-polynomials $q(\boldsymbol{x}), r(\boldsymbol{x})$ (quotient and remainder) such that:

$$p(\boldsymbol{x}) = \max(q(\boldsymbol{x}) + d(\boldsymbol{x}), r(\boldsymbol{x}))$$

However! The above is not always feasible (non-trivially).

Approximate Division: We relax the requirements, so that the polynomials we want to find satisfy:

$$p(\boldsymbol{x}) \geq \max(q(\boldsymbol{x}) + d(\boldsymbol{x}), r(\boldsymbol{x}))$$

We also require that $q(\boldsymbol{x}), r(\boldsymbol{x})$ satisfy the above maximally.

ICASSP 2024 KOREA

Algorithm for Approximate Maxpolynomial Division

1. Let C be the set of possible vectors $\boldsymbol{c}$ by which we can h-shift $\mathrm{Newt}(d)$ (each of which corresponds to a linear term in q).
2. We raise the shifted version of $\mathrm{ENewt}(d)$ as high as possible so that it still lies below $\mathrm{ENewt}(p)$, and we mark the vertical shift as q_c.
3. We set the quotient equal to:

$$q(\boldsymbol{x}) = \max_{\boldsymbol{c}\in C}(q_{\boldsymbol{c}} + \boldsymbol{c}^T\boldsymbol{x})$$

and add all terms not covered by a h-shift $\boldsymbol{c}$ to the remainder $r(\boldsymbol{x})$.

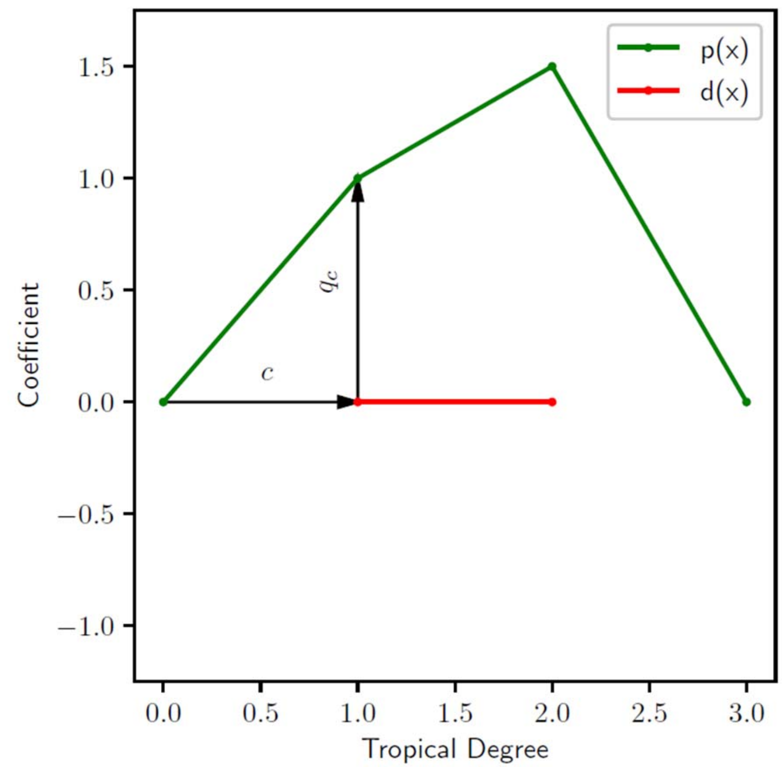


Figure: Division Method
Division of $p(x) = \max(3x, 2x + 1.5, x + 1, 0)$ by $d(x) = \max(x, 0)$.

ICASSP 2024 KOREA

Division Example (1)

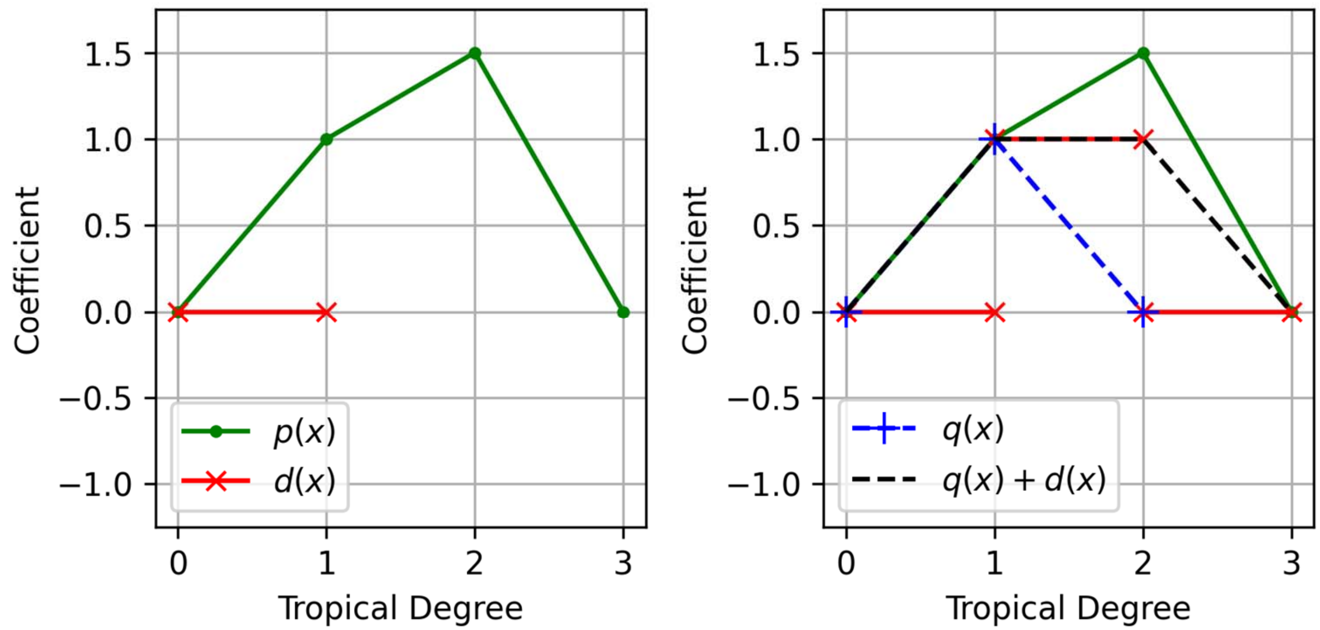


Figure: Division of $p(x) = \max(3x, 2x + 1.5, x + 1, 0)$ by $d(x) = \max(x, 0)$.

Note: The Newton Polytope of the divisor is raised as much as possible, but it cannot match the polytope of the dividend exactly. Thus, only 3 out of the 4 vertices are perfectly matched.

Division Example (2)

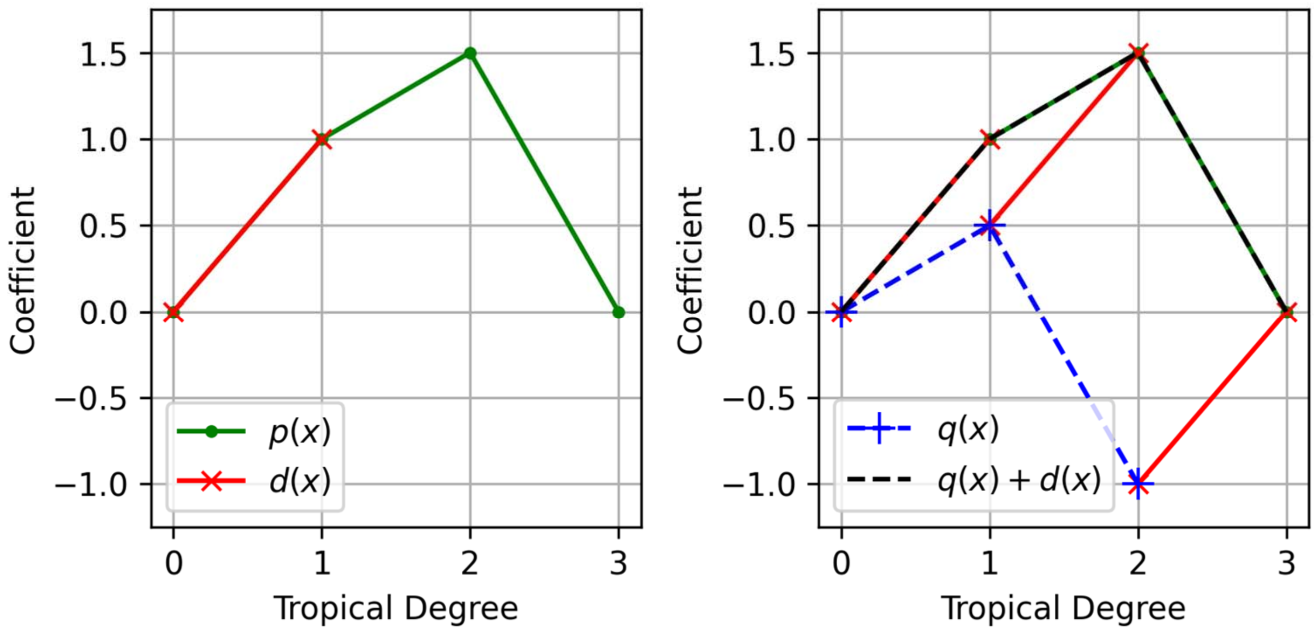


Figure: Division of $p(x) = \max(3x, 2x + 1.5, x + 1, 0)$ by $d(x) = \max(x + 1, 0)$.

Note: In this case, the polytope of the divisor can match that of the dividend perfectly, so all vertices are covered.

ICASSP 2024 KOREA

Application to Neural Network Minimization

General idea: Our algorithm seeks to minimize the network by matching the most important vertices of the ENewton Polytopes of its maxpolynomials.

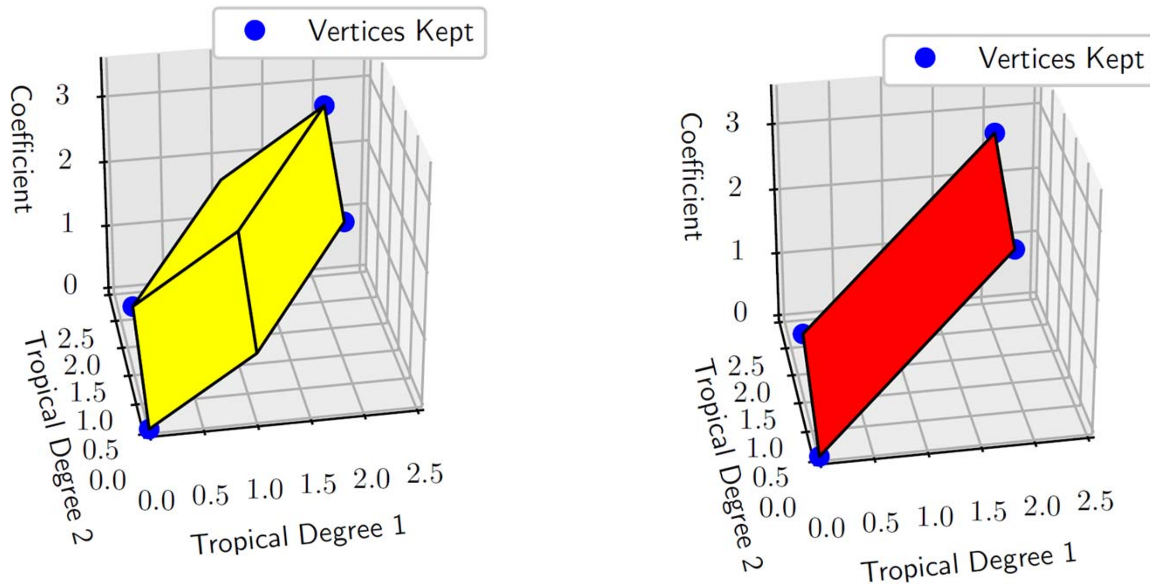
2-layer 1-output NN:

The NNs considered are the difference of two maxpolynomials.

For each of the two (+,-) maxpolynomials $p(\boldsymbol{x})$ of the network, we first find a divisor $d(\boldsymbol{x})$. This is done by:

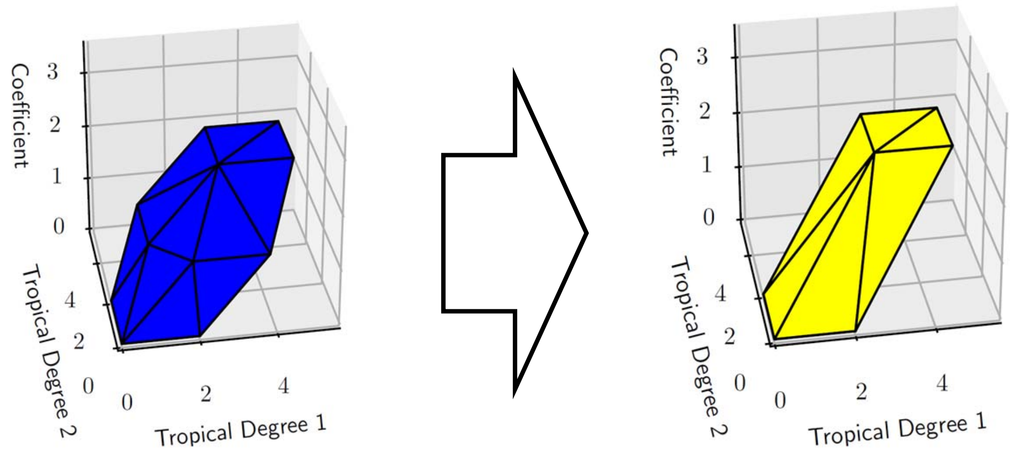
Finding the most important vertices of $\mathrm{ENewt}(p)$, via the weights of the network (based on which combination of neurons is activated).

Method for Single Output Neuron



- Final polytope (right) is precisely under the original (left).
- The process is a “smoothing” of the original polytope.
  (From the 8 vertices of the original-yellow polytope we keep only the 4 blue which comprise the vertices of the final-red polytope.)

Properties of Trop. Div. Approximation Method



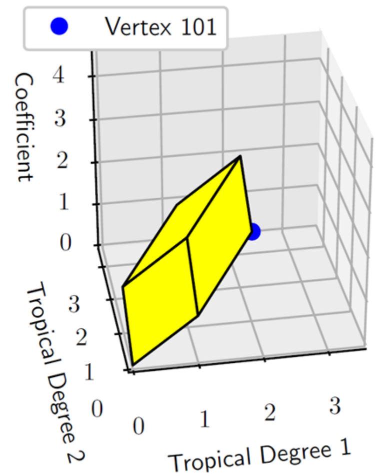
Original Network Polytope *Approximate Network Polytope*

1. Approximate polytope contains only vertices of the original.
2. The input samples activating the chosen vertices have the same output in the two networks.
3. At least $\frac{N}{\sum_{j=0}^{d}\binom{n}{j}} O(\log n')$ samples retain their output

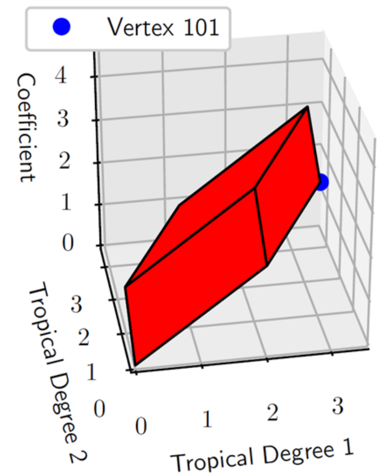
(N is # of samples, n and n' the # of neurons in hidden layer before and after the approximation). Note: this is not a tight bound.

ICASSP 2024 KOREA

Extension with Multiple Output Neurons



Upper hull of polytope, Neuron 1

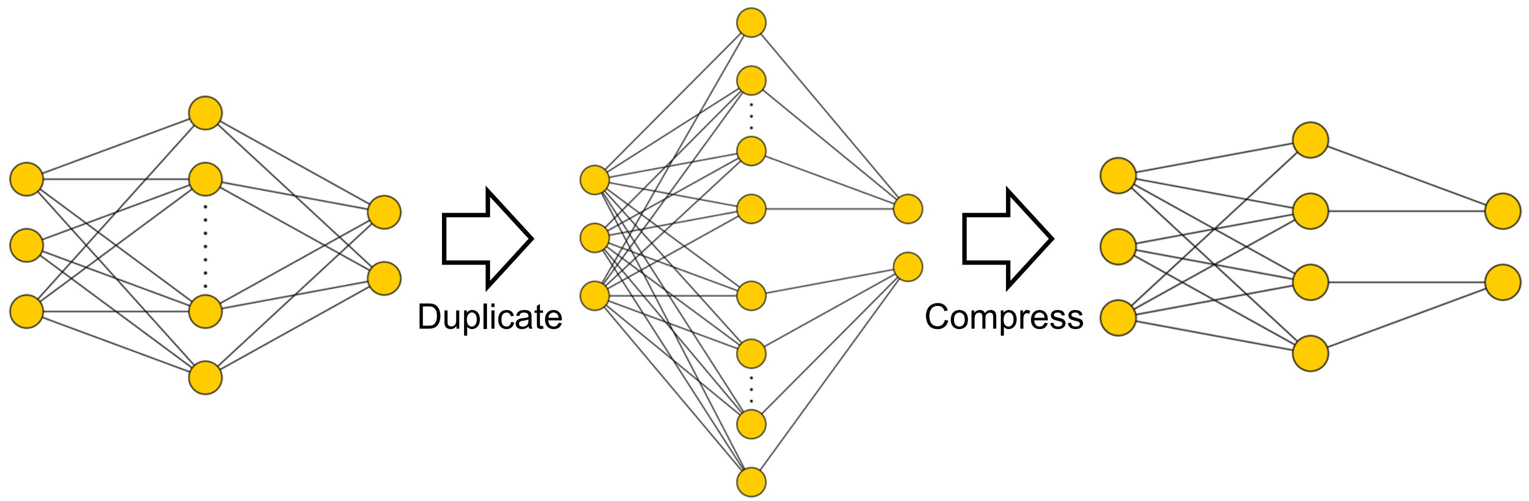


Upper hull of polytope, Neuron 2

- What we have: Multiple polytopes (one pair for each output neuron), interconnected (Minkowski sums of same hidden neurons but with different scaling weights).
- What we want: Simultaneous approximation of all polytopes.

ICASSP 2024 KOREA

Trop. Div. method for Multiple Outputs: One-Vs-All Framework



Experiments: Trop. Division NN Minimization

MNIST Dataset

Neurons Kept	TropDiv Method, Avg. Accuracy	TropDiv Method, St. Deviation
Original	98.604	0.027
75%	**96.560**	**1.245**
50%	**96.392**	**1.177**
25%	**95.154**	**2.356**
10%	**93.748**	**2.572**
5%	92.928	2.589

Fashion-MNIST Dataset

Neurons Kept	TropDiv Method, Avg. Accuracy	TropDiv Method, St. Deviation
Original	88.658	0.538
75%	**83.556**	2.885
50%	**83.300**	2.799
25%	**82.224**	2.845
10%	**80.430**	3.267

ICASSP 2024 KOREA

[G. Smyrnis & P. Maragos, "*Multiclass Neural Net Minimization, Tropical Newton Polytope Approximation*", ICML 2020]

Revisiting Tropical Division

Reference:

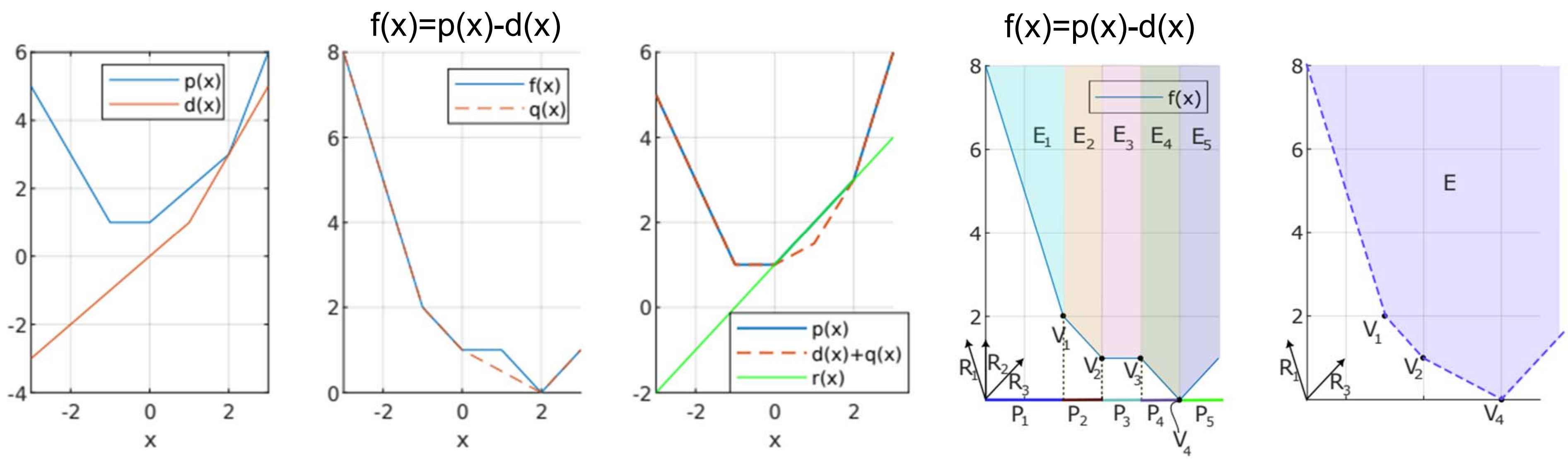
- I. Kordonis and P. Maragos, "*Revisiting Tropical Polynomial Division: Theory, Algorithms and Application to Neural Networks*", ArXiv 2023

ICASSP 2024 KOREA

Tropical Division: Real Coefficients

- There is a unique quotient-remainder pair (in the case of real coefficients)
- Exact Algorithm

  Compute the best convex underapproximation of the difference between the dividend and the divisor



Tropical Division: Approximate Algorithm

Alternating Minimization Scheme

- *Phase 1*: Partition the data $\boldsymbol{x}_j$ into subsets J_i where affine term i is maximal in the quotient
- *Phase 2*: Maximize

$$\begin{aligned} \underset{\boldsymbol{a}_i, b_i}{\text{maximize}} \quad & \sum_{i=1}^{m} \sum_{j \in J_i} \boldsymbol{a}_i^T \boldsymbol{x}_j + b_i \\ \text{subject to} \quad & \boldsymbol{a}_i^T \boldsymbol{x}_j + b_i \leq f(\boldsymbol{x}_j), \\ & \boldsymbol{a}_i \in C \end{aligned}$$

Search for a quotient in the form

$$q(\boldsymbol{x}) = \bigvee_{i=1}^{m} \boldsymbol{a}_i^T \boldsymbol{x}_j + b_i$$

Set of data, for which term i is maximal

$$\boldsymbol{a}_i^T \boldsymbol{x}_j + b_i \geq \boldsymbol{a}_{i'}^T \boldsymbol{x}_j + b_{i'}$$

Set of slopes $\boldsymbol{a}$ such that

$$\mathrm{Newt}(d) \oplus \{\boldsymbol{a}\} \subset \mathrm{Newt}(p)$$

- Iterate

The optimization problem of Phase 2 is a linear program

Results

- Solutions closer to the actual quotient
- Tested in simple datasets, gives improved results for neural network compression

ICASSP 2024 KOREA

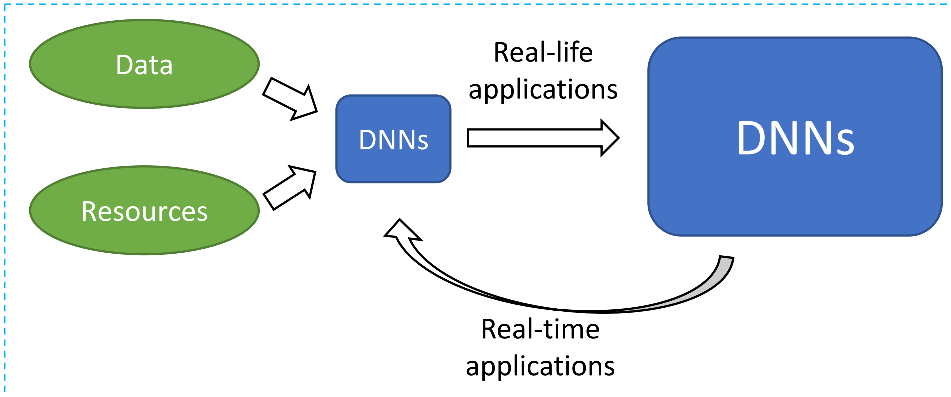
Minimization of Neural Nets via Newton Polytope Approximation

Reference:

- P. Misiakos, G. Smyrnis, G. Retsinas and P. Maragos, "*Neural Network Approximation based on Hausdorff distance of Tropical Zonotopes*", Proc. ICLR 2022.

ICASSP 2024 KOREA

Neural Network Compression



SoA architectures improve accuracy by adding complexity!

✓ *e.g. Increasing depth/width/connectivity*

Optimize/compress a model with respect to:

■ **#parameters** ■ **FLOPS**

■ *memory footprint* ■ *parallelization*

Solutions:
Bottleneck layers, Shared Weights, Tensor Decomposition, *Quantization, Pruning/Sparsification*

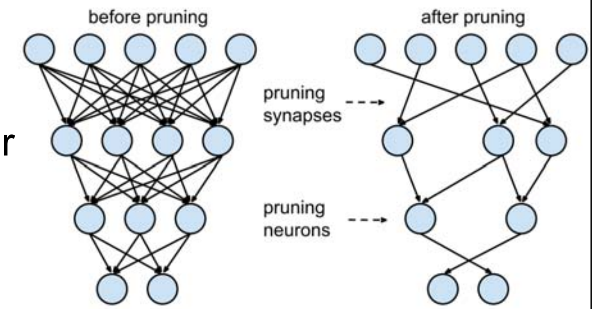
Pruning: Find weights/neurons with the least contribution

✓ Pruning individual weights vs channels/neurons

Two notable approaches:

- Minimum magnitude
- Minimum inducing error

Iterative process:
1) Prune 2) Re-train



S. Han et al. "Learning both weights and connections for efficient neural network", NIPS 2015

Pruning via Zonotope Approximation

P. Misiakos,…, P. Maragos, "Neural Network Approximation based on Hausdorff Distance of Tropical Zonotopes", ICLR, 2022

Approximately equal polytopes ⇒ *Approximately* equivalent polynomials

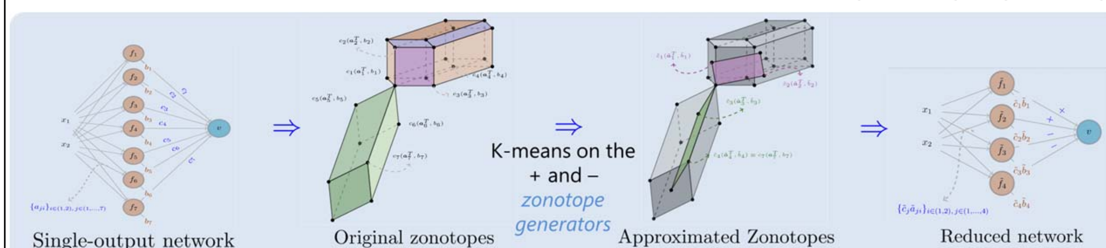
ReLU NNs ≡ **Tropical rational maps**
[*Zhang et al., 2018*]

Polynomials & Polytopes equivalence
[*Charisopoulos and Maragos, 2018*]

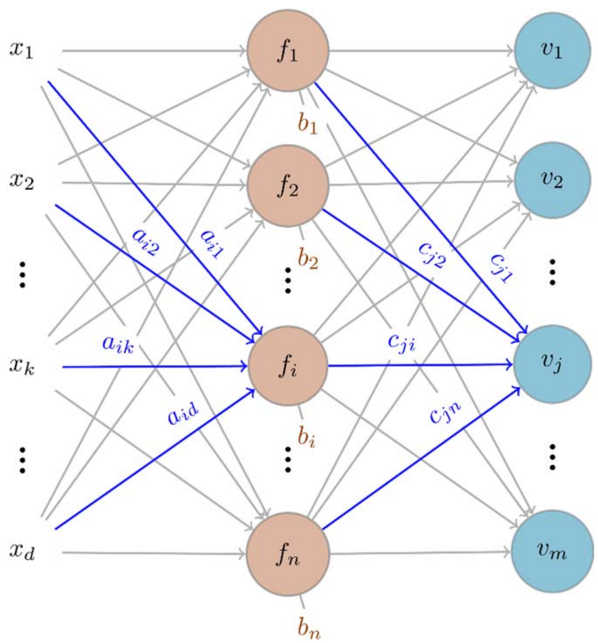
Theorem: NN with 1 hidden layer *Hausdorff distance of zonotopes*

$$\max_{x\in\mathcal{B}} \|v(\boldsymbol{x}) - \tilde{v}(\boldsymbol{x})\|_1 \le \rho\cdot\left(\sum_{j=1}^{m} \mathcal{H}\left(P_j,\tilde{P}_j\right) + \mathcal{H}\left(Q_j,\tilde{Q}_j\right)\right)$$

Positive and negative *zonotopes*: $P_j = \mathrm{ENewt}(p_j)$ $Q_j = \mathrm{ENewt}(q_j)$



Neural Network Tropical Geometry: Polynomials



1 hidden layer with ReLU activations

$i-$ th hidden layer neuron

$$f_i(\boldsymbol{x}) = \max\left(\boldsymbol{a}_i^T \boldsymbol{x} + b_i, 0\right)$$

Tropical *polynomial*

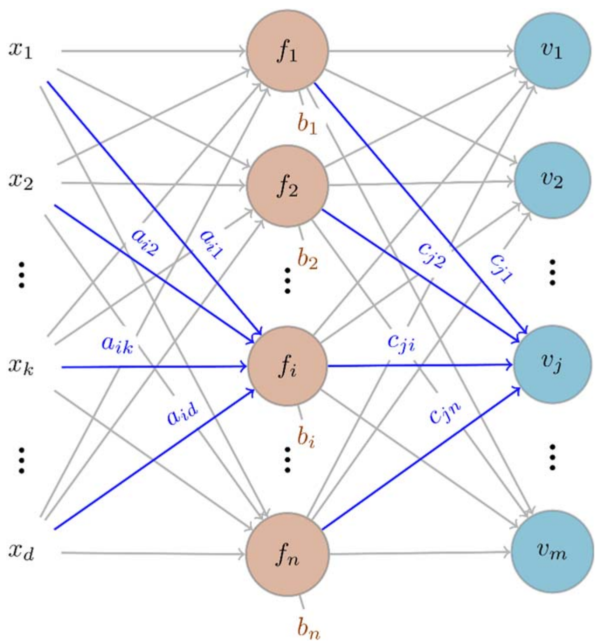
$j-$ th output layer neuron

$$\begin{aligned} v_j(\boldsymbol{x}) &= \sum_{i=1}^{n} c_{ji} f_i(x) \\ &= \sum_{c_{ji}>0} |c_{ji}| f_i(\boldsymbol{x}) - \sum_{c_{ji}<0} |c_{ji}| f_i(\boldsymbol{x}) \\ &= p_j(\boldsymbol{x}) - q_j(\boldsymbol{x}) \end{aligned}$$

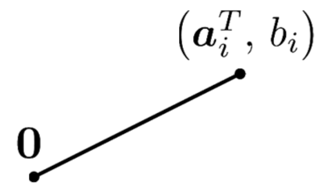
Tropical *rational* function

ICASSP 2024 KOREA

Neural Network Tropical Geometry: Polytopes



$$f_i(\boldsymbol{x}) = \max\left(\boldsymbol{a}_i^T \boldsymbol{x} + b_i, 0\right)$$



$\text{ENewt}\left(f_i\right)$ *is a linear segment*

$$v_j(\boldsymbol{x}) = \sum_{c_{ji}>0} |c_{ji}| f_i(\boldsymbol{x}) - \sum_{c_{ji}<0} |c_{ji}| f_i(\boldsymbol{x})$$
$$= p_j(\boldsymbol{x}) - q_j(\boldsymbol{x})$$

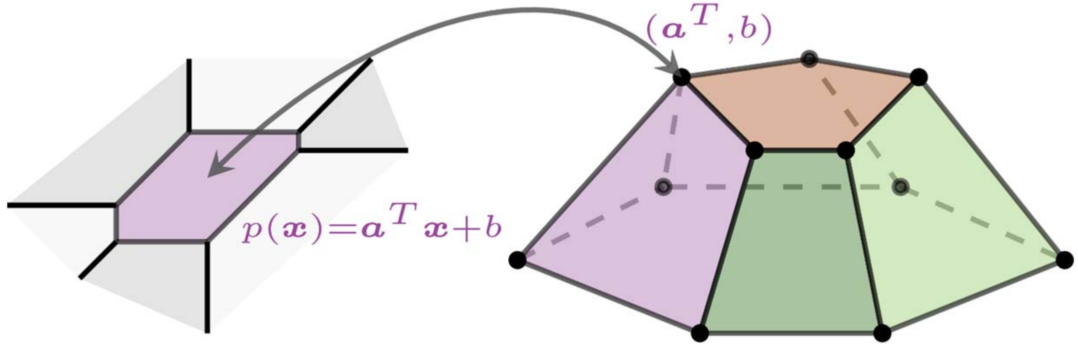
$P_j = \text{ENewt}\left(p_j\right)$

$Q_j = \text{ENewt}\left(q_j\right)$

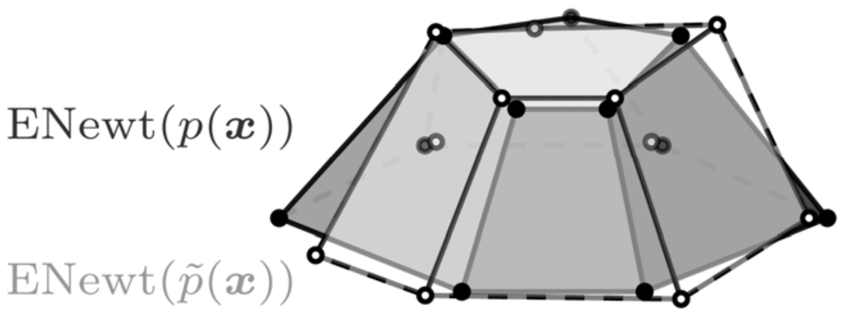
Positive and Negative *zonotopes – or polytopes for deeper NNs*

$c_{ji}\left(\boldsymbol{a}_i^T,\, b_i\right)$ *Generators* of the zonotopes

Approximate Extended Newton Polytopes



linear regions $\overset{1-1}{\longleftrightarrow}$ *vertices of the upper envelope of the extended Newton polytope*



$\overset{?}{\Longrightarrow}$ $p(\boldsymbol{x}) \approx \tilde{p}(\boldsymbol{x})$

Approximate extended Newton polytopes

Approximate tropical polynomials

Approximating Tropical Polynomials

Proposition Let $p, \tilde{p} \in \mathbb{R}_{\max}[\boldsymbol{x}]$ and consider the polytopes $P = \mathrm{ENewt}(p)$, $\tilde{P} = \mathrm{ENewt}(\tilde{p})$. Then,

$$\max_{x \in \mathcal{B}} |p(\boldsymbol{x}) - \tilde{p}(\boldsymbol{x})| \leq \rho \cdot \mathcal{H}\left(P, \tilde{P}\right)$$

Hausdorff distance of polytopes

ICASSP 2024 KOREA

Neural Network Approximation Theorem

Theorem: Consider two neural networks $v, \tilde{v}$ with output size m and $P_j, Q_j, \tilde{P}_j, \tilde{Q}_j$ be the positive and negative extended Newton polytopes of $v, \tilde{v}$ respectively. Then,

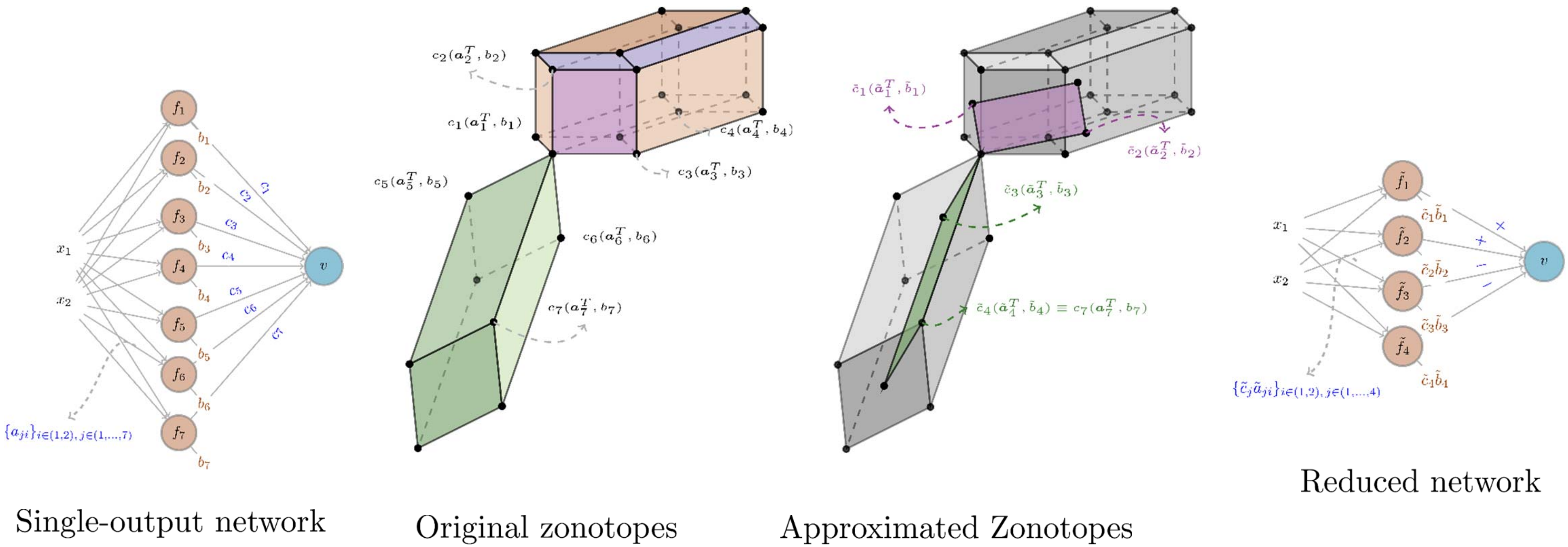
$$\max_{x \in \mathcal{B}} \|v(\boldsymbol{x}) - \tilde{v}(\boldsymbol{x})\|_1 \leq \rho \cdot \left(\sum_{j=1}^{m} \mathcal{H}\left(P_j, \tilde{P}_j\right) + \mathcal{H}\left(Q_j, \tilde{Q}_j\right) \right)$$

Approximately equal polytopes $\Rightarrow$ *Approximately* equivalent networks

ICASSP 2024 KOREA

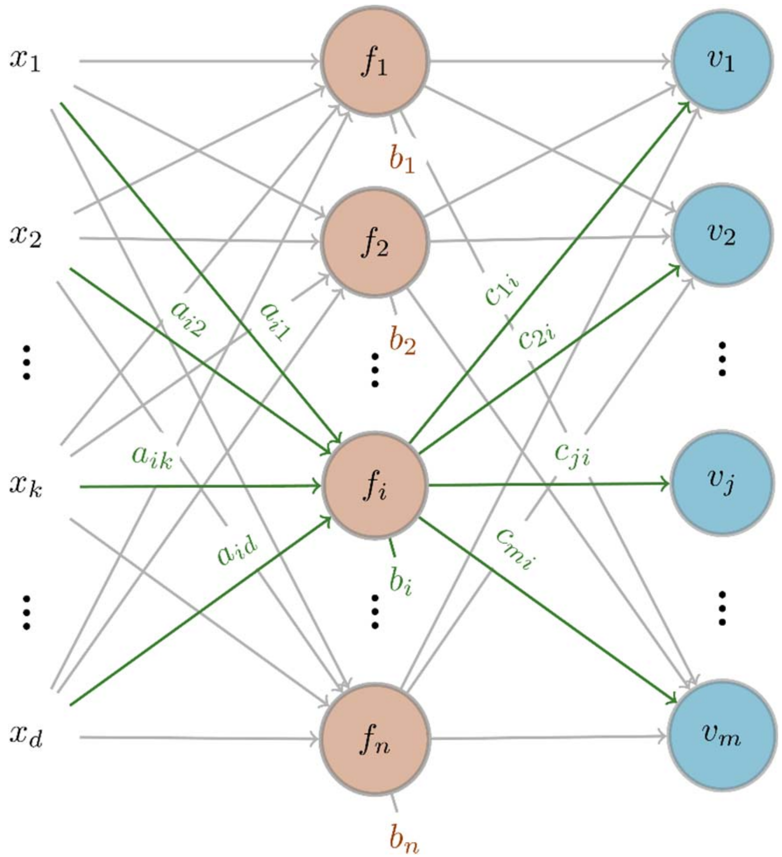
Zonotope K-Means

K-means on the positive and negative *zonotope generators*



ICASSP 2024 KOREA

Neural Path K-means



Generalization for multi-output networks

K-means on the vectors associated with the *neural paths*

Performance Results: Comparison with tropical division

Binary Classification Experiments

Percentage of Remaining Neurons	MNIST 3/5			MNIST 4/9		
	Smyrnis et al., 2020	Zonotope K-means	Neural Path K-means	Smyrnis et al., 2020	Zonotope K-means	Neural Path K-means
100% (Original)	99.18 ± 0.27	99.38 ± 0.09	99.38 ± 0.09	99.53 ± 0.09	99.53 ± 0.09	99.53 ± 0.09
5%	99.12 ± 0.37	99.42 ± 0.07	99.25 ± 0.04	98.99 ± 0.09	99.52 ± 0.09	99.48 ± 0.15
1%	99.11 ± 0.36	99.39 ± 0.05	99.32 ± 0.03	99.01 ± 0.09	99.46 ± 0.05	99.35 ± 0.17
0.5%	99.18 ± 0.36	99.41 ± 0.05	99.22 ± 0.11	98.81 ± 0.09	99.35 ± 0.24	98.84 ± 1.18
0.3%	99.18 ± 0.36	99.25 ± 0.37	99.19 ± 0.41	98.81 ± 0.09	98.22 ± 1.38	98.22 ± 1.33

Multiclass Classification Experiments

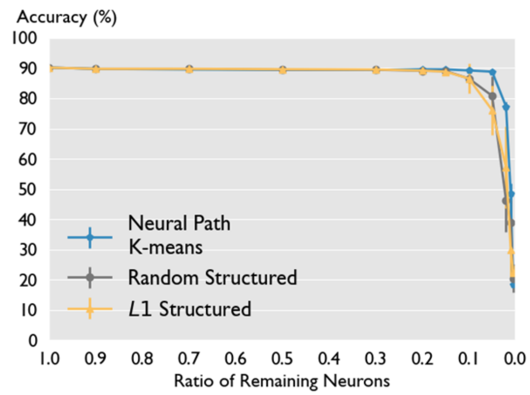
Percentage of Remaining Neurons	MNIST		Fashion-MNIST	
	Smyrnis and Maragos, 2020	Neural Path K-means	Smyrnis and Maragos, 2020	Neural Path K-means
100% (Original)	98.60 ± 0.03	98.61 ± 0.11	88.66 ± 0.54	89.52 ± 0.19
50%	96.39 ± 1.18	98.13 ± 0.28	83.30 ± 2.80	88.22 ± 0.32
25%	95.15 ± 2.36	98.42 ± 0.42	82.22 ± 2.85	86.67 ± 1.12
10%	93.48 ± 2.57	96.89 ± 0.55	80.43 ± 3.27	86.04 ± 0.94
5%	92.93 ± 2.59	96.31 ± 1.29	–	83.68 ± 1.06

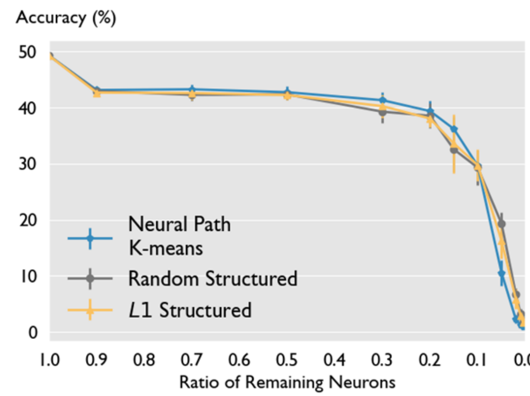
[P. Misiakos, G. Smyrnis, G. Retsinas and P. Maragos, "*Neural Network Approximation based on Hausdorff Distance of Tropical Zonotopes*", Proc. ICLR 2022]

Comparison with Baselines

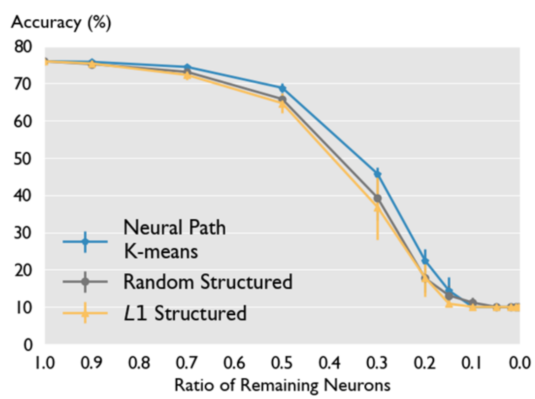
CIFAR10 | CIFAR100

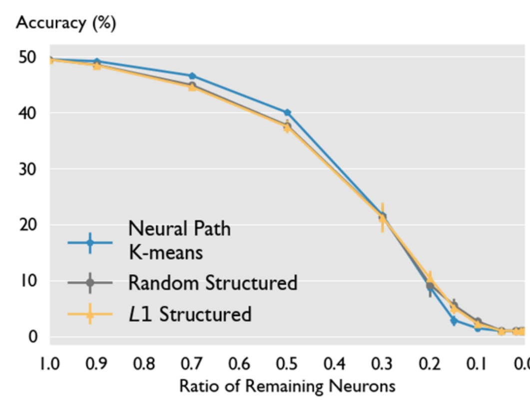
CIFAR-VGG



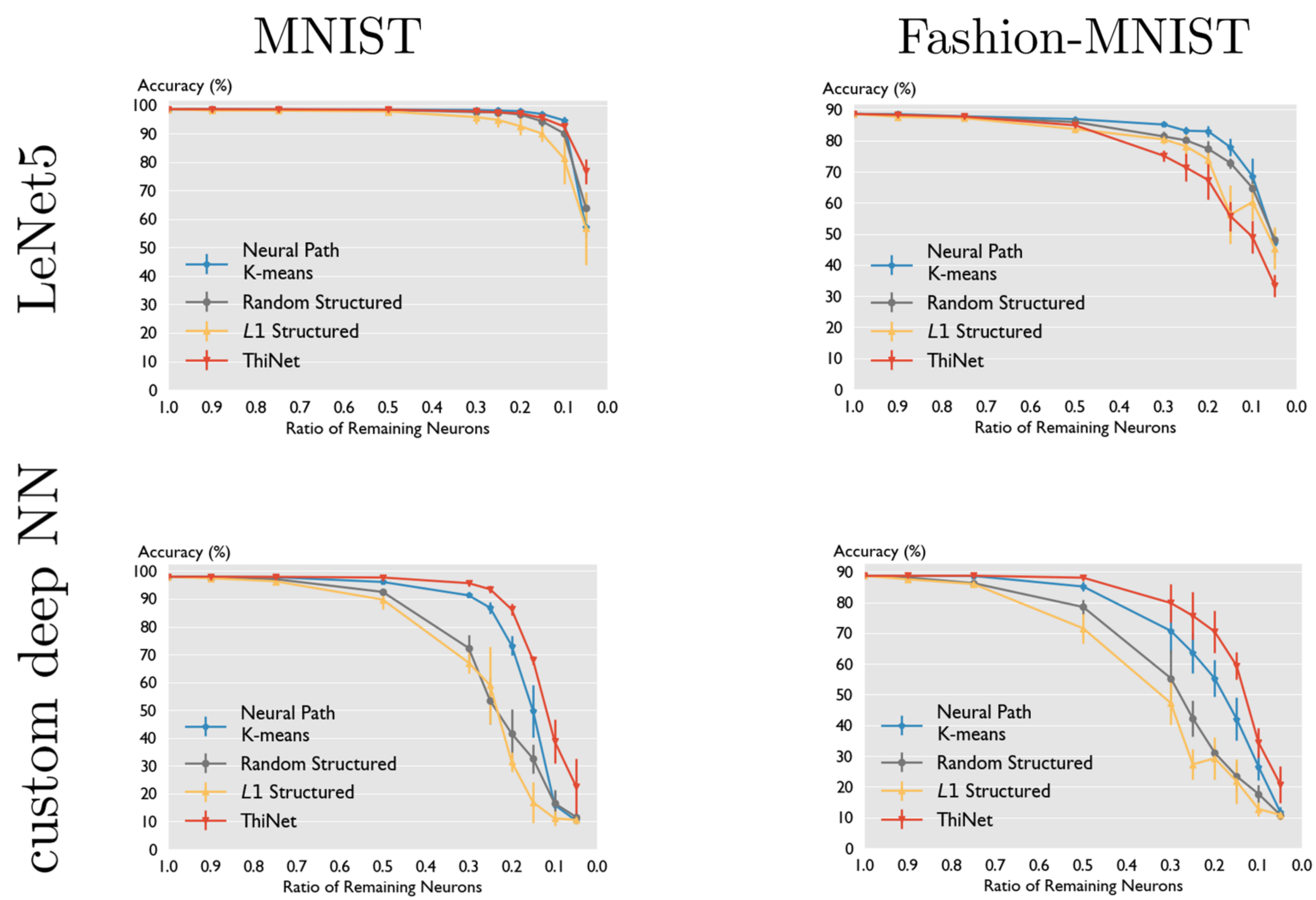


AlexNet





Comparison with ThiNet and Baselines



ICASSP 2024 KOREA

Optimal Solutions of Max-Plus Equations and Sparsity

References:

- R. Cuninghame-Green, *Minimax Algebra*, Springer-Verlag, 1979.
- P. Butkovic, *Max-Linear Systems: Theory and Algorithms*, Springer-Verlag, 2010.
- A. Tsiamis and P. Maragos, "*Sparsity in Max-plus Algebra*", Discrete Events Dynamic Systems, 2019.
- N. Tsilivis, A. Tsiamis and P. Maragos, "*Sparsity in Max-plus Algebra And Applications in Multivariate Convex Regression*", ICASSP, 2021.
- N. Tsilivis, A. Tsiamis and P. Maragos, "*Sparse Approximate Solutions to Max-Plus Equations*", Int'l Conf. Discrete Geometry and Mathematical Morphology, 2021.
- N. Tsilivis, A. Tsiamis and P. Maragos, "*Toward a Sparsity Theory on Weighted Lattices*", Journal of Mathematical Imaging and Vision, 2022.

Solve Max-plus Equations

- **Problems**:
  (1) Exact problem: Solve $\delta_A(\mathbf{x}) = \mathbf{A} \boxplus \mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \overline{\mathbb{R}}^{m\times n}, \ \mathbf{b} \in \overline{\mathbb{R}}^m$

  (2) Approximate Constrained: Min $\|\mathbf{A} \boxplus \mathbf{x} - \mathbf{b}\|_{p=1\dots\infty}$ s.t. $\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}$

- **Theorem**: (a) The **greatest (sub)solution** of (1) and unique solution of (2) is

$$\hat{\mathbf{x}} = \varepsilon_A(\mathbf{b}) = \mathbf{A}^* \boxplus' \mathbf{b} = [\bigwedge_i b_i - a_{ij}], \quad \mathbf{A}^* \triangleq -\mathbf{A}^T$$

  and yields the **Greatest Lower Estimate (GLE)** of data $\mathbf{b}$:

$$\delta_A(\varepsilon_A(\mathbf{b})) = \mathbf{A} \boxplus (\mathbf{A}^* \boxplus' \mathbf{b}) \leq \mathbf{b}$$

  (b) **Min Max Absolute Error (MMAE) unconstrained unique solution**:

$$\tilde{\mathbf{x}} = \hat{\mathbf{x}} + \mu, \quad \mu = \|\mathbf{A} \boxplus \hat{\mathbf{x}} - \mathbf{b}\|_\infty / 2$$

- **Geometry**: Operators δ, ε are vector dilation and erosion, and
  the GLE $\mathbf{b} \mapsto \delta\varepsilon(\mathbf{b})$ is an opening (lattice projection).

- **Complexity**: $O(mn)$

ICASSP 2024 KOREA

Sparsest Solution to Max-Plus Equation

[Tsiamis & Maragos, DEDS 2019]

- A sparse vector $\boldsymbol{x} \in \mathbb{R}^n_{\max}$ has many $-\infty$ elements.
- Let $\text{supp}(\boldsymbol{x})$ be the support (the set of finite indices)
- We solve the following problems:

Exact solution

$$\min_{\boldsymbol{x} \in \mathbb{R}^n_{\max}} \quad |\text{supp}(\boldsymbol{x})| \qquad \text{subject to} \quad \boldsymbol{A} \boxplus \boldsymbol{x} = \boldsymbol{b}$$

Approximate solution

$$\min_{\boldsymbol{x} \in \mathbb{R}^n_{\max}} \quad |\text{supp}(\boldsymbol{x})| \qquad \text{subject to} \quad \|\boldsymbol{b} - \boldsymbol{A} \boxplus \boldsymbol{x}\|_1 \leq \epsilon, \quad \boldsymbol{A} \boxplus \boldsymbol{x} \leq \boldsymbol{b}$$

- NP-complete problem (~minimum set cover). Use greedy algorithms.
- Submodularity tools provide suboptimality bounds.
- Extensions to other Lp norms [Tsilivis, Tsiamis & Maragos, DGMM 2021]

Submodularity

- A set function f : U → R is called submodular if

$$f(A \cup \{k\}) - f(A) \geq f(B \cup \{k\}) - f(B) \quad \forall A \subseteq B \subseteq U,\ k \notin B.$$

- Idea of diminishing returns – natural in resource allocation problems.
- Often supplies fast, greedy, algorithms.

Sparsest Solution to Max-Plus Equation – General Norms

- Extensions to other Lp norms [Tsilivis, Tsiamis & Maragos, DGMM 2021]

$$\min_{\mathbf{x} \in \mathbb{R}^n_{\max}} |\mathrm{supp}(\mathbf{x})|, \text{ s.t. } \|\mathbf{b} - \mathbf{A} \boxplus \mathbf{x}\|_p^p \leq \epsilon, \quad \mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}. \tag{4}$$

- Greedy algorithm, as in p=1 – similar analysis.
- Provides heuristic for sparse solutions without the monotonicity constraint:

$$\mathbf{x}_{\mathrm{SMMAE}} = \mathbf{x}^* + \frac{\|\mathbf{b} - \mathbf{A} \boxplus \mathbf{x}^*\|_\infty}{2},$$

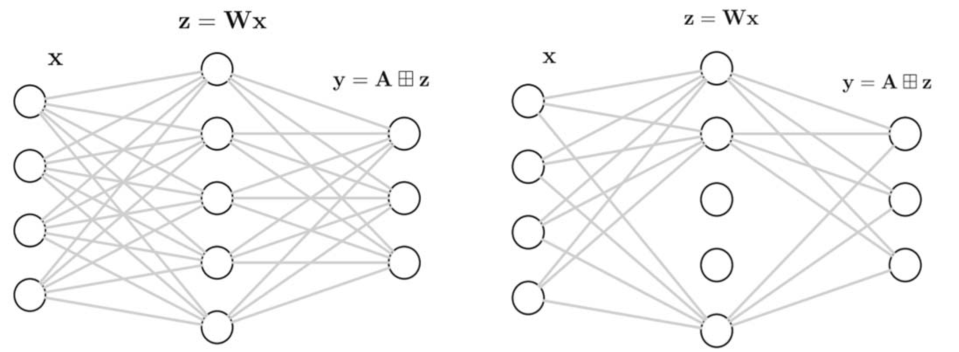
where $\mathbf{x}^*$ *is a solution of problem (4) with fixed* (p, ϵ).

- **Best** approximation error among all vectors with same *support*.
- Applications:
  - Morphological Neural Networks Minimization
  - Convex Regression

ICASSP 2024 KOREA

Morphological Neural Networks Minimization

- Sparse Solutions to Max-Plus Equations: neuron pruning in Morphological Neural Networks.



(a) A simple Max-plus block with $d = 4, n = 5, k = 3$.

(b) The same Max-plus block, after pruning two neurons from its first layer.

- Experiments on image classification datasets:

Same performance,
Less neurons

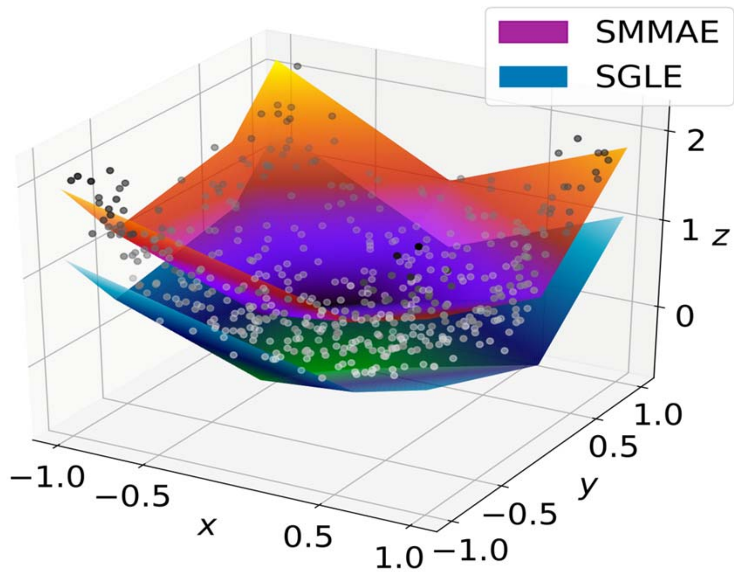
	MNIST		FashionMNIST	
	64	128	64	128
Full model	92.21	92.17	79.27	83.37
Pruned ($n = 10$)	92.21	92.17	79.27	83.37

[Tsilivis, Tsiamis & Maragos, DGMM 2021]

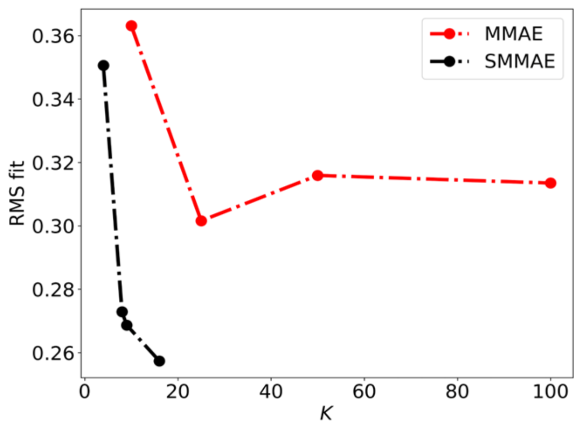
ICASSP 2024 KOREA

Multivariate Convex Regression

- Convex functions as piecewise linear $p(\mathbf{x}) = \bigvee_{k=1}^{K} \mathbf{a}_k^\intercal \mathbf{x} + b_k,$
- Approximation from data by solving max-plus systems of equations.
- Sparsity = Few affine regions.



- Improved results over non-sparse approximation:



[Tsilivis, Tsiamis & Maragos, ICASSP, 2021]

Tropical Approximation

References:

- I. Kordonis, E. Theodosis, G. Retsinas, P. Maragos, "*Matrix Factorization in Tropical and Mixed Tropical-Linear Algebras*", Proc. ICASSP, 2024.

Mixed Tropical-Linear Approximation

Mixed Tropical-Linear Map: Linear Transformation followed by a max-plus

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}_T \boxplus (\boldsymbol{A}_L \boldsymbol{x})$$

Tropical Polynomial Map: Maximum of a set of affine terms

$$p(\boldsymbol{x}) = \bigvee_{i=1}^{K} (\boldsymbol{a}_i^T \boldsymbol{x} + b_i) = [b_1 \ \dots \ b_n] \boxplus \begin{bmatrix} \boldsymbol{a}_1^T \boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_n^T \boldsymbol{x} \end{bmatrix}$$

Approximation problems involving tropical mappings

$$\boldsymbol{y}_k \approx \boldsymbol{f}(\boldsymbol{x}_k)$$

Mixed Tropical-Linear Approximation: Formulations

Mixed Tropical-Linear Map

$$f(\boldsymbol{x}) = \boldsymbol{A}_T \boxplus (\boldsymbol{A}_L \boldsymbol{x})$$

Approximate

$$\boldsymbol{y}_k \approx \boldsymbol{f}(\boldsymbol{x}_k)$$

Problems and Research Directions:

Max-plus regression: Given a set of input vectors $\boldsymbol{x}_k$ a set of target vectors $\boldsymbol{y}_k$ and matrix $\boldsymbol{A}_L$, find the tropical matrix $\boldsymbol{A}_T$ that best fits the data

Tropical inversion: Given matrices $\boldsymbol{A}_L$, $\boldsymbol{A}_T$, and the target $\boldsymbol{y}$, find the input $\boldsymbol{x}$

Tropical regression: Given a set of input-target pairs, find the mapping matrices $\boldsymbol{A}_L$, $\boldsymbol{A}_T$

Tropical compression: Given a set of targets $\boldsymbol{y}_k$, find a set of input vectors and a pair of matrices that best fits the data

Relationships among these problems; develop algorithms for their soln and applications

ICASSP 2024 KOREA

Tropical Inversion: Relation to Other Problems

Tropical inversion: Given the matrices $\boldsymbol{A_T}, \boldsymbol{A_L}$ and the output $\boldsymbol{y}$ solve

$$\underset{\boldsymbol{x}\in\mathbb{R}^d}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{A}_T \boxplus (\boldsymbol{A}_L \boldsymbol{x})\|$$

Interesting special cases:

- Bellman's equation
- Tropical Factorization
- Tropical Regression

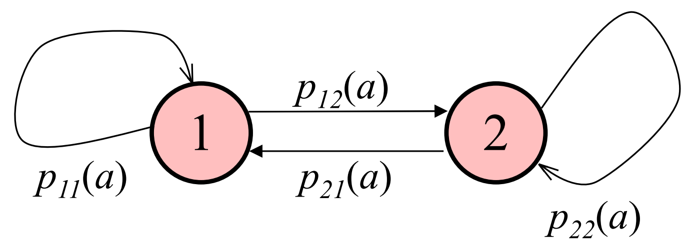
ICASSP 2024 KOREA

Bellman's equation as tropical inversion - 1

Consider a Markov Decision Process

- State s in $\{1,2\}$, Action a in $\{1,2\}$, Reward $r(s,a)$
- *Total discounted reward:*

$$\sum_{k=0}^{\infty} \gamma^k r(s_k, a_k)$$



$p_{12}(a)$: Probability of moving from state s=1 to state s'=2 under action a

Bellman's Equation:

total expected reward starting from state s :

$$V_s = \max_{a \in \{1,2\}} \left(r(s,a) + \gamma \sum_{s'=1}^{2} p_{s,s'}(a) V_{s'} \right), \;\; s = 1, 2$$

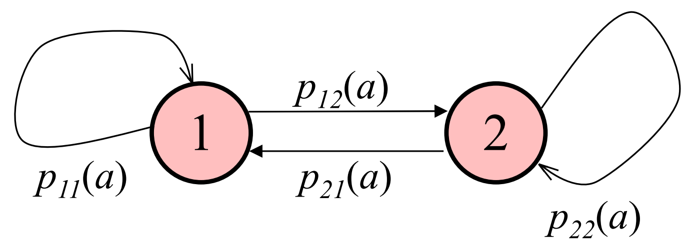
Bellman’s equation as tropical inversion - 2

Consider a Markov Decision Process

- State s in $\{1,2\}$, Action a in $\{1,2\}$, Reward $r(s,a)$

  Total discounted reward with factor γ:

$$\sum_{k=0}^{\infty} \gamma^k r(s_k, a_k)$$



$p_{12}(a)$: Probability of moving from state s=1 to state s'=2 under action a

Bellman’s Equation

$$V_s = \max_{a\in\{1,2\}} (r(s,a) + \gamma \sum_{s'=1}^{2} p_{s,s'}(a) V_{s'}), \;\; s = 1,2$$

↓ Substituting

$$V_1 = \max(r_{1,1} + \gamma(0.9V_1 + 0.1V_2), r_{1,2} + \gamma(0.2V_1 + 0.8V_2)),$$
$$V_2 = \max(r_{2,1} + \gamma(0.4V_1 + 0.6V_2), r_{2,2} + \gamma(0.1V_2 + 0.9V_3)),$$

↓ Subtracting V_s from both sides

$$0 = \max\big(r_{1,1} + [0.9\gamma - 1 \quad 0.1\gamma]\boldsymbol{V},\; r_{1,2} + [0.2\gamma - 1 \quad 0.8\gamma]\boldsymbol{V}\big)$$
$$0 = \max\big(r_{1,1} + [0.4\gamma \quad 0.6\gamma - 1]\boldsymbol{V},\; r_{1,2} + [0.1\gamma \quad 0.9\gamma - 1]\boldsymbol{V}\big)$$

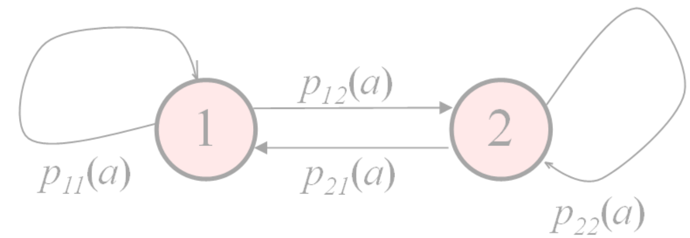
Transition probability matrices

$$\boldsymbol{P}_1 = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}, \boldsymbol{P}_2 = \begin{bmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{bmatrix}$$

Value as a vector

$$\boldsymbol{V} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

Bellman's equation as tropical inversion - 3



$$\boldsymbol{P}_1 = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}, \boldsymbol{P}_2 = \begin{bmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{bmatrix}$$

Bellman's Equation

$$V_s = \max_{a \in \{1,2\}} (r(s,a) + \gamma \sum_{s'=1}^{2} p_{s,s'}(a) V_{s'}), \ \ s = 1, 2$$

$$\downarrow$$

$$0 = \max\left(r_{1,1} + \begin{bmatrix} 0.9\gamma - 1 & 0.1\gamma \end{bmatrix} \boldsymbol{V}, \ r_{1,2} + \begin{bmatrix} 0.2\gamma - 1 & 0.8\gamma \end{bmatrix} \boldsymbol{V}\right)$$
$$0 = \max\left(r_{1,1} + \begin{bmatrix} 0.4\gamma & 0.6\gamma - 1 \end{bmatrix} \boldsymbol{V}, \ r_{1,2} + \begin{bmatrix} 0.1\gamma & 0.9\gamma - 1 \end{bmatrix} \boldsymbol{V}\right)$$

Write in terms of a Tropical Inversion Problem

$$\mathbf{0} = \boldsymbol{A}_T \boxplus (\boldsymbol{A}_L \boldsymbol{V})$$

where

$$\boldsymbol{A}_L = \begin{bmatrix} 0.9\gamma - 1 & 0.1\gamma \\ 0.2\gamma - 1 & 0.8\gamma \\ 0.4\gamma & 0.6\gamma - 1 \\ 0.1\gamma & 0.9\gamma - 1 \end{bmatrix}, \boldsymbol{A}_T = \begin{bmatrix} r_{1,1} & r_{1,2} & -\infty & -\infty \\ -\infty & -\infty & r_{2,1} & r_{2,2} \end{bmatrix}$$

The Tropical Inversion Problem: The Double Projection Algorithm

Tropical inversion problem

Find $\boldsymbol{x}$ such that

$$f(\boldsymbol{x}) = \boldsymbol{A}_T \boxplus (\boldsymbol{A}_L \boldsymbol{x})$$

is as close as possible to a target value $\boldsymbol{y}$

Max-plus inversion problem

Find $\boldsymbol{v}$ such that

$$\boldsymbol{A}_T \boxplus \boldsymbol{v}$$

as close as possible to a target $\boldsymbol{y}$

Project to the tropical cone

$$S' = \{\boldsymbol{A}_T \boxplus \boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^m\}$$

$$\boldsymbol{v} = \boldsymbol{A}_T^* \boxplus' \boldsymbol{y} + \|(\boldsymbol{A}_T \boxplus (\boldsymbol{A}_T^* \boxplus' \boldsymbol{y}) - \boldsymbol{y}\|_\infty / 2$$

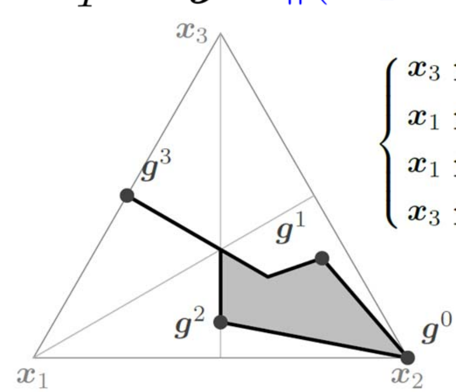
Project to the linear subspace

$$S'' = \{\boldsymbol{A}_L \boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^d\}$$

$$P_S(\boldsymbol{y}) = \boldsymbol{A}_L^+ \boldsymbol{v}$$

Geometric Problem: Describe set S

$$S = \{\boldsymbol{A}_T \boxplus (\boldsymbol{A}_L \boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$$



$$\begin{cases} \boldsymbol{x}_3 \le \boldsymbol{x}_1 + 2 \\ \boldsymbol{x}_1 \le \max(\boldsymbol{x}_2, \boldsymbol{x}_3) \\ \boldsymbol{x}_1 \le \boldsymbol{x}_3 + 2 \\ \boldsymbol{x}_3 \le \max(\boldsymbol{x}_1, \boldsymbol{x}_2 - 1) \end{cases}$$

A set like S'.
[Allamigeon et al. 2010]

Tropical Matrix Factorization (TMF): Algorithm

$$\underset{\boldsymbol{C},\boldsymbol{D}}{\text{minimize}} \quad \|\boldsymbol{Y} - \boldsymbol{C} \boxplus \boldsymbol{D}\|_F$$

Early work: De Schutter and De Moor 1990's
More recent: Karaev et. al 2010's, Omanović et. al 2020's

Algorithms:

- Exact (using Extended Linear Complementarity Problem)
- Gradient-based
- Max-plus based

Our:

- Improved Gradient-based
- Compares favourably with the literature

Algorithms for Tropical Compression

Combine algorithms for TMF with ordinary low rank approximation

[I.K., E.T., G.R, P.M. "*Matrix Factorization in Tropical and Mixed Tropical-Linear Algebras*", ICASSP2024]

The Tropical Compression Problem

Problem: Given a set of vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ determine appropriate matrices $\boldsymbol{A}_L, \boldsymbol{A}_T$ and datapoints $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^d$ that solve

$$\underset{\boldsymbol{x}_i, \boldsymbol{A}_L, \boldsymbol{A}_T}{\text{minimize}} \quad \sum_{i=1}^{N} \|\boldsymbol{y}_i - \boldsymbol{A}_T \boxplus (\boldsymbol{A}_L \boldsymbol{x}_i)\|_2^2$$

Multi-person utility learning: Model the preferences of users, based on past choices

- *Users*: They have a preference according to the features of the items (unknown function)
- *Items*: They have unknown features $\boldsymbol{x}_i$

Reduction to a Constrained Approximate TMF: For $\boldsymbol{X} = [\boldsymbol{x}_1 \ \ldots \ \boldsymbol{x}_N]$, $\boldsymbol{Y} = [\boldsymbol{y}_1 \ \ldots \ \boldsymbol{y}_N]$ the problem is written as:

$$\begin{aligned} &\underset{\boldsymbol{C}, \boldsymbol{D}}{\text{minimize}} && \|\boldsymbol{Y} - \boldsymbol{C} \boxplus \boldsymbol{D}\|_F \\ &\text{subj. to} && \text{rank}(\boldsymbol{D}) \leq d \end{aligned}$$

ICASSP 2024 KOREA

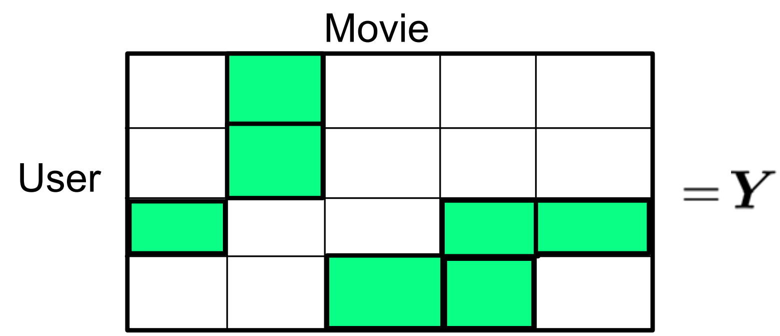
Application of Tropical Compression in Recommender Systems

Movielens 1M Dataset

- Set of Users
- Set of N Movies

$$\underset{\boldsymbol{x}_j, \boldsymbol{A}_L, \boldsymbol{A}_T}{\text{minimize}} \quad \sum_{j=1}^{N} \|\boldsymbol{y}_j - \boldsymbol{A}_T \boxplus (\boldsymbol{A}_L \boldsymbol{x}_j)\|_2^2$$

Implicit Feedback



Use Tropical Compression on Y

- Get the matrices and the estimated features
- Compute the probability of person i watch movie j
- Results better that well-regularized linear matrix factorization

  (Hit Rate @10 → Success %: 74.3 vs 73.1)

ICASSP 2024 KOREA

[I.K., E.T., G.R, P.M. "*Matrix Factorization in tropical and Mixed Tropical-Linear Algebras*", ICASSP2024]

Conclusions

- Interesting and diverse approximation problems use Tropical Maps.
  Several motivations.
- Interesting Geometry/Combinatorics
- Exact solutions are difficult. Focus on approximate algorithms
- Central role of Tropical Matrix Factorization in analysis and algorithms
- Applications, e.g. Recommender Systems

Tropical Regression and Piecewise-Linear Surface Fitting

Main References:

- P. Maragos and E. Theodosis, "*Multivariate Tropical Regression and Piecewise-Linear Surface Fitting*", *Proc. ICASSP*, 2020.
- P. Maragos, V. Charisopoulos and E. Theodosis, "*Tropical Geometry and Machine Learning*", *Proceedings of the IEEE*, 2021.

Related:

- A. Magnani and S. Boyd, "*Convex piecewise-linear fitting*," Optim. Eng., 2009.
- J. Hook, "*Linear regression over the max-plus semiring: Algorithms and applications*," ArXiv 2017.
- A. Ghosh et al., "*Max-Affine Regression: Parameter Estimation for Gaussian Designs*", IEEE T-Info. Theory, 2022.

Optimal Regression for Fitting Euclidean vs Tropical Lines

Problem: Fit a curve to data $(x_i, y_i), \quad i = 1, \ldots, m$

Euclidean:

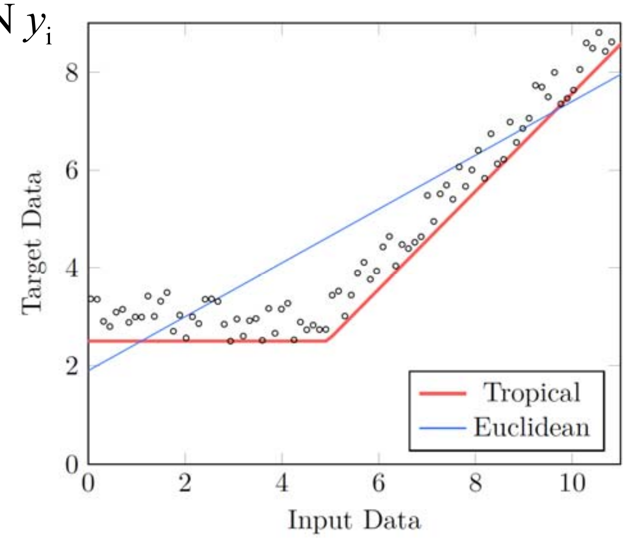
Fit a straight line $y = ax + b$ by minimizing ℓ_2-norm of error:

$$a = \frac{\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)/m}{\sum (x_i)^2 - \left(\sum x_i\right)^2/m} \;,\; b = \frac{1}{m}\sum_i y_i - a x_i$$

Tropical:

Fit a tropical line $y = \max(a + x, b)$ by minimizing some ℓ_p-norm of error:

Greatest Subsolution: $a = \mathrm{MIN}_i \; y_i - x_i \;,\; b = \mathrm{MIN}_i \; y_i$



Solve Max-plus Equations

- **Problems**:

  (1) Exact problem: Solve $\delta_A(\mathbf{x}) = \mathbf{A} \boxplus \mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \overline{\mathbb{R}}^{m\times n}, \ \mathbf{b} \in \overline{\mathbb{R}}^{m}$

  (2) Approximate Constrained: Min $\|\mathbf{A} \boxplus \mathbf{x} - \mathbf{b}\|_{p=1\ldots\infty}$ s.t. $\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}$

- **Theorem**: (a) The **greatest (sub)solution** of (1) and unique solution of (2) is

$$\hat{\mathbf{x}} = \varepsilon_A(\mathbf{b}) = \mathbf{A}^* \boxplus' \mathbf{b} = [\bigwedge_i b_i - a_{ij}], \quad \mathbf{A}^* \triangleq -\mathbf{A}^T$$

  and yields the **Greatest Lower Estimate (GLE)** of data $\mathbf{b}$:

  Lattice Projection: $\delta_A(\varepsilon_A(\mathbf{b})) = \mathbf{A} \boxplus (\mathbf{A}^* \boxplus' \mathbf{b}) \leq \mathbf{b}$

  (b) **Min Max Absolute Error (MMAE) unconstrained unique solution**:

$$\tilde{\mathbf{x}} = \hat{\mathbf{x}} + \mu, \quad \mu = \|\mathbf{A} \boxplus \hat{\mathbf{x}} - \mathbf{b}\|_\infty / 2$$

- **Geometry**: Operators δ, ε are vector dilation and erosion, and the GLE $\mathbf{b} \mapsto \delta\varepsilon(\mathbf{b})$ is an opening (lattice projection).

- **Complexity**: $O(mn)$

Sparse solutions: [Tsiamis & Maragos 2019], [Tsilivis et al. 2021]

Optimally Fitting Tropical Lines to Data

Problem: Fit a tropical line $y = \max(a + x, b)$ to noisy data (x_i, f_i), $i = 1, \ldots, m$, where $f_i = y_i + \text{error}$ by minimizing $\ell_{1,\ldots,\infty}$ norm of error:

Greatest Subsolution (GLE): $\hat{w} = (\hat{a}, \hat{b})$, $\hat{a} = \text{MIN}_i\, f_i - x_i$, $\hat{b} = \text{MIN}_i\, f_\text{i}$

Min Max Abs. Error (MMAE) Solution: $\tilde{w} = \hat{w} + \mu$, $\mu = \|\,\text{GLE error}\,\|_\infty /2$

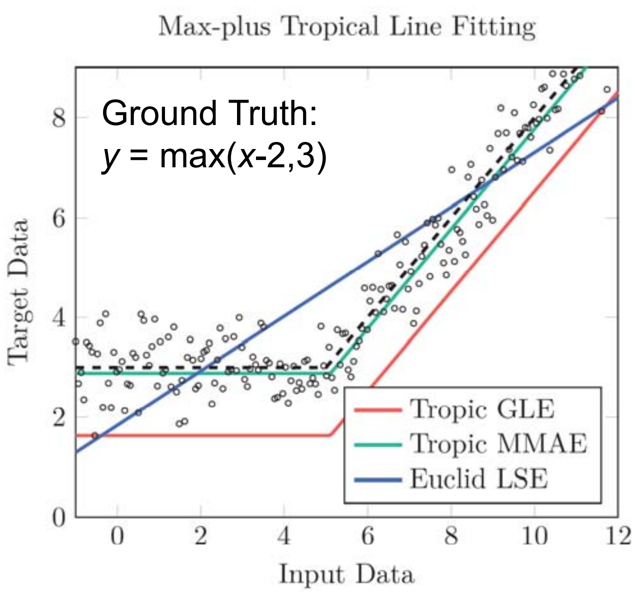
$$\underbrace{\begin{bmatrix} x_1 & 0 \\ \vdots & \vdots \\ x_m & 0 \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \Longrightarrow \underbrace{\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}}_{\hat{\mathbf{w}}} = \underbrace{\begin{bmatrix} \bigwedge_i f_i - x_i \\ \bigwedge_i f_i \end{bmatrix}}_{\mathbf{X}^* \boxplus' \mathbf{f}}$$

Numerical Examples of Optimally Fitting Tropical Lines to Data

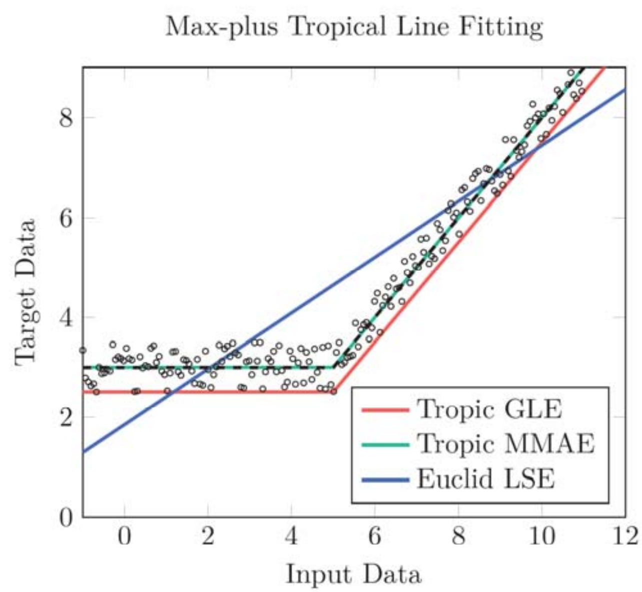
Problem: Fit a tropical line $y = \max(a + x, b)$ to noisy data (x_i, f_i), $i = 1, \ldots, m = 200$, where $f_i = y_i + \text{error}$ by minimizing $\ell_{1,\ldots,\infty}$ of error:

Greatest Subsolution (GLE): $\hat{w} = (\hat{a}, \hat{b})$, $\hat{a} = \text{MIN}_i\, f_i - x_i$, $\hat{b} = \text{MIN}_i\, f_i$

Min Max Abs. Error (MMAE) Solution: $\tilde{w} = \hat{w} + \mu$, $\mu = \|\text{ GLE error }\|_\infty /2$



(a) T-line with Gaussian Noise



(b) T-line with Uniform Noise

ICASSP 2024 KOREA

Optimal Fitting 1D Max-Plus Tropical Polynomials to Data

We wish to fit a tropical polynomial $f(x)$ to given data $(x_i, f_i) \in \mathbb{R}^2$, $i = 1, ..., m$,

$$f(x) = \max(a_0x + b_0, a_1x + b_1, a_2x + b_2, \ldots, a_Kx + b_K) = \bigvee_{k=0}^{K} a_kx + b_k$$

where $a_k \in \mathbb{Z}$, $b_k \in \mathbb{R}$, and $f_i = f(x_i) + \text{error}$, by minimizing the ℓ_1 error norm. For example, if $a_k = k-1$ we have a K-degree tropical polynomial curve:

$$f(x) = \max(b_0, x + b_1, 2x + b_2, \ldots, Kx + b_K)$$

The equations to solve for finding the optimal parameters $\mathbf{b}$ become:

$$\underbrace{\begin{bmatrix} a_0x_1 & a_1x_1 & a_2x_1 & \cdots & a_Kx_1 \\ a_0x_2 & a_1x_2 & a_2x_2 & \cdots & a_Kx_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_0x_m & a_1x_m & a_2x_m & \cdots & a_Kx_m \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}}_{\mathbf{b}} = \underbrace{\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}}$$

Optimal solution for minimum ℓ_1 error

$$\begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_K \end{bmatrix} = \hat{\mathbf{b}} = \mathbf{X}^* \boxplus' \mathbf{f} = \begin{bmatrix} -a_0x_1 & -a_0x_2 & \cdots & -a_0x_m \\ -a_1x_1 & -a_1x_2 & \cdots & -a_1x_m \\ \vdots & \vdots & \vdots & \vdots \\ -a_Kx_1 & -a_Kx_2 & \cdots & -a_Kx_m \end{bmatrix} \boxplus' \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} = \begin{bmatrix} \bigwedge_{i=1}^{m} f_i - a_0x_i \\ \bigwedge_{i=1}^{m} f_i - a_1x_i \\ \vdots \\ \bigwedge_{i=1}^{m} f_i - a_Kx_i \end{bmatrix}$$

ICASSP 2024 KOREA

Optimal Fitting Max-Plus Tropical Planes to Data

Problem: Fit a tropical plane $z = \max(a + x, b + y, c)$ to noisy data (x_i, y_i, f_i), where $f_i = z_i$+error, $i = 1, \ldots, m = 100,$ by minimizing $\ell_{1,\ldots,\infty}$ norm of error:

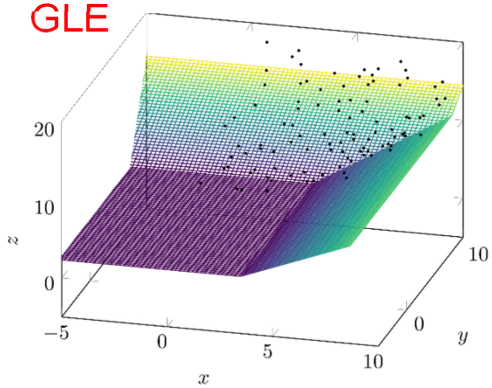
Greatest Subsolution (GLE): $\hat{w} = (\hat{a}, \hat{b}, \hat{c})$

Min Max Abs. Error (MMAE) Solution: $\tilde{w} = \hat{w} + \mu, \quad \mu = \| \text{GLE error} \|_\infty /2$

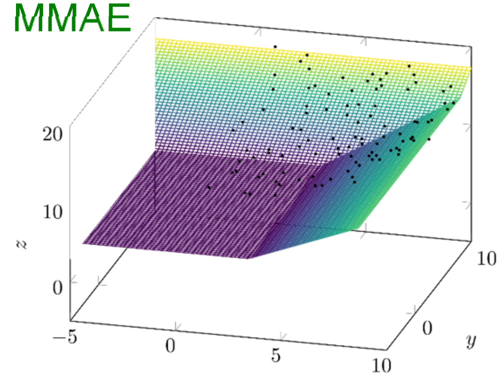
$$\underbrace{\begin{bmatrix} x_1 & y_1 & 0 \\ \vdots & \vdots & \vdots \\ x_m & y_m & 0 \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \Longrightarrow \underbrace{\begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix}}_{\hat{\mathbf{w}}} = \underbrace{\begin{bmatrix} \bigwedge_i f_i - x_i \\ \bigwedge_i f_i - y_i \\ \bigwedge_i f_i \end{bmatrix}}_{\mathbf{X}^* \boxplus' \mathbf{f}}$$

Ground Truth:
$z = \max(x + 5, y + 7, 9)$
Noise: $N(0,1)$

GLE



MMAE



Optimal Fitting 2D Higher-degree Tropical Polynomials to Data

Data (noisy paraboloid):

3D tuples $(x_i, y_i, f_i) \in \mathbb{R}^3$

$f_i = x_i^2 + y_i^2 + \varepsilon_i,$

$(x_i, y_i) \sim \text{Unif}[-1,1]$

$\varepsilon_i \sim \mathcal{N}(0, 0.25^2)$

Model:

Fit K-rank 2D trop. polynomial

$p(x,y) = \mathrm{MAX}_{k=1}^{K} \{a_k x + b_k y + c_k\}$

by minimizing error $\| f_i - p(x_i, y_i) \|_\infty$

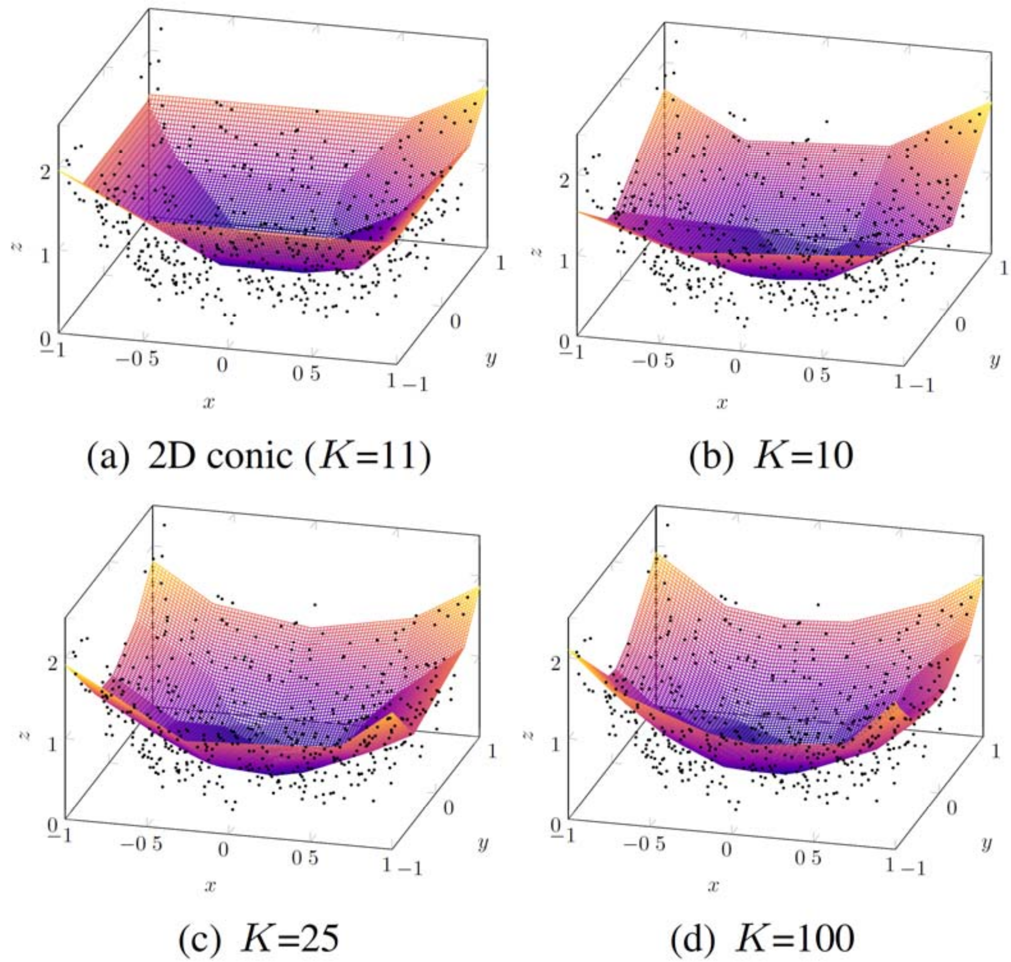
Estimation algorithm:

$K-$means on data gradients $\rightarrow a_k, b_k$

solve max-plus eqns $\rightarrow c_k$

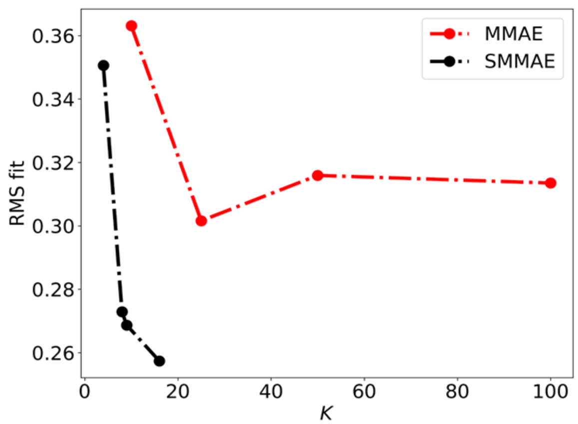
Complexity: $\approx$ Linear

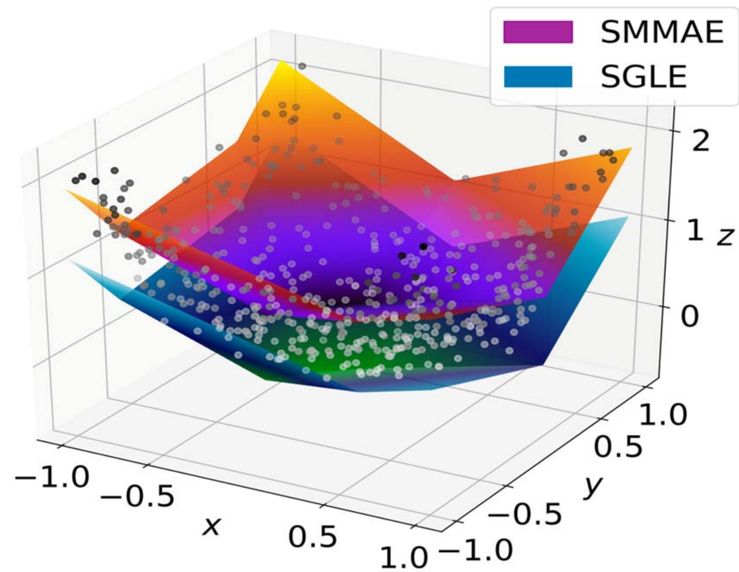
O(#data, #dimensions)



Multivariate Convex Regression

- Convex functions as piecewise linear $p(\mathbf{x}) = \bigvee_{k=1}^{K} \mathbf{a}_k^{\intercal}\mathbf{x} + b_k,$
- Approximation from data by solving max-plus systems of equations.
- Sparsity = Few affine regions.
- Improved results over non-sparse approximation:



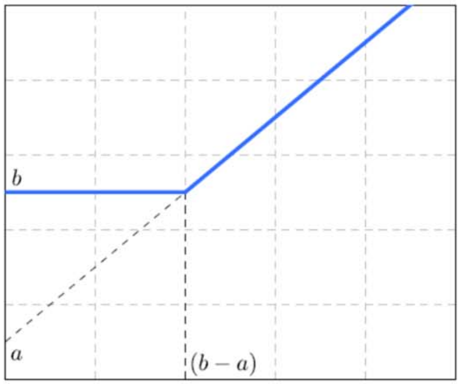


[Tsilivis, Tsiamis & Maragos, ICASSP, 2021]

Generalized Tropical Versions of Lines & Planes over Max-* Algebras

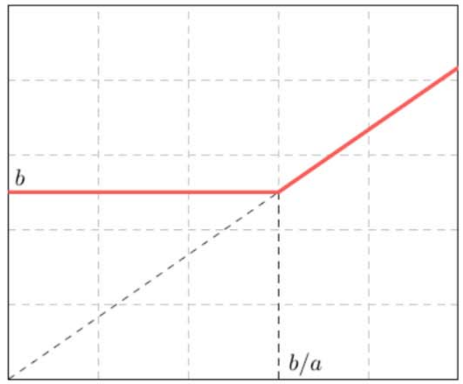
Max-plus Tropical Line

$y = \max(a + x, b)$



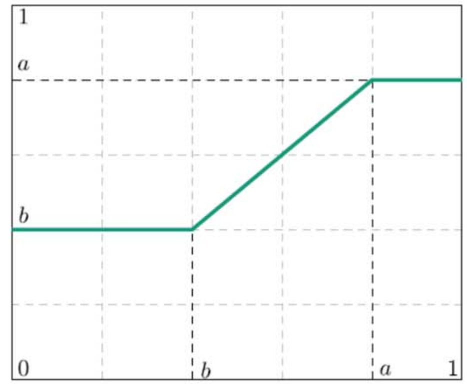
Max-product Tropical Line

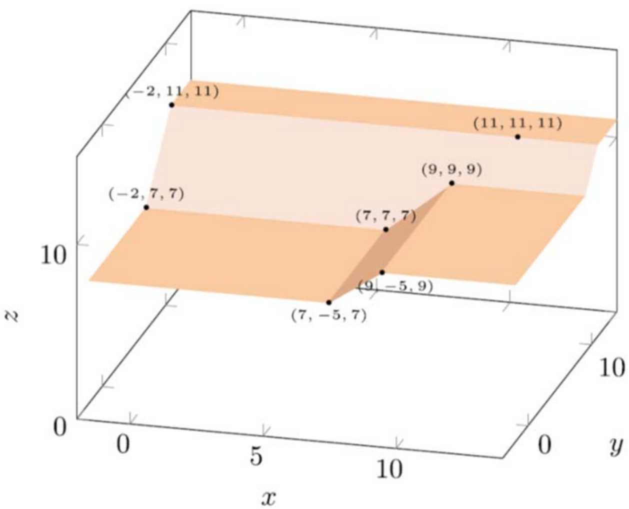
$y = \max(a \cdot x, b)$



Max-min Tropical Line

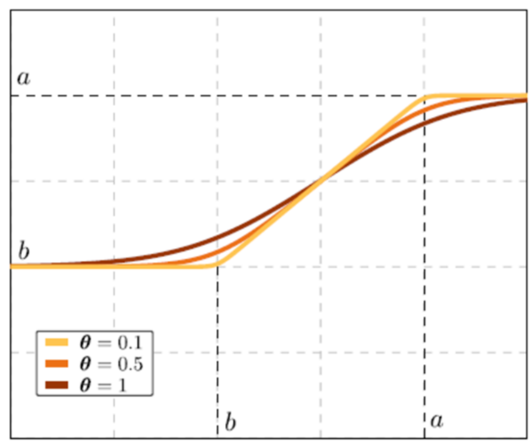
$y = \max(\min(a, x), b)$





Max–min plane $z = \max(9 \wedge x, 11 \wedge y, 7)$.

SoftMax-SoftMin Tropical Line



Conclusions

- **Tropical Geometry**, and its underlying **max-plus algebra**, provide principled and insightful tools for analysis of NNs with PWL activations and other ML systems.
- **NNs** with nonlinear max/min-plus nodes: similar performance and superior compression ability compared to linear counterparts. Trained via CCP or SGD/Adam.
- **Tropical Regression**: Tropical Polynomials for multidimensional data fitting using PWL functions. Low-complexity algorithm from optimal solutions of max-plus eqns.
- **NN Minimization**: TG offers effective and insightful tools for compression of NNs.
- **Mixed Tropical-Linear Approximation** can solve several types of optimization problems.
- **Future work**: extensions to deeper networks, nonconvex settings, and to more general functions using max-* algebra on weighted lattices. Tropical Approximation theory & applications.

Thank you for your attention

Collaborators and References

TG&ML: Vasilis Charisopoulos, Manos Theodosis

NN Minimization:
Georgios Smyrnis, Panos Misiakos, George Retsinas, Nikos Dimitriadis

Tropical Approximation: Ioannis Kordonis

Tropical Sparsity: Anastasios Tsiamis, Nikos Tsilivis

For more information, demos, and current results:

http://robotics.ntua.gr and http://cvsp.cs.ntua.gr

Bibliography - 1

- M. Akian, S. Gaubert, V. Nitica, and I. Singer, “Best approximation in max-plus semimodules,” Linear Algebra Appl., vol. 435, no. 12, pp. 3261–3296, Dec. 2011.
- M. Akian, S. Gaubert, and A. Guterman, “Tropical polyhedra are equivalent to mean payoff games,” Int. J. Algebra Comput., vol. 22, no. 1, Feb. 2012, Art. no. 1250001.
- M. Akian, S. Gaubert, Y. Qi and O. Saadi, “Tropical Linear Regression And Mean Payoff Games: Or, How To Measure The Distance To Equilibria”, ArXiv 2021.
- M. Alfarra, A. Bibi, H. Hammoud, M. Gaafar, and B. Ghanem, “On the decision boundaries of neural networks: A tropical geometry perspective,” 2020, arXiv:2002.08838. [Online].
- X. Allamigeon, S. Gaubert and E. Goubault, “The Tropical Double Description Method”, SSTACS 2010.
- R. Arora, A. Basu, P. Mianjy, and A. Mukherjee, “Understanding deep neural networks with rectified linear units,” in Proc. Int. Conf. Learn. Represent., 2018.
- D. Avis and K. Fukuda, “Reverse search for enumeration,” Discrete Appl. Math., vol. 65, nos. 1–3, pp. 21–46, Mar. 1996.
- N. S. Aybat and Z. Wang, “A parallel method for large scale convex regression problems,” in 53rd IEEE Conference on Decision and Control. IEEE, 2014, pp. 5710–5717.
- F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, Synchronization and Linearity: An Algebra for Discrete Event Systems. Hoboken, NJ, USA: Wiley, 2001.
- Gabor Balazs, "Convex Regression: Theory, Practice, and Applications," University of Alberta, 2016.
- D. Bertsimas, and N. Mundru, "Sparse convex regression“, INFORMS Journal on Computing 33.1 (2021): 262-279.
- G. Birkhoff, Lattice Theory, AMS, 1967.

Bibliography - 2

- T. S. Blyth, Lattices and Ordered Algebraic Structures. New York, NY, USA: Springer-Verlag, 2005.
- T. S. Blyth and M. F. Janowitz, Residuation Theory. New York, NY, USA: Pergamon, 1972.
- A. Bora, A. Jalal, E. Price, and A. G. Dimakis, “Compressed sensing using generative models,” in Proc. Int. Conf. Mach. Learn. (ICML), 2017, pp. 537–546.
- G. Borgefors, “Distance transformations in arbitrary dimensions,” Comput. Vis., Graph., Image Process., vol. 27, no. 3, pp. 321–345, Sep. 1984.
- S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, “A tutorial on geometric programming,” Optim. Eng., vol. 8, no. 1, pp. 67–127, May 2007.
- S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- R. W. Brockett and P. Maragos, “Evolution equations for continuous-scale morphological filtering,” IEEE Trans. Signal Process., vol. 42, no. 12, pp. 3377–3386, Dec. 1994.
- P. Butkoviˇc, H. Schneider and S. Sergeev, “Generators, extremals and bases of max cones”, Linear Algebra and its Applications 421 (2007) 394–406.
- P. Butkoviˇc, Max-Linear Systems: Theory and Algorithms. London, U.K.: Springer-Verlag, 2010.
- G. C. Calafiore, S. Gaubert, and C. Possieri, “Log-sum-exp neural networks and posynomial models for convex and log-log-convex data,” IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 5, pp. 1–12, Mar. 2020.
- G. C. Calafiore, S. Gaubert, and C. Possieri, “A universal approximation result for difference of log-sum-exp neural networks,” IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 12, pp. 5603–5612, Dec. 2020.
- V. Charisopoulos and P. Maragos, “Morphological perceptrons: Geometry and training algorithms,” in Proc. Int. Symp. Math. Morphol., in Lecture Notes in Computer Science, vol. 10225, J. Angulo, Ed. Cham, Switzerland: Springer, 2017, pp. 3–15.

Bibliography - 3

- V. Charisopoulos and P. Maragos, “A tropical approach to neural networks with piecewise linear activations,” 2018, arXiv:1805.08749. [Online]. Available: http://arxiv.org/abs/1805.08749
- W. Chen and R. Mazumder, “Multivariate convex regression at scale,” 2020, arXiv:2005.11588. [Online]. Available: http://arxiv.org/ abs/2005.11588
- L. O. Chua and A.-C. Deng, “Canonical piecewise-linear representation,” IEEE Trans. Circuits Syst., vol. 35, no. 1, pp. 101–111, Jan. 1988.
- G. Cohen, D. Dubois, J. Quadrat, and M. Viot, “A linear-system-theoretic view of discrete-event processes and its use for performance evaluation in manufacturing,” IEEE Trans. Autom. Control, vol. 30, no. 3, pp. 210–220, Mar. 1985.
- G. Cohen, S. Gaubert, and J.-P. Quadrat, “Duality and separation theorems in idempotent semimodules,” Linear Algebra Appl., vol. 379, pp. 395–422, Mar. 2004.
- R. A. Crowell, “The tropical division problem and the Minkowski factorization of generalized permutahedra,” arXiv preprintarXiv:1908.00241, 2019.
- M. A. Cueto, J. Morton, and B. Sturmfels, “Geometry of the restricted Boltzmann machine,” in Algebraic Methods in Statistics and Probability II (Contemporary Mathematics), vol. 516, M. A. G. Viana and H. P. Wynn, Eds. Providence, RI, USA: AMS, 2010, pp. 135–153.
- R. Cuninghame-Green, Minimax Algebra. New York, NY, USA: Springer-Verlag, 1979.
- R. A. Cuninghame-Green, “Projections in minimax algebra,” Math. Program., vol. 10, no. 1, pp. 111–123, Dec. 1976.
- R. A. Cuninghame-Green and K. Cechlárová, “Residuation in fuzzy algebra and some applications,” Fuzzy Sets Syst., vol. 71, no. 2, pp. 227–239, Apr. 1995.

Bibliography - 4

- A. Damle and Y. Sun, "A geometric approach to archetypal analysis and nonnegative matrix factorization," Technometrics, vol. 59, no. 3, pp. 361–370, Jul. 2017.
- M. Develin and B. Sturmfels, "Tropical convexity", Doc. Math., 9:1–27 (electronic), 2004.
- S. S. Dey, G. Wang, and Y. Xie, "An approximation algorithm for training one-node ReLU neural network," 2018, arXiv:1810.03592. [Online]. Available: http://arxiv.org/abs/1810.03592
- L. Dorst and R. Van den Boomgaard, "Morphological signal processing and the slope transform," Signal Process., vol. 38, no. 1, pp. 79–98, Jul. 1994.
- A. Dunbar and L. Ruthotto, "An alternating minimization algorithm for regression with tropical rational functions."
- Y. Eldar and G. Kutyniok, Compressed Sensing: Theory and Applications. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- P. Felzenszwalb and D. Huttenlocher, "Efficient Belief Propagation for Early Vision", Int'l J. Comp. Vision, 70(1), 41–54, 2006.
- W. Fenchel, "On conjugate convex functions," Can. J. Math., vol. 1, no. 1, pp. 73–77, Feb. 1949.
- M. Fickus, D. G. Mixon, A. A. Nelson, and Y. Wang, "Phase retrieval from very few measurements," Linear Algebra Appl., vol. 449, pp. 475–499, May 2014.
- G. Franchi, A. Fehri, and A. Yao, "Deep morphological networks," Pattern Recognit., vol. 102, Jun. 2020, Art. no. 107246.
- K. Fukuda, "From the zonotope construction to the Minkowski addition of convex polytopes," J. Symbolic Comput., vol. 38, no. 4, pp. 1261–1272, Oct. 2004.
- K. Fukuda, "Lecture: Polyhedral computation, spring 2016," Institute for Operations Research and Institute of Theoretical Computer Science, ETH Zurich. https://inf. ethz.ch/personal/fukudak/lect/pclect/notes2015/PolyComp2015. pdf, 2016.
- K. Fukuda et al., "Frequently asked questions in polyhedral computation," ETH, Zurich, Switzerland, 2004.

Bibliography - 5

- S. Gaubert and R. D. Katz, “The Minkowski theorem for max-plus convex sets”, Linear Algebra and its Applications 421 (2007) 356–369.
- S. Gaubert and R. D. Katz, “Minimal half-spaces and external representation of tropical polyhedra,” J. Algebr. Combinatorics, vol. 33, no. 3, pp. 325–348, May 2011.
- S. Gaubert and M. Plus, “Methods and applications of (max,+) linear algebra,” in Proc. Symp. Theor. Aspects Comput. Sci., 1997, pp. 261–282.
- A. Ghosh, A. Pananjady, A. Guntuboyina, and K. Ramchandran, “Max-affine regression: Provable, tractable, and near-optimal statistical estimation,” 2019, arXiv:1906.09255. [Online]. Available: http://arxiv.org/abs/1906.09255
- A. Ghosh, A. Pananjady, A. Guntuboyina, and K. Ramchandran, “Max-affine regression: Parameter Estimation for Gaussian Designs”, IEEE Trans. Info. Theory, March 2022.
- M. Gondran and M. Minoux, Graphs, Dioids and Semirings: New Models and Algorithms. New York, NY, USA: Springer, 2008.
- I. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, “Maxout networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2013, pp. 1319–1327.
- P. Gritzmann and B. Sturmfels, “Minkowski addition of polytopes: Computational complexity and applications to Gröbner bases,” SIAM J. Discrete Math., vol. 6, no. 2, pp. 246–269, May 1993.
- C. Güzelis and I. C. Göknar, “A canonical representation for piecewise-affine maps and its applications to circuit analysis,” IEEE Trans. Circuits Syst., vol. 38, no. 11, pp. 1342–1354, Nov. 1991.
- S. Han, J. Pool, J. Tran, and W. Dally, “Learning both weights and connections for efficient neural network,” in Proc. Adv. Neural Inf. Process. Syst., 2015.
- P. Hand and V. Voroninski, “Global guarantees for enforcing deep generative priors by empirical risk,” in Proc. Conf. Learn. Theory, 2018, pp. 970–978.

Bibliography - 6

- L. A. Hannah and D. B. Dunson, “Multivariate convex regression with adaptive partitioning,” 2011, arXiv:1105.1924. [Online]. Available: http://arxiv.org/abs/1105.1924
- L. A. Hannah and D. B. Dunson, “Ensemble methods for convex regression with applications to geometric programming based circuit design,” in Proc. Int. Conf. Mach. Learn. (ICML), 2012.
- F. M. Harper and J. A. Konstan, “The movielens datasets: History and context,” Acm transactions on interactive intelligent systems (tiis), vol. 5, no. 4, pp. 1–19, 2015.
- Y. He, J. Lin, Z. Liu, H. Wang, L.-J. Li, and S. Han, “AMC: AutoML for model compression and acceleration on mobile devices,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 784–800.
- B. Heidergott, G. J. Olsder, and J. van der Woude, Max Plus at Work: Modeling and Analysis of Synchronized Systems: A Course on Max-Plus Algebra and Its Applications. Princeton, NJ, USA: Princeton Univ. Press, 2006.
- H. Heijmans, Morphological Image Operators. New York, NY, USA: Academic, 1994.
- H. J. A. M. Heijmans and P. Maragos, “Lattice calculus of the morphological slope transform,” Signal Process., vol. 59, no. 1, pp. 17–42, May 1997.
- C. Hildreth, “Point estimates of ordinates of concave functions,” J. Amer. Stat. Assoc., vol. 49, no. 267, pp. 598–619, Sep. 1954.
- G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

Bibliography - 7

- W. Hoburg, P. Kirschen, and P. Abbeel, “Data fitting with geometric-programming-compatible softmax functions,” Optim. Eng., vol. 17, no. 4, pp. 897–918, Dec. 2016.
- C. A. Holloway, “On the estimation of convex functions,” Oper. Res., vol. 27, no. 2, pp. 401–407, 1979.
- J. Hook, “Linear regression over the max-plus semiring: Algorithms and applications,” 2017, arXiv:1712.03499. [Online]. Available: http://arxiv.org/abs/1712.03499
- J. Hook, “Min-plus algebraic low rank matrix approximation: a new method for revealing structure in networks,” arXiv preprint arXiv:1708.06552, 2017
- T. Hori and A. Nakamura, Speech Recognition Algorithms Using Weighted Finite-State Transducers. San Rafael, CA, USA: Morgan & Claypool, 2013.
- A. I. Humayun, R. Balestriero, G. Balakrishnan and R. G. Baraniuk, “SplineCam: Exact Visualization and Characterization of Deep Network Geometry and Decision Boundaries”, Proc. CVPR 2023.
- I. Itenberg, G. Mikhalkin, and E. I. Shustin, Tropical algebraic geometry. Springer Science & Business Media, 2009, vol. 35.
- M. Joswig. Tropical Halfspaces. In Combinatorial and computational geometry, vol. 52 of Math. Sci. Res. Inst. Publ., pp. 409–431, Cambridge Univ. Press, 2005.
- M. Joswig, “The Cayley trick for tropical hypersurfaces with a view toward Ricardian economics,” in Homological and Computational Methods in Commutative Algebra. Springer, 2017, pp. 107–128
- M. Joswig, Essentials of tropical combinatorics. American Mathematical Society, 2021, vol. 219.
- P. Julian, “The complete canonical piecewise-linear representation: Functional form minimal degenerate intersections,” IEEE Trans. Circuits Syst. I, Fundam. Theory Appl., vol. 50, no. 3, pp. 387–396, Mar. 2003.

Bibliography - 8

- C. Kahlert and L. O. Chua, “The complete canonical piecewise-linear representation. I. The geometry of the domain space,” IEEE Trans. Circuits Syst. I, Fundam. Theory Appl., vol. 39, no. 3, pp. 222–236, Mar. 1992.
- S. Karaev and P. Miettinen, “Capricorn: An algorithm for subtropical matrix factorization,” in Proceedings of the 2016 SIAM international conference on data mining. SIAM, 2016, pp. 702–710.
- S. Karaev and P. Miettinen, “Cancer: Another algorithm for subtropical matrix factorization,” in Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part II 16. Springer, 2016, pp. 576–592.
- S. Karaev and P. Miettinen, “Algorithms for approximate subtropical matrix factorization,” Data Mining and Knowledge Discovery, vol. 33, pp. 526–576, 2019.
- S. Karaev, “Matrix factorization over dioids and its applications in data mining,” 2019.
- J. Kim, L. Vandenberghe, and C.-K.-K. Yang, “Convex piecewise-linear modeling method for circuit optimization via geometric programming,” IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 29, no. 11, pp. 1823–1827, Nov. 2010.
- I. Kordonis, P. Maragos, and G.P. Papavassilopoulos. "Stochastic stability in max-product and max-plus systems with Markovian jumps." Automatica 92, pp 123-132, 2018
- I. Kordonis and P. Maragos, “Revisiting tropical polynomial division: Theory, algorithms and application to neural networks,” preprint arXiv:2306.15157, 2023
- I. Kordonis, E. Theodosis, G. Retsinas, P. Maragos, “Matrix Factorization in Tropical and Mixed Tropical-Linear Algebras,” arXiv:2309.13914, 2023

Bibliography - 9

- I. Kordonis, P. Maragos “The Tropical Inversion Problem: Unifying Approximation Problems from Diverse Fields”, ArXiv.
- F. Lauer, “On the complexity of piecewise affine system identification,” Automatica, vol. 62, pp. 148–153, Dec. 2015.
- C.-Y. Lee, A. L. Johnson, E. Moreno-Centeno, and T. Kuosmanen, “A more efficient algorithm for convex nonparametric least squares,” European Journal of Operational Research, vol. 227, no. 2, pp. 391–400, 2013.
- E. Lim and P. W. Glynn, “Consistency of multidimensional convex regression,” Operations Research, vol. 60, no. 1, pp. 196–208, 2012.
- B. Lin and N. M. Tran, “Linear and rational factorization of tropical polynomials,” arXiv preprint arXiv:1707.03332, 2017.
- J.-N. Lin, H.-Q. Xu, and R. Unbehauen, “A generalization of canonical piecewise-linear functions,” IEEE Trans. Circuits Syst. I, Fundam. Theory Appl., vol. 41, no. 4, pp. 345–347, Apr. 1994.
- G. L. Litvinov, “Maslov dequantization, idempotent and tropical mathematics: A brief introduction,” J. Math. Sci., vol. 140, no. 3, pp. 426–444, Jan. 2007.
- G. L. Litvinov, V. P. Maslov, and G. B. Shpiz, “Idempotent functional analysis: An algebraic approach,” Math. Notes, vol. 69, no. 5, pp. 696–729, 2001.
- Y. Lucet, “What shape is your conjugate? A survey of computational convex analysis and its applications,” SIAM J. Optim., vol. 20, no. 1, pp. 216–250, Jan. 2009.
- J.-H. Luo, J. Wu, and W. Lin, “ThiNet: A filter level pruning method for deep neural network compression,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 5058–5066.

Bibliography - 10

- A. Maas, A. Hannun, and A. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in Proc. Int. Conf. Mach. Learn. (ICML), 2013, p. 3.
- D. Maclagan and B. Sturmfels, Introduction to Tropical Geometry. Providence, RI, USA: AMS, 2015.
- A. Magnani and S. P. Boyd, “Convex piecewise-linear fitting,” Optim. Eng., vol. 10, no. 1, pp. 1–17, Mar. 2009.
- P. Manurangsi and D. Reichman, “The computational complexity of training ReLU(s),” 2018, arXiv:1810.04207. [Online]. Available: http://arxiv.org/abs/1810.04207
- P. Maragos, “Morphological systems: Slope transforms and max-min difference and differential equations,” Signal Process., vol. 38, no. 1, pp. 57–77, Jul. 1994.
- P. Maragos, “Lattice image processing: A unification of morphological and fuzzy algebraic systems,” J. Math. Imag. Vis., vol. 22, nos. 2–3, pp. 333–353, May 2005.
- P. Maragos, “Morphological filtering for image enhancement and feature detection,” in Image and Video Processing Handbook, A. Bovik, Ed., 2nd ed. Amsterdam, The Netherlands: Elsevier, 2005, pp. 135–156.
- P. Maragos, “Dynamical systems on weighted lattices: General theory,” Math. Control, Signals, Syst., vol. 29, no. 4, pp. 1–49, Dec. 2017.
- P. Maragos, “Tropical geometry, mathematical morphology and weighted lattices,” in Proc. Int. Symp. Math. Morphol., in Lecture Notes in Computer Science, vol. 11564, B. Burgeth, Ed. Cham, Switzerland: Springer, 2019, pp. 3–15.
- P. Maragos and E. Theodosis, “Multivariate tropical regression and piecewise-linear surface fitting,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2020, pp. 3822–3826.

Bibliography - 11

- V. P. Maslov, “On a new superposition principle for optimization problems,” Russian Math. Surv., vol. 42, no. 3, pp. 39–48, 1987.
- R. Mazumder, A. Choudhury, G. Iyengar, and B. Sen, “A computational framework for multivariate convex regression and its variants,” Journal of the American Statistical Association, vol. 114, no. 525, pp. 318–331, 2019.
- W. M. McEneaney, Max-Plus Methods for Nonlinear Control and Estimation. Cambridge, MA, USA: Birkhaüser, 2006.
- F. Meyer, Topographic Tools for Filtering and Segmentation 1 & 2. Hoboken, NJ, USA: Wiley, 2019.
- G. Mikhalkin, “Enumerative tropical algebraic geometry in R2,” J. Amer. Math. Soc., vol. 18, no. 2, pp. 313–377, 2005.
- M. Mohri, “Weighted automata algorithms,” in Handbook Weighted Automata. Berlin, Germany: Springer, 2009, pp. 213–254.
- M. Mohri, F. Pereira, and M. Riley, “Weighted finite-state transducers in speech recognition,” Comput. Speech Lang., vol. 16, no. 1, pp. 69–88, Jan. 2002.
- R. Mondal, S. Sundar Mukherjee, S. Santra, and B. Chanda, “Dense morphological network: An universal function approximator,” 2019, arXiv:1901.00109. [Online]. Available: http://arxiv.org/abs/1901.00109
- G. F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, “On the number of linear regions of deep neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2014.
- J.-J. Moreau, “Inf-convolution, Sous-additivité, Convexité des Fonctions Numériques,” J. Mathématiques Pures Appliquées, vol. 49, pp. 109–154, 1970.

Bibliography - 12

- A. Omanovic, H. Kazan, P. Oblak, and T. Curk, “Sparse data embedding and prediction by tropical matrix factorization,” BMC bioinformatics, vol. 22, no. 1, pp.1–18, 2021.
- A. Omanovic, P. Oblak, and T. Curk, “Application of tropical semiring for matrix factorization,” Uporabna informatika, vol. 28, no. 4, 2020.
- A. Omanovic, P. Oblak, and T. Curk, “Faststmf: “Efficient tropical matrix factorization algorithm for sparse data,” arXiv preprint arXiv:2205.06619, 2022.
- L. Pachter and B. Sturmfels, “Tropical geometry of statistical models,” Proc. Nat. Acad. Sci. USA, vol. 101, no. 46, pp. 16132–16137, Nov. 2004.
- L. Pachter and B. Sturmfels, “Parametric inference for biological sequence analysis,” Proc. Nat. Acad. Sci. USA, vol. 101, no. 46, pp. 16138–16143, Nov. 2004.
- R. Pascanu, G. Montufar, and Y. Bengio, “On the number of response regions of deep feed forward networks with piece-wise linear activations,” 2013, arXiv:1312.6098. [Online]. Available: http://arxiv.org/abs/1312.6098
- J.-E. Perin, “Tropical semirings,” in Idempotency, J. Gunawardena, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1998, pp. 50–69.
- L. F. C. Pessoa and P. Maragos, “Neural networks with hybrid morphological/rank/linear nodes: A unifying framework with applications to handwritten character recognition,” Pattern Recognit., vol. 33, no. 6, pp. 945–960, Jun. 2000.
- M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein, “On the expressive power of deep neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2017, pp. 2847–2854.
- G. X. Ritter and P. Sussner, “An introduction to morphological neural networks,” in Proc. 13th Int. Conf. Pattern Recognit., 1996.

Bibliography - 13

- G. X. Ritter, P. Sussner, and J. L. Diza-de-Leon, “Morphological associative memories,” IEEE Trans. Neural Netw., vol. 9, no. 2, pp. 281–293, Mar. 1998.
- G. X. Ritter and G. Urcid, “Lattice algebra approach to single-neuron computation,” IEEE Trans. Neural Netw., vol. 14, no. 2, pp. 282–295, Mar. 2003.
- R. T. Rockafellar, Convex Analysis. Princeton, NJ, USA: Princeton Univ. Press, 1970.
- B. De Schutter and B. De Moor, “Minimal realization in the max algebra is an extended linear complementarity problem,” Systems &Control Letters, vol. 25, no. 2, pp. 103–111, 1995.
- B. De Schutter and B. De Moor, “The extended linear complementarity problem,” Mathematical Programming, vol. 71, no. 3, pp. 289–325, 1995.
- B. De Schutter and B. De Moor, “Matrix factorization and minimal state space realization in the max-plus algebra,” in Proceedings of the 1997 American control conference (Cat. No. 97CH36041), vol. 5. IEEE, 1997, pp. 3136–3140.
- J. Serra, Image Analysis and Mathematical Morphology. New York, NY, USA: Academic, 1982.
- J. Serra, Ed., Image Analysis and Mathematical Morphology, vol. 2. New York, NY, USA: Academic, 1988.
- T. Serra and S. Ramalingam, “Empirical bounds on linear regions of deep rectifier networks,” in Proc. AAAI Conf. Artif. Intell., 2020, pp. 5628–5635.
- T. Serra, C. Tjandraatmadja, and S. Ramalingam, “Bounding and counting linear regions of deep neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2018, pp. 4558–4566.
- Y. Shitov, “The complexity of tropical matrix factorization,” Advances in Mathematics, vol. 254, pp. 138–156, 2014.

Bibliography - 14

- A. Siahkamari, D. A. E. Acar, C. Liao, K. L. Geyer, V. Saligrama, and B. Kulis, “Faster algorithms for learning convex functions,” in International Conference on Machine Learning. PMLR, 2022, pp. 20 176–20 194.
- I. Simon, “On semigroups of matrices over the tropical semiring,” RAIRO Theor. Informat. Appl., vol. 28, nos. 3–4, pp. 277–294, 1994.
- G. Smyrnis, P. Maragos, and G. Retsinas, “Maxpolynomial division with application to neural network simplification,” in Proc. ICASSP IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2020, pp. 4192–4196.
- G. Smyrnis and P. Maragos, “Multiclass neural network minimization via tropical newton polytope approximation,” in Proc. Int. Conf. Mach. Learn. (ICML), 2020, pp. 9068–9077.
- N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” J. Mach. Learn. Res., vol. 1, no. 15, pp. 1929–1958, 2014.
- P. Sussner and E. L. Esmi, “Morphological perceptrons with competitive learning: Lattice-theoretical framework and constructive learning algorithm,” Inf. Sci., vol. 181, no. 10, pp. 1929–1950, May 2011.
- J. M. Tarela and M. V. Martínez, “Region configurations for realizability of lattice piecewise-linear models,” Math. Comput. Model., vol. 30, nos. 11–12, pp. 17–27, Dec. 1999.
- M. Telgarsky, “Benefits of depth in neural networks,” in Proc. Conf. Learn. Theory, 2016, pp. 1517–1539.
- E. Theodosis and P. Maragos, “Analysis of the viterbi algorithm using tropical algebra and geometry,” in Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun., Jun. 2018, pp. 1–5.

Bibliography - 15

- E. Theodosis and P. Maragos, “Tropical modeling of weighted transducer algorithms on graphs,” in Proc. ICASSP IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 8653–8657.
- A. Toriello and J. P. Vielma, “Fitting piecewise linear continuous functions,” Eur. J. Oper. Res., vol. 219, no. 1, pp. 86–95, May 2012.
- N. M. Tran and J. Wang, “Minimal representations of tropical rational signomials,” arXiv preprint arXiv:2205.05647, 2022.
- A. Tsiamis and P. Maragos, “Sparsity in max-plus algebra,” Discrete Events Dyn. Syst., vol. 29, no. 2, pp. 163–189, 2019.
- T. van den Boom and B. D. Schutter, “Modeling and control of switching max-plus-linear systems with random and deterministic switching,” Discrete Event Dyn. Syst., vol. 22, no. 2, pp. 293–332, 2012.
- O. Viro, “Dequantization of real algebraic geometry on logarithmic paper,” in Proc. Eur. Congr. Math., 2001, pp. 135–146.
- S. Wang, “General constructive representations for continuous piecewise-linear functions,” IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 51, no. 9, pp. 1889–1896, Sep. 2004.
- P.-F. Yang and P. Maragos, “Min-max classifiers: Learnability, design and application,” Pattern Recognition, vol. 28, no. 6, pp. 879–899, Jun. 1995.
- L. Zhang, G. Naitzat, and L.-H. Lim, “Tropical geometry of deep neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2018, pp. 5824–5832.
- Y. Zhang et al., “Max-plus operators applied to filter selection and model pruning in neural networks,” in Proc. Int. Symp. Math. Morphol., in Lecture Notes in Computer Science, B. Burgeth, Eds., vol. 11564. Cham, Switzerland: Springer, 2019, pp. 310–322.
- U. Zimmermann, Linear and Combinatorial Optimization in Ordered Algebraic Structures. Amsterdam, The Netherlands: North Holland, 1981.