

Tropical Geometry and Piecewise-Linear Approximation of Curves and Surfaces on Weighted Lattices

Petros Maragos and Emmanouil Theodosis

Abstract Tropical Geometry and Mathematical Morphology share the same max-plus and min-plus semiring arithmetic and matrix algebra. In this chapter we summarize some of their main ideas and common (geometric and algebraic) structure, generalize and extend both of them using weighted lattices and a max- \star algebra with an arbitrary binary operation \star that distributes over max, and outline applications to geometry, machine learning, and optimization. Further, we generalize tropical geometrical objects using weighted lattices. Finally, we provide the optimal solution of max- \star equations using morphological adjunctions that are projections on weighted lattices, and apply it to optimal piecewise-linear regression for fitting max- \star tropical curves and surfaces to arbitrary data that constitute polygonal or polyhedral shape approximations. This also includes an efficient algorithm for solving the convex regression problem of data fitting with max-affine functions.

1 Introduction

As stated in [56], tropical geometry is a “marriage between algebraic geometry and polyhedral geometry”. It is a relatively recent field in mathematics and computer science. However, the scalar arithmetic of its analytic part pre-existed in the form of max-plus and min-plus semiring arithmetic used in finite automata, nonlinear functional and image analysis, convex analysis, nonlinear control and optimization. In this chapter we explore the parts it shares with morphological image analysis, ex-

Petros Maragos

National Technical University of Athens, School of Electrical and Computer Engineering, Athens 15773, Greece. e-mail: maragos@cs.ntua.gr.

Emmanouil Theodosis

Harvard University, School of Engineering and Applied Sciences, Cambridge, MA 02138, USA. e-mail: etheodosis@g.harvard.edu.

tend both of them using weighted lattices, and apply max-plus algebra to optimally fitting of tropical curves and surfaces to data.

Combinations of max-plus, or its dual min-plus, arithmetic with corresponding nonlinear matrix algebra and signal convolutions have been used in operations research and scheduling [23]; discrete event systems, max-plus control and optimization [2, 3, 4, 14, 21, 30, 32, 37, 66, 85]; convex analysis [75, 55]; morphological image analysis [38, 61, 67, 76, 77]; nonlinear PDEs of the Hamilton-Jacobi type and vision scale-spaces [13, 39]; speech recognition and natural language processing [44, 69]; neural networks [17, 18, 33, 72, 73, 89, 90, 91]; idempotent mathematics (nonlinear functional analysis) [54, 53].

Max-plus (a.k.a. ‘max-sum’) arithmetic forms an idempotent semiring denoted as $(\mathbb{R}_{\max}, \max, +)$ where $\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$ and the real number addition and multiplication are replaced by the max and sum operations, respectively. As an idempotent semiring it is covered by the theory of dioids [32]. The dual min-plus semiring $(\mathbb{R}_{\min}, \min, +)$, where $\mathbb{R}_{\min} = \mathbb{R} \cup \{+\infty\}$, has been called ‘tropical semiring’ and has been used in finite automata [45, 78], speech and language recognition using graphical models [69], and tropical geometry [56, 68]. In idempotent mathematics [54, 53], in convex optimization [12], and in the general theory of dioids [32], they often use the limit of the *Log-Sum-Exp* approximation for the max and min operations:

$$\begin{aligned} \lim_{\theta \downarrow 0} \theta \cdot \log(e^{a/\theta} + e^{b/\theta}) &= \max(a, b) \\ \lim_{\theta \downarrow 0} (-\theta) \log(e^{-a/\theta} + e^{-b/\theta}) &= \min(a, b) \end{aligned} \quad (1)$$

where $\theta > 0$ is usually called a ‘temperature’ parameter. This approximation is at the heart of the ‘Maslov Dequantization’ [65] of real numbers, which generates a whole family of semirings $S_\theta = (\mathbb{R}_{\max}, +_\theta, \times_\theta)$, $\theta > 0$, whose operations of generalized ‘addition’ $+_\theta$ and ‘multiplication’ \times_θ are defined as

$$\begin{aligned} a +_\theta b &:= \theta \cdot \log(e^{a/\theta} + e^{b/\theta}) = \phi_\theta^{-1}[\phi_\theta(a) + \phi_\theta(b)] \\ a \times_\theta b &:= a + b \end{aligned} \quad (2)$$

with $\phi_\theta(a) := \exp(a/\theta)$. This makes S_θ isomorphic to the semiring of nonnegative real numbers $\mathbb{R}_{\geq 0}$ equipped with standard addition and multiplication. This isomorphism is enabled via the logarithmic mapping $a \mapsto \phi_\theta^{-1}(a) = \theta \log(a)$. In the limit $\theta \downarrow 0$, we get S_0 which is the max-plus semiring. Thus, $(\mathbb{R}_{\geq 0}, +, \times)$ is isomorphic to $(\mathbb{R}_{\max}, \max, +)$.

Max and min operations (or more generally supremum and infimum) form the algebra of lattices, which has been used to generalize Euclidean morphology [76], based on Minkowski set operations and their extensions to functions via level sets, to more general morphological operators on complete lattices [77, 38, 40, 5]. The scalar arithmetic of morphology on functions has been mainly flat; a few exceptions include the max-plus convolutions and related operations which have appeared in morphological image analysis [38, 60, 76, 79], image algebra [74], convex analysis and optimization [6, 55], and nonlinear dynamical systems [4, 63]. Such non-flat morphological operations and their generalizations to a max- \star algebra have been

systematized and extended using the theory of weighted lattices [62, 63]. This connects morphology with max-plus algebra and tropical geometry.

Tropical Geometry and the standard image operators of Mathematical Morphology share the same max-plus and min-plus semiring arithmetic and matrix algebra. In this chapter, whose preliminary version is based on [64], we begin in Section 2 with some elementary concepts from classic Euclidean image morphological operators based on Minkowski set and function addition, duality pairs in the form of adjunctions (a.k.a. residuation pairs) and their formalization using lattice theory. Then in Section 3 we show how approximation (1) converts the linear heat PDE modeling the Gaussian scale-space in computer vision to PDEs generating multi-scale max-plus morphological operators. We continue in Section 4 with elementary concepts and objects of tropical geometry. Then, in Section 5 we summarize the theory of weighted lattices, which form nonlinear vector spaces, and use them to extend the max-plus mathematical morphology and tropical geometry using a max- \star algebra with an arbitrary binary operation \star that distributes over max. Further, we generalize tropical geometrical objects using weighted lattices. Finally, in Section 6 we outline the optimal solution of max- \star equations using morphological adjunctions (a.k.a. residuation pairs) that are projections on weighted lattices, and apply it in Section 7 to optimal convex piecewise-linear regression for fitting max- \star tropical curves and surfaces to arbitrary data that constitute polygonal or polyhedral shape approximations. Throughout the chapter, we also outline applications to numerical geometry, machine learning, and optimization.

Notation: For maximum (or supremum) and minimum (or infimum) operations we use the well-established lattice-theoretic symbols of \vee and \wedge . We do *not* use the notation (\oplus, \otimes) for $(\max, +)$ or $(\min, +)$ which is frequently used in max-plus algebra, because in image analysis i) the symbol \oplus is extensively used for Minkowski set operations and max-plus convolutions, and ii) \otimes is unnecessarily confusing compared to the classic symbol $+$ of addition. We use roman letters for functions, signals and their arguments and greek letters mainly for operators. Also, boldface roman letters for vectors (lowcase) and matrices (capital). If $\mathbf{M} = [m_{ij}]$ is a matrix, its (i, j) -th element is denoted as m_{ij} or as $[\mathbf{M}]_{ij}$. Similarly, $\mathbf{x} = [x_i]$ denotes a column vector, whose i -th element is denoted as $[\mathbf{x}]_i$ or simply x_i .

2 Elements of Max-plus Morphology and Flat Lattices

We view images, signals, and vectors as elements of *complete lattices* $(\mathcal{L}, \vee, \wedge)$, where \mathcal{L} is the set of lattice elements equipped with two binary operations, \vee and \wedge , which denote the lattice supremum and infimum respectively. Each of these operations induces a partial ordering \leq ; e.g. for any $X, Y \in \mathcal{L}$, $X \leq Y \iff Y = X \vee Y$. The lattice operations satisfy many properties, as summarized in Table 1. Conversely, a set \mathcal{L} equipped with two binary operations \vee and \wedge that satisfy properties (L1, L1')–(L5, L5') is a lattice whose supremum is \vee , the infimum is \wedge , and partial ordering \leq is given by (L6). Completeness means that the supremum and infimum of any (even

infinite) subset of \mathcal{L} exists and belongs to \mathcal{L} . A lattice $(\mathcal{L}, \vee, \wedge)$ contains two weaker substructures: a sup-semilattice (\mathcal{L}, \vee) that satisfies properties (L1–L4) and an inf-semilattice (\mathcal{L}, \wedge) that satisfies properties (L1'–L4'). Examples of complete lattices we use in computer vision include i) the lattice of Euclidean shapes, i.e. subsets of \mathbb{R}^n , equipped with set union and intersection, and ii) the lattice $\text{Fun}(E, \overline{\mathbb{R}})$ of functions with (arbitrary) domain E and values in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, equipped with the pointwise supremum and pointwise infimum of extended real numbers. For data processing, we also consider operators $\psi : \mathcal{L} \rightarrow \mathcal{M}$ between two complete lattices. The set of all such operators becomes a complete lattice if equipped with the supremum, infimum, and partial ordering defined pointwise for the operators' outputs.

Monotone Operators:

A lattice operator $\psi : \mathcal{L} \rightarrow \mathcal{M}$ is called *increasing* or *isotone* if it is order preserving; i.e. if, for any $X, Y \in \mathcal{L}$, $X \leq Y \implies \psi(X) \leq \psi(Y)$. (We use the same symbol for the partial order in \mathcal{L} and \mathcal{M} although they may be different, hoping that the difference will be clear from the context.) Examples of increasing operators are the lattice homomorphisms which preserve suprema and infima. If a lattice homomorphism is also a bijection, then it becomes an automorphism. Four fundamental types of increasing operators are: *dilations* δ and *erosions* ε that satisfy respectively $\delta(\bigvee_i X_i) = \bigvee_i \delta(X_i)$ and $\varepsilon(\bigwedge_i X_i) = \bigwedge_i \varepsilon(X_i)$ over arbitrary (possibly infinite) collections; *openings* α that are increasing, idempotent ($\alpha^2 = \alpha$), and antiextensive ($\alpha \leq \text{id}$), where id denotes the identity operator; *closings* β that are increasing, idempotent, and extensive ($\beta \geq \text{id}$).

A lattice operator ψ is called *decreasing* or *antitone* if it is order-inverting, i.e. $X \leq Y \implies \psi(X) \geq \psi(Y)$. Dual homomorphisms interchange suprema with infima and hence are decreasing operators. For example, *anti-dilations* δ^a satisfy $\delta^a(\bigvee_i X_i) = \bigwedge_i \delta^a(X_i)$. A lattice dual automorphism is a bijection that interchanges suprema with infima. For example, a *negation* v is a dual automorphism that is also involutive, i.e. $v^2 = \text{id}$.

Residuation and Adjunctions:

An increasing operator $\psi : \mathcal{L} \rightarrow \mathcal{M}$ between two complete lattices is called *residuated* [10, 9] if there exists an increasing operator $\psi^\sharp : \mathcal{M} \rightarrow \mathcal{L}$ such that

$$\psi\psi^\sharp \leq \text{id} \leq \psi^\sharp\psi \quad (3)$$

ψ^\sharp is called the **residual** of ψ , is unique, and is the closest to being an inverse of ψ . Specifically, the residuation pair (ψ, ψ^\sharp) can solve inverse problems of the type $\psi(X) = Y$ 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* in the sense that

$$\hat{X} = \psi^\sharp(Y) = \bigvee \{X : \psi(X) \leq Y\} \quad (4)$$

On complete lattices an increasing operator ψ is residuated (resp. a residual ψ^\sharp) if and only if it is a dilation (resp. erosion). Equivalently, ψ is residuated if $\psi^\sharp(Y)$,

defined as in (4), exists for each Y . The residuation theory has been used for solving inverse problems (mainly in matrix algebra) over the extended max-plus semiring $(\mathbb{R}, \vee, +)$ or other complete idempotent semirings which as lattices are made complete [24, 23, 4, 22].

A pair (δ, ε) of two operators $\delta : \mathcal{L} \rightarrow \mathcal{M}$ and $\varepsilon : \mathcal{M} \rightarrow \mathcal{L}$ between two complete lattices is called **adjunction**¹ if

$$\delta(X) \leq Y \iff X \leq \varepsilon(Y) \quad \forall X \in \mathcal{L}, Y \in \mathcal{M} \quad (5)$$

In any adjunction, δ is a dilation and ε is an erosion. The double inequality (5) is equivalent to the inequality (3) satisfied by a residuation pair of increasing operators if we identify the residuated map ψ with δ and its residual ψ^\sharp with ε . Further, from (5) or (3) it follows that any adjunction (δ, ε) automatically yields an opening $\alpha = \delta\varepsilon$ and a closing $\beta = \varepsilon\delta$, where the composition of two operators is written as an operator product. To view (δ, ε) as an adjunction instead of a residuation pair has the advantage of the additional geometrical intuition and visualization afforded by the dilation and erosion operators in image and shape analysis.

There is a one-to-one correspondence between the two operators of an adjunction; e.g., given a dilation δ , there is a unique erosion

$$\varepsilon(Y) = \delta^\sharp(Y) = \bigvee \{X \in \mathcal{L} : \delta(X) \leq Y\} \quad (6)$$

such that (δ, ε) is an adjunction, and conversely. Thus, dilations and erosions on complete lattices always come in pairs. In any adjunction (δ, ε) , ε is called the *adjoint erosion* (a.k.a. *upper adjoint*) of δ , whereas δ is the *adjoint dilation* (a.k.a. *lower adjoint*) of ε .

Example 1 (a) A classic example of a morphological set adjunction is the pair of Minkowski set addition \oplus and subtraction \ominus : for $X, B \subseteq \mathbb{R}^n$

$$\begin{aligned} \delta_B(X) &= X \oplus B := \{\mathbf{x} \in \mathbb{R}^n : B_{+\mathbf{x}} \cap X \neq \emptyset\} \\ \varepsilon_B(X) &= X \ominus B := \{\mathbf{x} \in \mathbb{R}^n : B_{+\mathbf{x}} \subseteq X\} \end{aligned} \quad (7)$$

where $B^s = \{-\mathbf{b} : \mathbf{b} \in B\}$ and $B_{+\mathbf{x}} := \{\mathbf{b} + \mathbf{x} : \mathbf{b} \in B\}$.

(b) A classic example of a morphological signal adjunction is the pair of Minkowski function addition \oplus and subtraction \ominus : for $f, g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$

¹ As explained in [5, 38, 40], the adjunction is related to *Galois connection*, which is a pair of two decreasing maps ψ and ϕ between complete lattices, such that $Y \leq \psi(X) \iff X \leq \phi(Y)$; if this holds, both ψ and ϕ are anti-dilations and their compositions $\psi\phi$ and $\phi\psi$ are closings [1]. As a name, ‘adjunction’ was introduced in [31] as equivalent to an isotone Galois connection. The advantage of residuations and adjunctions over Galois connections is that the former can form new adjunctions via composition, whereas this is not the case with (antitone) Galois connections. Several authors define residuations (ψ, ψ^\sharp) on partially ordered sets; this however may not guarantee the general existence of (4) and the fact that the (ψ, ψ^\sharp) are a dilation and erosion respectively, unless the underlying sets become complete lattices.

$$\begin{aligned}\delta_g(f)(\mathbf{x}) &= f \oplus g(\mathbf{x}) := \sup\{f(\mathbf{y} - \mathbf{x}) + g(\mathbf{y}) : \mathbf{y} \in \mathbb{R}^n\} \\ \varepsilon_g(f)(\mathbf{x}) &= f \ominus g(\mathbf{x}) := \inf\{f(\mathbf{x} - \mathbf{y}) - g(\mathbf{y}) : \mathbf{y} \in \mathbb{R}^n\}\end{aligned}\quad (8)$$

Thus, $f \oplus g$ is the supremal (max-plus) convolution of f by g and $f \ominus g$ is the infimal convolution of $f(\mathbf{x})$ by $-g(-\mathbf{x})$.

3 Tropical Dequantization and Vision Scale-Space PDEs

By following the procedure in Maslov [65], we show in this section how the transformation of (1) converts the classic linear heat PDE to a nonlinear PDE that generates multiscale erosions (min-plus convolutions). Consider the 1D linear heat PDE

$$\frac{\partial U}{\partial t} = \frac{\theta}{2} \frac{\partial^2 U}{\partial x^2} \quad (9)$$

which models a homogeneous linear diffusion. It is also well known in computer vision because it models the Gaussian scale-space since its solution $U(x, t)$, $t \geq 0$, is the multiscale convolution of some initial function $f(x) = U(x, 0)$ with multiscale Gaussians $G_\sigma(x) = \exp(-x^2/2\sigma^2)/(\sigma\sqrt{2\pi})$ of variance equal to $\sigma^2 = \theta t$:

$$U(x, t) = \frac{1}{\sqrt{2\pi\theta t}} \int_{\mathbb{R}} f(x-y) \exp(-\frac{y^2}{2\theta t}) dy \quad (10)$$

As shown by Maslov [65], the substitution $U = \exp(-W/\theta)$ converts the heat PDE to Hopf's nonlinear equation:

$$\frac{\partial W}{\partial t} + \frac{1}{2} \left(\frac{\partial W}{\partial x} \right)^2 - \frac{\theta}{2} \frac{\partial^2 W}{\partial x^2} = 0 \quad (11)$$

The heat PDE (10) obeys a linear superposition; i.e., if $u_i(x, t)$ is its solution for initial condition $f_i(x)$, $i = 1, 2$, and if $f(x) = a_1 f_1(x) + a_2 f_2(x)$, the total solution becomes $U(x, t) = a_1 u_1(x, t) + a_2 u_2(x, t)$. However, the solution $W = -\theta \log U$ of the nonlinear PDE (11) obeys the following nonlinear superposition:

$$W(x, t) = -\theta \log[\exp(-\frac{b_1 + w_1}{\theta}) + \exp(-\frac{b_2 + w_2}{\theta})] = -\theta \log[c_1 u_1 + c_2 u_2] \quad (12)$$

where $w_i(x, t) = -\theta \log u_i(x, t)$, $i = 1, 2$, are solutions of (11) and $c_i = \exp(-b_i/\theta)$. If the heat diffusivity constant θ becomes very small, we can perform Maslov's dequantization as in (1) to convert the above log-sum-exp superposition to a tropical (min-plus) superposition $W = \min(b_1 + w_1, b_2 + w_2)$. Further, the limit of the PDE (11) for $\theta \rightarrow 0$ yields another nonlinear PDE

$$\frac{\partial S}{\partial t} + \frac{1}{2} \left(\frac{\partial S}{\partial x} \right)^2 = 0 \quad (13)$$

which models the multiscale weighted erosion $g(x) \ominus k_t(x)$ of an initial function $g(x) = S(x, 0)$ by multiscale parabolas $k_t(x) = -x^2/2t$:

$$S(x, t) = g(x) \ominus k_t(x) = \bigwedge_y g(x - y) + y^2/2t \quad (14)$$

Thus, classic morphological PDEs [13] can be obtained from the linear heat PDE (modeling Gaussian scale space) via tropicalization.

4 Elements of Tropical Geometry

We first present some simple examples of tropical² curves and surfaces which result from tropicalizing the polynomials that analytically describe their Euclidean counterparts; here ‘tropicalization’ means replacing sum with max or min and multiplications with additions. Then, we explain this tropicalization as a dequantization of real algebraic geometry.

4.1 Examples of Tropical Polynomial Curves and Surfaces

Tropical Polynomial Curves:

Consider the analytic expressions for a Euclidean line, parabola, and cubic curve:

$$\begin{aligned} p_1(x) &= a \cdot x + b, & p_2(x) &= a \cdot x^2 + b \cdot x + c, \\ p_3(x) &= a \cdot x^3 + b \cdot x^2 + c \cdot x + d, \end{aligned} \quad (15)$$

The equations for their corresponding max-plus tropical polynomials are:

$$\begin{aligned} p_1^{\max}(x) &= \max(a + x, b), & p_2^{\max}(x) &= \max(a + 2 \cdot x, b + x, c), \\ p_3^{\max}(x) &= \max(a + 3 \cdot x, b + 2 \cdot x, c + x, d), \end{aligned} \quad (16)$$

The equations for the min-plus case are identical as in (16) by replacing max with min. The graphs of all the above can be seen in Fig. 1.

Tropical Polynomial Surfaces:

Consider the equations of the following tropical planes represented as 2D max-plus and min-plus polynomial of degree 1:

² The adjective ‘tropical’ was coined by French mathematicians, including Dominique Perrin and Jean-Eric Pin, to honor their Brazilian colleague Imre Simon who was one of the pioneers of min-plus algebra as applied to automata. However, we give it an alternative and substantial meaning in connection with its Greek origin word ‘τροπικός’, which comes from the Greek word ‘τροπή’ which means “turn” or “changing the way/direction”, to literally express the fact that tropical curves and surfaces bend and turn.

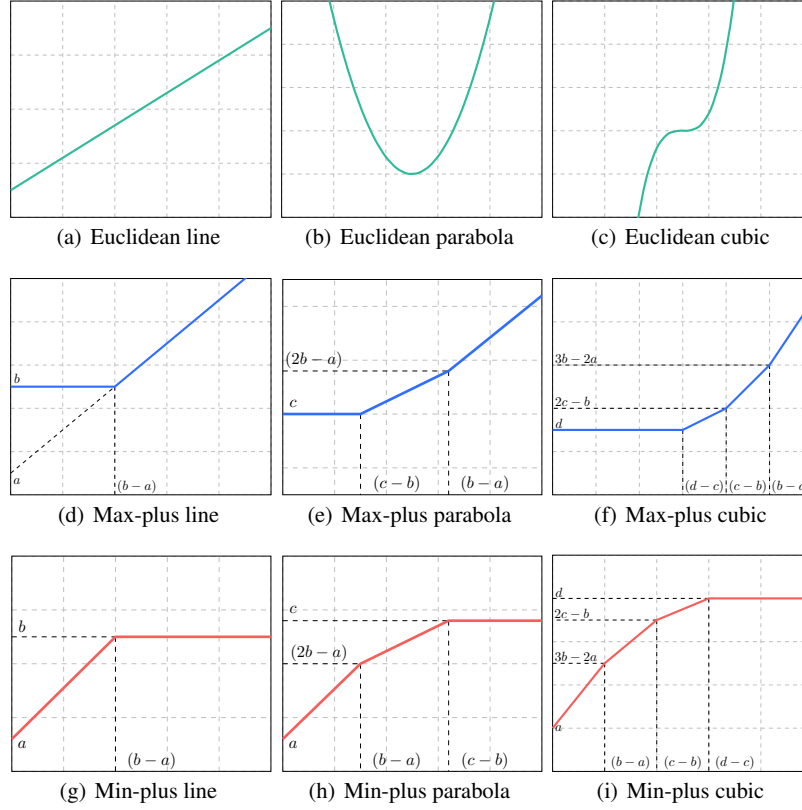


Fig. 1 Euclidean and tropical 1D polynomials up to 3rd degree.

$$f_1(x, y) = \max(0 + x, 2 + y, 7), \quad f_2(x, y) = \min(5 + x, 7 + y, 9), \quad (17)$$

whose graphs can be seen as surfaces in Fig. 2.

As a next example, to the general Euclidean conic polynomial

$$p_{\text{e-conic}}(x, y) = ax^2 + bxy + cy^2 + dx + ey + f \quad (18)$$

there corresponds the following two-variable max-plus tropical polynomial of degree 2:

$$p_{\text{t-conic}}(x, y) = \max(a + 2x, b + x + y, c + 2y, d + x, e + y, f) \quad (19)$$

Its min-plus version is shown in Fig. 3.

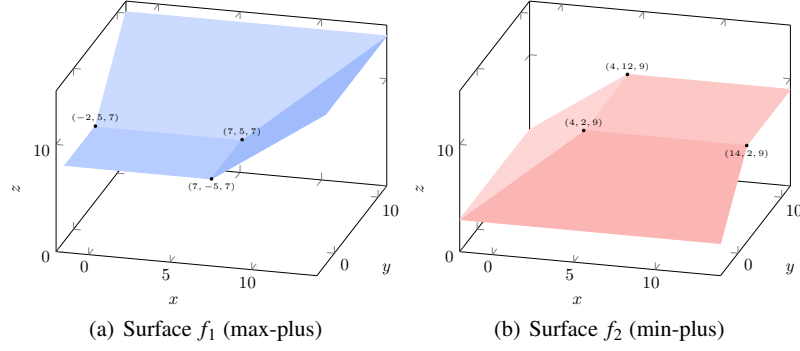


Fig. 2 Surfaces of the two tropical planes in (17).

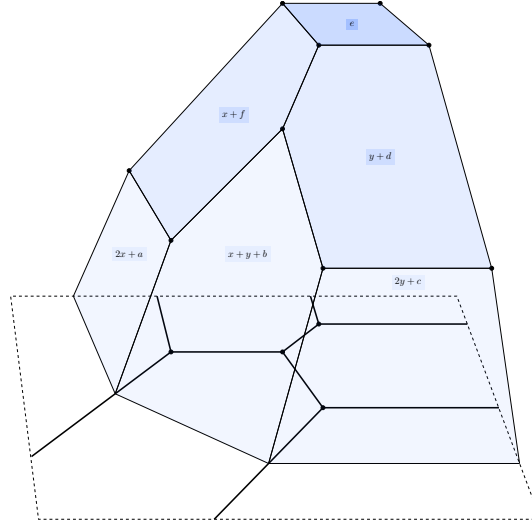


Fig. 3 Surface (graph) of the 2D min-plus tropical polynomial function $p(x, y) = \min(a + 2x, b + x + y, c + 2y, d + x, e + y, f)$ and its tropical quadratic curve. (Inspired by Fig. 1.3.2 of [56].)

4.2 Tropical Polynomials as Dequantization of Algebraic Geometry

The algebraic side of tropical geometry [56] results from a transformation of analytic Euclidean geometry where the traditional arithmetic of the real field $(\mathbb{R}, +, \times)$ involved in the analytic expressions of geometric objects is replaced by the arithmetic of the min-plus tropical semiring $(\mathbb{R}_{\min}, \wedge, +)$; some authors use its max-plus dual semiring $(\mathbb{R}_{\max}, \vee, +)$. We use both semirings as dual parts of the weighted lattice $(\mathbb{R}, \vee, \wedge, +)$ (explained in Sec. 5). This transformation converts Euclidean objects into polygonal lines on the plane and polyhedra in higher dimensions. A geometric explanation and visualization of this transformation is obtained from

Viro's graphing of polynomial curves on log-log paper [86]. Consider the monomial curve $v = cu^a$, $c > 0$, on the positive quadrant of the (u, v) plane and consider the log-log transformation of both coordinates composed with a uniform scaling by $\theta > 0$: $x = \theta \log u$, $y = \theta \log v$. Then, on the (x, y) plane the curve becomes the line $y = b/\theta + ax$, where $b = \log c$. If we have a K -term polynomial curve $v = P(u) = \sum_{k=1}^K c_k u^{a_k}$ with $c_k = \exp(b_k) > 0$ and $a_k \in \mathbb{R}$ (i.e. a posynomial [11]) then we convert it to

$$P_\theta(x) = \theta \log \left[\sum_{k=1}^K \exp(b_k/\theta) \exp(a_k x/\theta) \right] \quad (20)$$

As $\theta \downarrow 0$ this yields via Maslov dequantization a K -term **1D max-plus tropical polynomial**

$$p(x) = \max_{k=1}^K [b_k + a_k x] \quad (21)$$

While each $P_\theta(x)$ is a smooth function, their limit $p(x)$ is a max-affine function and represents a *piecewise-linear (PWL)* convex function. Note: if we perform dequantization with negative exponents as in (1) we obtain a min-plus polynomial which is a PWL concave function.

The above procedure extends to multiple dimensions or higher degrees and shows us the way to tropicalize any classic n -variable polynomial (linear combination of power monomials) $\sum_k c_k u_1^{a_{k1}} \cdots u_n^{a_{kn}}$ defined over $\mathbb{R}_{>0}^n$ where $c_k > 0$ and $\mathbf{a}_k = (a_{k1}, \dots, a_{kn})^T$ is traditionally some nonnegative integer³ vector but herein we allow $\mathbf{a}_k \in \mathbb{R}^n$: replace the sum with max and log the individual terms so that the multiplicative coefficients become additive and the powers become real multiples of the indefinite log variables. Thus, a general n -variable max-plus polynomial $p: \mathbb{R}^n \rightarrow \mathbb{R}$ has the expression:

$$p(\mathbf{x}) = \bigvee_{k=1}^K b_k + \mathbf{a}_k^T \mathbf{x}, \quad \mathbf{x} = (x_1, \dots, x_n)^T \quad (22)$$

where $K = \text{rank}(p)$ is the number of terms of p . Its graph (hypersurface) is a max of K hyperplanes with intercepts $b_k = \log c_k \in \mathbb{R}$ and real slope vectors $\mathbf{a}_k \in \mathbb{R}^n$. The degree of p is $|\mathbf{a}| = \max_k |\mathbf{a}_k|$ where $|\mathbf{a}_k| = |a_{k1}| + \cdots + |a_{kn}|$. Thus, the curves or surfaces of real algebraic geometry, which is essentially polynomial geometry, become via dequantization the graphs of *convex* PWL functions represented by tropical (max-plus) polynomials.

³ Traditionally, 'tropical polynomials' assume that the parameters a_{ki} are nonnegative integers. If we also allow negative integers, we get 'Laurent tropical polynomials'. As in [14], we allow any real coefficients; this may be called 'tropical posynomials' [16].

4.3 Tropical Curves and Newton Polytopes

To the zero set of a classic polynomial there corresponds the *tropical curve or surface* of a max-plus tropical polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$

$$\mathcal{V}(p) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \text{more than one terms of } p(\mathbf{x}) \text{ attain the max}\} \quad (23)$$

The above also defines the tropical curve of min-plus polynomials by replacing max with min. Thus, $\mathcal{V}(p)$ consists of the singularity points (of non-differentiability) of $p(\mathbf{x})$. Examples are shown in Fig. 4 for degree-1 tropic polynomials and in Fig. 3 for a degree-2 polynomial. The max-plus line $y = \max(a+x, b)$ of Fig. 1 and the tropical curve of the max-plus polynomial $\max(a+x, b+y, c)$ of Fig. 4 are special cases of the general family of the 12 max-plus line types of \mathbb{R}_{\max}^2 given in [22] by

$$\max(a+x, b+y, c) = \max(a'+x, b'+y, c'), \quad a, b, c, a', b', c' \in \mathbb{R}_{\max}, \quad (24)$$

where not all the coefficients are needed. This is a tropical version of the Euclidean line equation $ax + by + c = 0$.

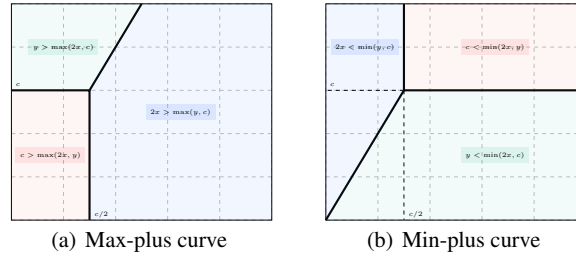


Fig. 4 Tropical curve of the max-polynomial $p(x,y) = \max(2x,y,c)$ left and its dual min-polynomial $p'(x,y) = \min(2x,y,c)$ right. Best viewed in color.

Another interesting geometric object related to a max-plus polynomial p is its *Newton polytope* which is the convex hull (denoted by $\text{conv}(\cdot)$) of the set of points represented by its slope coefficient vectors:

$$\text{New}(p) \triangleq \text{conv}\{\mathbf{a}_k : k = 1, \dots, \text{rank}(p)\} \quad (25)$$

This satisfies several important properties [17]:

$$\text{New}(p_1 \vee p_2) = \text{conv}(\text{New}(p_1) \cup \text{New}(p_2)) \quad (26)$$

$$\text{New}(p_1 + p_2) = \text{New}(p_1) \oplus \text{New}(p_2) \quad (27)$$

Examples are shown in Fig. 5. Thus, the Newton polytope of the sum (resp. max) of two tropical polynomials is the Minkowski sum (resp. the convex hull of the union) of their individual polytopes.

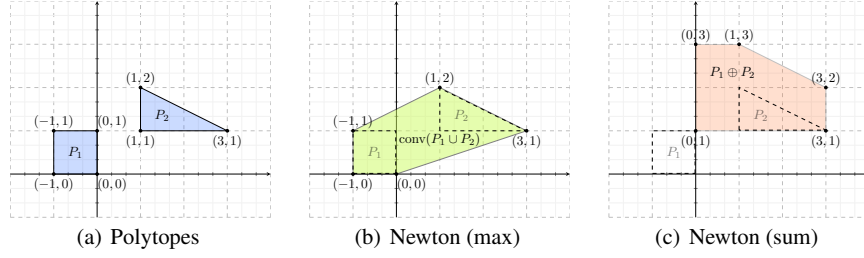


Fig. 5 Newton polytopes of (a) two max-polynomials $p_1(x, y) = \max(x + y, 3x + y, x + 2y)$ and $p_2(x, y) = \max(0, -x, y, y - x)$, (b) their max $p_1 \vee p_2$, and (c) their sum $p_1 + p_2$.

4.4 Tropical Halfspaces and Polytopes

In pattern analysis problems on Euclidean spaces \mathbb{R}^{n+1} we often use halfspaces $\mathcal{H}(\mathbf{a}, b) := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} \leq b\}$, polyhedra (finite intersections of halfspaces), and polytopes (compact polyhedra formed as the convex hull of a finite set of points). Replacing linear inner products $\mathbf{a}^T \mathbf{x}$ with max-plus versions yields *tropical halfspaces* [29] with parameters $\mathbf{a} = [a_i], \mathbf{b} = [b_i] \in \mathbb{R}^{n+1}$:

$$\mathcal{T}(\mathbf{a}, \mathbf{b}) \triangleq \{\mathbf{x} \in \mathbb{R}^n_{\max} : \max(a_{n+1}, \bigvee_{i=1}^n a_i + x_i) \leq \max(b_{n+1}, \bigvee_{i=1}^n b_i + x_i)\} \quad (28)$$

where $\min(a_i, b_i) = -\infty \forall i$. Thus, for each i , only one coefficient is needed either in the left or in the right side of inequality (28). Replacing max with min yields tropical halfspaces with dual boundaries that are min-plus hyperplanes. Examples of polytopes in the plane that are polygonal regions formed by min-plus tropical halfplanes are shown in Fig. 6. Obviously, their separating boundaries are tropical lines. Such regions in multiple dimensions were used in [17, 18, 89] as morphological perceptrons.

As an example in 3D space, in Fig. 7 we can see two different views of the intersection of the tropical halfspaces corresponding to the two tropical polynomial in (17). This is a polytope that is the polyhedral region formed by intersecting the halfspace above the surface of the 2D max-plus polynomial f_1 with the halfspace below the surface of the min-plus polynomial f_2 .

We note from Fig. 6 and Fig. 7 that the number of tropical boundaries required to form polytopes, which could serve as decision regions in pattern classification problems, is smaller than the number of linear boundaries. See, for instance, the polytope R_P in Fig. 6(b). This observation remains valid in higher dimensions too; namely, decision regions can be formed with fewer tropical lines or hyper-planes (and hence with fewer parameters) than their Euclidean counterparts because any single tropical line or hyper-plane bends and turns and hence provides more than one straight line edge or flat face.

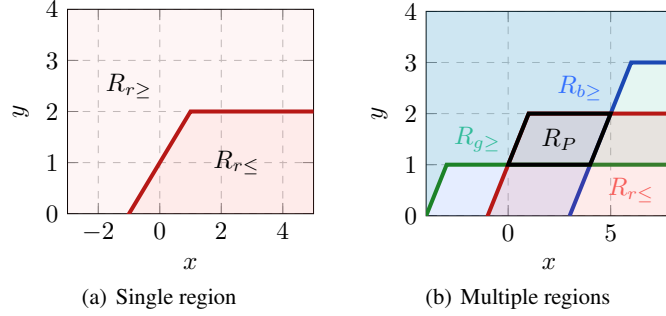


Fig. 6 Regions $R_{c\geq}$ and $R_{c\leq}$ formed by min-plus tropical halfspaces in \mathbb{R}^2 , where c denotes the color of the tropical boundary and ≥ 0 (resp. ≤ 0) the set of points above (resp. below) the boundary. (a) The red boundary is the min-plus tropical line $y = \min(1+x, 2)$. (b) The green and blue boundaries are respectively the tropical lines $y = \min(4+x, 1)$ and $y = \min(x-3, 3)$. R_P is the polytope formed by the intersection of three tropical halfplanes. Best viewed in color.

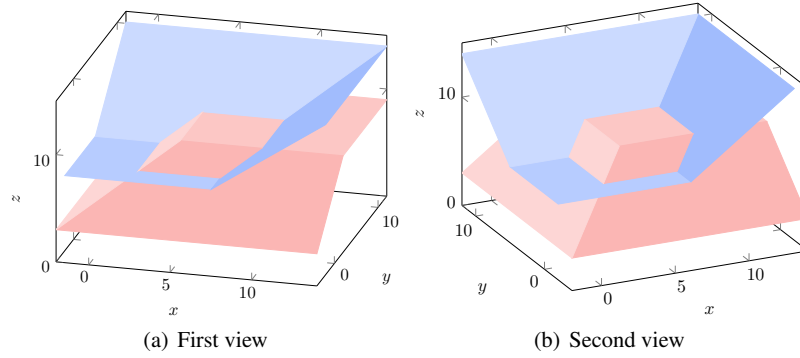


Fig. 7 Intersection of halfspaces of the 2D max-plus and min-plus tropical polynomials in (17). Best viewed in color.

5 Weighted Lattices: Nonlinear Vector Spaces and Extensions of Tropical Algebra and Geometry

5.1 Clodum: Extending Tropical Scalar Arithmetic

A lattice $(\mathcal{K}, \vee, \wedge)$ is often endowed with a third binary operation, called symbolically the ‘multiplication’ \star , under which (\mathcal{K}, \star) is a group, or a monoid, or just a semigroup [7]. Even if we have only a sup-semilattice (\mathcal{K}, \vee) (i.e. an idempotent commutative semigroup) we can consider its supremum \vee as an idempotent ‘addition’ and equip it with an additional ‘multiplication’ operation \star so that the structure $(\mathcal{K}, \vee, \star)$ becomes an idempotent semiring. Such ordered monoids have been studied in detail in [7, 92, 32] and form the algebraic basis of max-plus algebra.

Consider now an algebra $(\mathcal{K}, \vee, \wedge, \star, \star')$ with four binary operations, which we call a *lattice-ordered double monoid*, where $(\mathcal{K}, \vee, \wedge)$ is a lattice, (\mathcal{K}, \star) is a monoid whose ‘multiplication’ \star distributes over \vee , and (\mathcal{K}, \star') is a monoid whose ‘multiplication’ \star' distributes over \wedge . These distributivities imply that both \star and \star' are increasing. To the above definitions we add the word *complete* if \mathcal{K} is a complete lattice and the distributivities involved are infinite. We call the resulting algebra a *complete lattice-ordered double monoid*, in short *clodum* [60, 62, 63]. Previous works on minimax or max-plus algebra have used alternative names for structures similar to the above definitions which emphasize semigroups and semirings instead of lattices [4, 23, 32]; see [63] for similarities and differences. We precisely define an algebraic structure $(\mathcal{K}, \vee, \wedge, \star, \star')$ to be a **clodum** if:

(C1) $(\mathcal{K}, \vee, \wedge)$ is a complete distributive lattice. Thus, it contains its least $\perp := \bigwedge \mathcal{K}$ and greatest element $\top := \bigvee \mathcal{K}$. The supremum \vee (resp. infimum \wedge) plays the role of a generalized ‘addition’ (resp. ‘dual addition’).

(C2) (\mathcal{K}, \star) is a monoid whose operation \star plays the role of a generalized ‘multiplication’ with identity (‘unit’) element e and is a dilation.

(C3) (\mathcal{K}, \star') is a monoid with identity e' whose operation \star' plays the role of a generalized ‘dual multiplication’ and is an erosion.

Remarks: (i) As a lattice, \mathcal{K} is not necessarily infinitely distributive, although herein all our examples will be such.

(ii) The clodum ‘multiplications’ \star and \star' do not have to be commutative.

(iii) The least (greatest) element \perp (\top) of \mathcal{K} is both the ‘zero’ element for the ‘addition’ \vee (\wedge) and an absorbing null for the ‘multiplication’ \star (\star').

(iv) We avoid degenerate cases by assuming that $\vee \neq \star$ and $\wedge \neq \star'$. However, \star may be the same as \star' , in which case we have a self-dual ‘multiplication’.

A clodum \mathcal{K} is called *self-conjugate* if it has a lattice negation $a \mapsto a^\neg$ such that

$$\left(\bigvee_i a_i\right)^\neg = \bigwedge_i a_i^\neg, \quad \left(\bigwedge_i b_i\right)^\neg = \bigvee_i b_i^\neg, \quad (a \star b)^\neg = a^\neg \star' b^\neg \quad (29)$$

The first two above properties are generalization of De Morgan’s laws in Boolean algebras. We assume that the suprema and infima in (29) may be over any (possibly infinite) collections.

If $\star = \star'$ over $G = \mathcal{K} \setminus \{\perp, \top\}$ where (G, \star) is a group and (G, \vee, \wedge) a conditionally complete lattice, then the clodum \mathcal{K} becomes a richer structure which we call a *complete lattice-ordered group*, in short **clog**. In any clog the distributivity between \vee and \wedge is of the infinite type and the ‘multiplications’ \star and \star' are commutative. Then, for each $a \in G$ there exists its ‘multiplicative inverse’ a^{-1} such that $a \star a^{-1} = e$. Further, the ‘multiplication’ \star and its self-dual \star' (which coincide over G) can be extended over the whole \mathcal{K} by involving the null elements. A clog becomes self-conjugate by setting $a^\neg = a^{-1}$ if $\perp < a < \top$, $\top^\neg = \perp$, and $\perp^\neg = \top$. In a clog \mathcal{K} the \star and \star' coincide in all cases with only one exception: the combination of the least and greatest elements; thus, we may occasionally denote the clog algebra as $(\mathcal{K}, \vee, \wedge, \star)$.

Example 2 (a) *Max-plus clog* $(\overline{\mathbb{R}}, \vee, \wedge, +, +')$: \vee/\wedge denote the standard sup/inf on $\overline{\mathbb{R}}$, $+$ is the standard addition on $\overline{\mathbb{R}}$ playing the role of a ‘multiplication’ \star with $+$ being the ‘dual multiplication’ \star' ; the operations $+$ and $+$ are identical for finite reals, but $a + (-\infty) = -\infty$ and $a +' (+\infty) = +\infty$ for all $a \in \overline{\mathbb{R}}$. The identities are $e = e' = 0$, the nulls are $\perp = -\infty$ and $\top = +\infty$, and the conjugation mapping is $a^\neg = -a$. Thus, $+$ and $+$ are respectively the ‘lower addition’ and ‘upper addition’ used in convex analysis [70].

(b) *Max-times clog* $([0, +\infty], \vee, \wedge, \times, \times')$: The identities are $e = e' = 1$, the nulls are $\perp = 0$ and $\top = +\infty$, and the conjugation mapping is $a^\neg = 1/a$. The scalar multiplications \times and \times' coincide over $(0, +\infty)$, but $a \times 0 = 0$ and $a \times' (+\infty) = +\infty$ for all $a \in [0, +\infty]$.

(c) *Max-min clodum* $([0, 1], \vee, \wedge, \min, \max)$: As ‘multiplications’ we have $\star = \min$ and $\star' = \max$. The identities and nulls are $e' = \perp = 0$, $e = \top = 1$. A possible conjugation mapping is $a^\neg = 1 - a$. Additional clodums that are not clogs are discussed in [60, 63] using more general fuzzy intersections and unions.

(d) *Max-softmin clodum* $(\overline{\mathbb{R}}, \vee, \wedge, +_{-\theta}, +_{\theta})$, $\theta > 0$: As ‘multiplication’ we have $\star = +_{-\theta}$ and as ‘dual multiplication’ $\star' = +_{\theta}$, defined in the log-sum-exp approximation (2). The identities and nulls are $e' = \perp = -\infty$, $e = \top = +\infty$, and the conjugation mapping is $a^\neg = -a$. By varying $\theta > 0$ we obtain a family of clodums whose ‘multiplications’ \star and \star' are smooth (‘soft’) versions of the min and max operations respectively. In the limit as $\theta \downarrow 0$ this family converges to a max-min clodum over $\overline{\mathbb{R}}$.

(e) *Matrix max- \star clodum*: $(\mathcal{K}^{n \times n}, \vee, \wedge, \boxtimes, \boxtimes')$ where $\mathcal{K}^{n \times n}$ is the set of $n \times n$ matrices with entries from a clodum \mathcal{K} , \vee/\wedge denote here elementwise matrix sup/inf, and \boxtimes, \boxtimes' denote max- \star and min- \star' matrix ‘multiplications’:

$$\mathbf{C} = \mathbf{A} \boxtimes \mathbf{B} = [c_{ij}], \quad c_{ij} = \bigvee_{k=1}^n a_{ik} \star b_{kj}, \quad \mathbf{D} = \mathbf{A} \boxtimes' \mathbf{B} = [d_{ij}], \quad d_{ij} = \bigwedge_{k=1}^n a_{ik} \star' b_{kj}$$

This is a clodum with non-commutative ‘multiplications’. For the max-plus clog $(\overline{\mathbb{R}}, \vee, \wedge, +, +')$, these matrix ‘multiplications’ are defined and denoted as

$$[\mathbf{A} \boxplus \mathbf{B}]_{ij} \triangleq \bigvee_{k=1}^n a_{ik} + b_{kj}, \quad [\mathbf{A} \boxplus' \mathbf{B}]_{ij} \triangleq \bigwedge_{k=1}^n a_{ik} +' b_{kj} \quad (30)$$

5.2 Complete Weighted Lattices: Nonlinear Vector Spaces

Consider a nonempty collection \mathcal{W} of mathematical objects, which will be our space; examples of such objects include the vectors in $\overline{\mathbb{R}}^n$ or signals in $\text{Fun}(E, \overline{\mathbb{R}})$. Also, consider a clodum $(\mathcal{K}, \vee, \wedge, \star, \star')$ of **scalars** with *commutative* operations \star, \star' and $\mathcal{K} \subseteq \overline{\mathbb{R}}$. We define *two internal operations* among vectors/signals X, Y in \mathcal{W} : their supremum $X \vee Y : \mathcal{W}^2 \rightarrow \mathcal{W}$ and their infimum $X \wedge Y : \mathcal{W}^2 \rightarrow \mathcal{W}$, which we denote using the same supremum symbol (\vee) and infimum symbol (\wedge) as in the

clodum, hoping that the differences will be clear to the reader from the context. Further, we define *two external operations* among any vector/signal X in \mathcal{W} and any scalar c in \mathcal{K} : a ‘scalar multiplication’ $c \star X : (\mathcal{K}, \mathcal{W}) \rightarrow \mathcal{W}$ and a ‘scalar dual multiplication’ $c \star' X : (\mathcal{K}, \mathcal{W}) \rightarrow \mathcal{W}$, again by using the same symbols as in the clodum. Now, we define \mathcal{W} to be a **weighted lattice** space over the clodum \mathcal{K} if for all $X, Y, Z \in \mathcal{W}$ and $a, b \in \mathcal{K}$ all the axioms of Table 1 hold. Note that, under axioms L1-L9 and their duals L1'-L9', \mathcal{W} is a distributive lattice with a least element O and a greatest element I . These axioms bear a striking similarity with those of a linear space. One difference is that the vector/signal addition (+) of linear spaces is now replaced by two dual superpositions, the lattice supremum (\vee) and infimum (\wedge); further, the scalar multiplication (\times) of linear spaces is now replaced by two operations \star and \star' which are dual to each other. Only one major property of linear spaces is missing from the weighted lattices: the existence of ‘additive inverses’. We define the space \mathcal{W} to be a **complete weighted lattice (CWL)** if (i) \mathcal{W} is closed under any (possibly infinite) suprema and infima, and (ii) the distributivity laws between the scalar operations \star (\star') and the supremum (infimum) are of the infinite type. Note that a commutative clodum is a complete weighted lattice over itself.

Table 1 Axioms of Weighted Lattices [63]

Sup-Semilattice	Inf-Semilattice	Description
L1. $X \vee Y \in \mathcal{W}$	L1'. $X \wedge Y \in \mathcal{W}$	Closure of \vee, \wedge
L2. $X \vee X = X$	L2'. $X \wedge X = X$	Idempotence of \vee, \wedge
L3. $X \vee Y = Y \vee X$	L3'. $X \wedge Y = Y \wedge X$	Commutativity of \vee, \wedge
L4. $X \vee (Y \vee Z) = (X \vee Y) \vee Z$	L4'. $X \wedge (Y \wedge Z) = (X \wedge Y) \wedge Z$	Associativity of \vee, \wedge
L5. $X \vee (X \wedge Y) = X$	L5'. $X \wedge (X \vee Y) = X$	Absorption between \vee, \wedge
L6. $X \leq Y \iff Y = X \vee Y$	L6'. $X \leq' Y \iff Y = X \wedge Y$	Consistency of \vee, \wedge with partial order \leq
L7. $O \vee X = X$	L7'. $I \wedge X = X$	Identities of \vee, \wedge
L8. $I \vee X = I$	L8'. $O \wedge X = O$	Absorbing Nulls of \vee, \wedge
L9. $X \vee (Y \wedge Z) = (X \vee Y) \wedge (X \vee Z)$	L9'. $X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z)$	Distributivity of \vee, \wedge
WL10. $a \star X \in \mathcal{W}$	WL10'. $a \star' X \in \mathcal{W}$	Closure of \star, \star'
WL11. $a \star (b \star X) = (a \star b) \star X$	WL11'. $a \star' (b \star' X) = (a \star' b) \star' X$	Associativity of \star, \star'
WL12. $a \star (X \vee Y) = a \star X \vee a \star Y$	WL12'. $a \star' (X \wedge Y) = a \star' X \wedge a \star' Y$	Distributive scalar-vector mult over vector sup/inf
WL13. $(a \vee b) \star X = a \star X \vee b \star X$	WL13'. $(a \wedge b) \star' X = a \star' X \wedge b \star' X$	Distributive scalar-vector mult over scalar sup/inf
WL14. $e \star X = X$	WL14'. $e' \star' X = X$	Scalar Identities
WL15. $\perp \star X = O$	WL15'. $\top \star' X = I$	Scalar Nulls
WL16. $a \star O = O$	WL16'. $a \star' I = I$	Vector Nulls

5.3 Vector and Signal Operators on Weighted Lattices

We focus on CWLs whose underlying set is a *space* $\mathcal{W} = \text{Fun}(E, \mathcal{K})$ of *functions* $f : E \rightarrow \mathcal{K}$ with values from a clodum $(\mathcal{K}, \vee, \wedge, \star, \star')$ of scalars as in Examples 2(a),(b),(c). Such functions include n -dimensional vectors if $E = \{1, 2, \dots, n\}$ or d -dimensional signals of continuous ($E = \mathbb{R}^d$) or discrete domain ($E = \mathbb{Z}^d$). Then, we extend *pointwise* the supremum, infimum, and scalar multiplications of \mathcal{K} to functions: e.g., for $F, G \in \mathcal{W}$, $a \in \mathcal{K}$ and $x \in E$, we define $(F \vee G)(x) := F(x) \vee G(x)$ and $(a \star F)(x) := a \star F(x)$. Further, the scalar operations \star and \star' , extended pointwise to functions, distribute over any suprema and infima, respectively. If the clodum \mathcal{K} is self-conjugate, then we can extend the conjugation $(\cdot)^\top$ to functions F pointwise: $F^\top(x) \triangleq (F(x))^\top$.

Elementary increasing operators on \mathcal{W} are those that act as **vertical translations** (in short V-translations) of functions. Specifically, pointwise ‘multiplications’ of functions $F \in \mathcal{W}$ by scalars $a \in \mathcal{K}$ yield the *V-translations* τ_a and *dual V-translations* τ'_a , defined by $[\tau_a(F)](x) := a \star F(x)$ and $[\tau'_a(F)](x) := a \star' F(x)$. A function operator ψ on \mathcal{W} is called **V-translation invariant** if it commutes with any V-translation τ , i.e., $\psi\tau = \tau\psi$. Similarly for dual translations.

Every function $F(x)$ admits a representation as a supremum of V-translated impulses placed at all points or as infimum of dual V-translated impulses:

$$F(x) = \bigvee_{y \in E} F(y) \star q_y(x) = \bigwedge_{y \in E} F(y) \star' q'_y(x) \quad (31)$$

where $q_y(x) = e$ at $x = y$ and \perp else, whereas $q'_y(x) = e'$ at $x = y$ and \top else. By using the V-translations and the representation of functions with impulses, we can build more complex increasing operators. We define operators δ as **dilation V-translation invariant (DVI)** and operators \mathcal{E} as **erosion V-translation invariant (EVI)** iff for any $c_i \in \mathcal{K}$, $F_i \in \mathcal{W}$

$$\text{DVI: } \delta\left(\bigvee_i c_i \star F_i\right) = \bigvee_i c_i \star \delta(F_i), \quad \text{EVI: } \mathcal{E}\left(\bigwedge_i c_i \star' F_i\right) = \bigwedge_i c_i \star' \mathcal{E}(F_i) \quad (32)$$

The structure of a DVI or EVI operator’s output is simplified if we express it via the operator’s impulse responses. Given a dilation δ on \mathcal{W} , its **impulse response map** is the map $H : E \rightarrow \text{Fun}(E, \mathcal{K})$ defined at each $y \in E$ as the output function $H(x, y)$ from δ when the input is the impulse $q_y(x)$. Dually, for an erosion operator \mathcal{E} we define its *dual impulse response map* H' via its outputs when excited by dual impulses: for $x, y \in E$

$$H(x, y) \triangleq \delta(q_y)(x), \quad H'(x, y) \triangleq \mathcal{E}(q'_y)(x) \quad (33)$$

Applying a DVI operator δ or an EVI operator \mathcal{E} to (31) and using the definitions in (33) yields the following unified representation, which is proven in [4, 58] for the max-plus case and in [60] for the more general max- \star and max- \star' cases.

Theorem 1 (a) An operator δ on \mathcal{W} is DVI iff its output can be expressed as

$$\delta(F)(x) = \bigvee_{y \in E} H(x, y) \star F(y) \quad (34)$$

(b) An operator \mathcal{E} on \mathcal{W} is EVI iff its output can be expressed as

$$\mathcal{E}(F)(x) = \bigwedge_{y \in E} H'(x, y) \star' F(y) \quad (35)$$

On signal spaces, the operations (34) and (35) are *shift-varying nonlinear convolutions*.

Weighted Lattice of Vectors:

Consider now the nonlinear vector space $\mathcal{W} = \mathcal{K}^n$, equipped with the pointwise partial ordering $\mathbf{x} \leq \mathbf{y}$, supremum $\mathbf{x} \vee \mathbf{y} = [x_i \vee y_i]$, and infimum $\mathbf{x} \wedge \mathbf{y} = [x_i \wedge y_i]$ between any vectors $\mathbf{x}, \mathbf{y} \in \mathcal{W}$. Then, $(\mathcal{W}, \vee, \wedge, \star, \star')$ is a complete weighted lattice. Elementary increasing operators are the *vector V-translations* $\tau_a(\mathbf{x}) = a \star \mathbf{x} = [a \star x_i]$ and their duals $\tau'_a(\mathbf{x}) = a \star' \mathbf{x}$, which ‘multiply’ a scalar a with a vector \mathbf{x} elementwise. A vector transformation on \mathcal{W} is called (dual) V-translation invariant if it commutes with any vector (dual) V-translation. By defining as ‘impulses’ the impulse vectors $\mathbf{q}_j = [q_j(i)]$ and their duals $\mathbf{q}'_j = [q'_j(i)]$, where the index j signifies the position of the identity, each vector $\mathbf{x} = [x_1, \dots, x_n]^T$ has a representation as a max of V-translated impulse vectors or as a min of V-translated dual impulse vectors. More complex examples of increasing operators on such vector spaces are the max- \star and the min- \star' ‘multiplications’ of a matrix \mathbf{A} with an input vector \mathbf{x} ,

$$\delta_{\mathbf{A}}(\mathbf{x}) \triangleq \mathbf{A} \boxtimes \mathbf{x}, \quad \mathcal{E}_{\mathbf{A}}(\mathbf{x}) \triangleq \mathbf{A} \boxtimes' \mathbf{x} \quad (36)$$

which are the prototypes of any vector transformation that obeys a sup- \star or an inf- \star' superposition.

Theorem 2 [62, 63] (a) Any vector transformation on the complete weighted lattice $\mathcal{W} = \mathcal{K}^n$ is DVI iff it can be represented as a matrix-vector max- \star product $\delta_{\mathbf{A}}(\mathbf{x}) = \mathbf{A} \boxtimes \mathbf{x}$ where $\mathbf{A} = [a_{ij}]$ with $a_{ij} = [\delta(\mathbf{q}_j)]_i$, $i, j = 1, \dots, n$.
 (b) Any vector transformation on \mathcal{K}^n is EVI iff it can be represented as a matrix-vector min- \star' product $\mathcal{E}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A} \boxtimes' \mathbf{x}$ where $\mathbf{A} = [a_{ij}]$ with $a_{ij} = [\mathcal{E}(\mathbf{q}'_j)]_i$.

Note that the above theorem also holds for vector transformations between CWLs of different dimensionality, say from \mathcal{K}^n to \mathcal{K}^m , in which case the corresponding matrix $\mathbf{A} \in \mathcal{K}^{m \times n}$ is rectangular. Given such a vector dilation $\delta(\mathbf{x}) = \mathbf{A} \boxtimes \mathbf{x} : \mathcal{K}^n \rightarrow \mathcal{K}^m$, there corresponds a unique erosion $\mathcal{E} : \mathcal{K}^m \rightarrow \mathcal{K}^n$ (equal to the residual operator δ^\sharp) so that (δ, \mathcal{E}) is a *vector adjunction*, i.e. $\delta(\mathbf{x}) \leq \mathbf{y} \iff \mathbf{x} \leq \mathcal{E}(\mathbf{y})$. We can find the adjoint vector erosion by decomposing both vector operators based on *scalar operators* (η, ζ) that form a *scalar adjunction* on \mathcal{K} :

$$\eta(a, v) \leq w \iff v \leq \zeta(a, w) \quad (37)$$

If we use as scalar ‘multiplication’ a commutative binary operation $\eta(a, v) = a \star v$ that is a dilation on \mathcal{K} , its scalar adjoint erosion becomes

$$\zeta(a, w) = \sup\{v \in \mathcal{K} : a \star v \leq w\} \quad (38)$$

which is a (possibly non-commutative) binary operation on \mathcal{K} . Then, the original vector dilation $\delta(\mathbf{x}) = \mathbf{A} \boxtimes \mathbf{x}$ is decomposed as

$$[\delta(\mathbf{x})]_i = \bigvee_{j=1}^n \eta(a_{ij}, x_j) = \bigvee_{j=1}^n a_{ij} \star x_j, \quad i = 1, \dots, m \quad (39)$$

whereas its adjoint vector erosion (i.e. the residual δ^\sharp of δ) is decomposed as

$$[\delta^\sharp(\mathbf{y})]_j = [\mathcal{E}(\mathbf{y})]_j = \bigwedge_{i=1}^m \zeta(a_{ij}, y_i), \quad j = 1, \dots, n \quad (40)$$

The latter can be written as a min- ζ matrix-vector multiplication

$$\mathcal{E}(\mathbf{y}) = \mathbf{A}^T \square'_\zeta \mathbf{y} \quad (41)$$

where the symbol \square'_ζ denotes the following nonlinear product of a matrix $\mathbf{A} = [a_{ij}]$ with a matrix $\mathbf{B} = [b_{ij}]$:

$$[\mathbf{A} \square'_\zeta \mathbf{B}]_{ij} \triangleq \bigwedge_k \zeta(a_{ik}, b_{kj})$$

Further, if $\mathcal{K} = (\vee, \wedge, \star, \star')$ is a *clog*, then $\zeta(a, w) = a^\neg \star' w$ and hence

$$\mathcal{E}(\mathbf{y}) = \mathbf{A}^* \boxtimes' \mathbf{y}, \quad [\mathcal{E}(\mathbf{y})]_j = \bigwedge_{i=1}^m a_{ij}^\neg \star' y_i, \quad j = 1, \dots, n \quad (42)$$

where $\mathbf{A}^* = [a_{ji}^\neg]$ is the *adjoint* (i.e. conjugate transpose) of $\mathbf{A} = [a_{ij}]$.

Weighted Lattice of Signals:

Consider the set $\mathcal{W} = \text{Fun}(E, \mathcal{K})$ of all signals $f : E \rightarrow \mathcal{K}$ with domain $E = \mathbb{R}^d$ or \mathbb{Z}^d and values from \mathcal{K} . The signal translations are the operators $\tau_{y,v}(f)(x) = f(x - y) \star v$ and their duals. A signal operator on \mathcal{W} is called *(dual) translation invariant* iff it commutes with any such (dual) translation. This translation-invariance contains both a vertical translation and a horizontal translation (shift). Consider now operators Δ on \mathcal{W} that are dilations and translation-invariant. Then Δ is both DVI in the sense of (32) and shift-invariant. We call such operators **dilation translation-invariant (DTI)** systems. Applying Δ to an input signal f decomposed as supremum of translated impulses yields its output as the sup- \star convolution \otimes of the input with the system’s impulse response $h = \Delta(q)$, where $q(x) = e$ if $x = 0$ and \perp else:

$$\Delta(f)(x) = (f \circledast h)(x) = \bigvee_{y \in E} f(y) \star h(x-y) \quad (43)$$

Conversely, every sup- \star convolution is a DTI system. As done for the vector operators, we can also build signal operator pairs (Δ, \mathcal{E}) that form adjunctions. Given Δ we can find its adjoint \mathcal{E} from scalar adjunctions (η, ζ) . Thus, by (37) and (38), if $\eta(h, f) = h \star f$, the adjoint signal erosion becomes

$$\mathcal{E}(g)(y) = \bigwedge_{x \in E} \zeta[h(x-y), g(x)] \quad (44)$$

Further, if \mathcal{K} is a clog, then

$$\mathcal{E}(g)(y) = \bigwedge_{x \in E} g(x) \star' h^\neg(x-y) \quad (45)$$

5.4 CWL Generalizations of Tropical Lines and Planes

In the same way that weighted lattices generalize max-plus morphology and extend it to other types of clodum arithmetic, we can extend the basic objects of max-plus tropical geometry (i.e. tropical lines and planes) to other max- \star geometric objects. For example, over a clodum $(\mathcal{K}, \vee, \wedge, \star, \star')$, we can generalize max-plus tropical lines $y = \max(a+x, b)$ as $y = \max(a \star x, b)$ and similarly tropical planes as $z = \max(a \star x, b \star y, c)$. Figure 8 shows some generalized tropical lines where the \star operation is sum, product, and min.

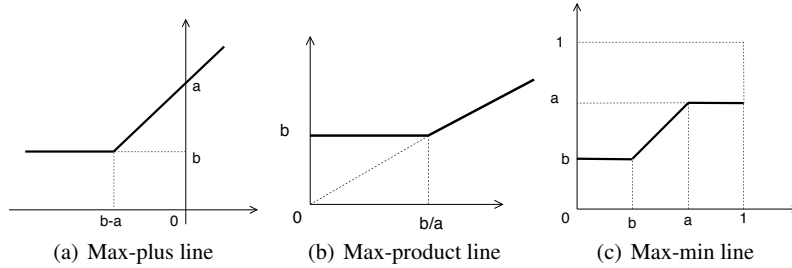


Fig. 8 Max- \star tropical lines $y = \max(a \star x, b)$: (a) $y = \max(a+x, b)$, (b) $y = \max(a \cdot x, b)$, (c) $y = \max(a \wedge x, b)$.

Further, we can generalize max-plus halfspaces (28) to max- \star tropical halfspaces:

$$\mathcal{T}(\mathbf{a}, \mathbf{b}) \triangleq \{\mathbf{x} \in \mathcal{K}^n : \mathbf{a}^T \boxtimes \begin{bmatrix} \mathbf{x} \\ e \end{bmatrix} \leq \mathbf{b}^T \boxtimes \begin{bmatrix} \mathbf{x} \\ e \end{bmatrix}\} \quad (46)$$

Examples of max-plus tropical halfspaces are shown in Fig. 6 and Fig. 7. The slopes of their bounding line segments or faces are either zero or equal to 1. Max-product halfspaces can give boundaries that are piecewise-linear but have arbitrary slopes. Max-min halfspaces have piecewise-linear boundaries with more corner points or edges; see example in Fig. 8(c).

Finally, a totally different generalization results if we replace the ‘multiplication’ \star in a generalized tropical line with the (log-sum-exp) softmin operation of Example 2(d), in which case the line segments of a tropical line will become smooth exponential curves.

6 Solving Max- \star Equations and Optimization

6.1 ℓ_p Optimal Subsolutions of Max- \star Equations

Consider a scalar clodum $(\mathcal{K}, \vee, \wedge, \star, \star')$, a matrix $\mathbf{A} \in \mathcal{K}^{m \times n}$ and a vector $\mathbf{b} \in \mathcal{K}^m$. The set of solutions of the max- \star equation

$$\mathbf{A} \boxtimes \mathbf{x} = \mathbf{b} \quad (47)$$

over \mathcal{K} is either empty or forms an idempotent semigroup under vector \vee , because if $\mathbf{x}_1, \mathbf{x}_2$ are two solutions then $\mathbf{x}_1 \vee \mathbf{x}_2$ is also a solution. A related problem in applications of max-plus algebra to scheduling is when a vector \mathbf{x} represents start times, a vector \mathbf{b} represents finish times, and the matrix \mathbf{A} represents processing delays. Then, if $\mathbf{A} \boxtimes \mathbf{x} = \mathbf{b}$ does not have an exact solution, it is possible to find the optimum \mathbf{x} such that we minimize a norm of the earliness subject to zero lateness. We generalize this problem from max-plus to max- \star algebra. The optimum will be the solution of the following constrained minimization problem:

$$\text{Minimize } \|\mathbf{A} \boxtimes \mathbf{x} - \mathbf{b}\|_p \text{ s.t. } \mathbf{A} \boxtimes \mathbf{x} \leq \mathbf{b} \quad (48)$$

where the norm $\|\cdot\|_p$ is any ℓ_p norm with $p = 1, 2, \dots, \infty$. While the two above problems have been solved in [23] for the max-plus case and for $p = 1$ or $p = \infty$, we provide next a more general result using adjunctions for the general case when \mathcal{K} is just a clodum or a general clog and $\|\cdot\|_p$ is any Minkowski norm.

Theorem 3 ([63]) *Consider a vector dilation $\delta(\mathbf{x}) = \mathbf{A} \boxtimes \mathbf{x} : \mathcal{K}^n \rightarrow \mathcal{K}^m$ over a clodum \mathcal{K} and let \mathcal{E} be its adjoint vector erosion. (a) If Eq. (47) has a solution, then*

$$\hat{\mathbf{x}} = \mathcal{E}(\mathbf{b}) = \mathbf{A}^T \square'_\zeta \mathbf{b} = \left[\bigwedge_{i=1}^m \zeta(a_{ij}, b_i) \right] \quad (49)$$

is its greatest solution, where ζ is the scalar adjoint erosion of \star as in (38). (b) If \mathcal{K} is a clog, the solution (49) becomes

$$\hat{\mathbf{x}} = \mathbf{A}^* \boxtimes' \mathbf{b} = \left[\bigwedge_{i=1}^m a_{ij} \star' b_i \right] \quad (50)$$

(c) The solution to the optimization problem (48) for any ℓ_p norm $\|\cdot\|_p$ is generally (49), or (50) in the case of a clog.

A main idea for solving (48) is to consider vectors \mathbf{x} that are *subsolutions* in the sense that $\delta(\mathbf{x}) = \mathbf{A} \boxtimes \mathbf{x} \leq \mathbf{b}$ and find the greatest such subsolution $\hat{\mathbf{x}} = \mathcal{E}(\mathbf{b})$, which yields either the greatest exact solution of (47) or an optimum subsolution in the sense of (48). This creates a lattice projection onto the $\max\text{-}\star$ span of the columns of \mathbf{A} via the opening $\delta(\mathcal{E}(\mathbf{b})) \leq \mathbf{b}$ that best approximates \mathbf{b} from below. Also, note that since $\mathbf{y} = \delta(\mathcal{E}(\mathbf{b})) = [y_i]$ is the greatest lower estimate of $\mathbf{b} = [b_i]$, $b_i - y_i$ is nonnegative and minimum for all i , and hence the norm $\|\mathbf{b} - \mathbf{y}\|_p$ is minimum for any $p = 1, 2, \dots, \infty$.

As a final note, in the $\max\text{-plus}$ case it is also possible to search and find *sparse solutions* of either the exact equation (47) or the approximate problem (48), as done in [84], where sparsity here means a large number of $-\infty$ values in the solution vector.

6.2 Projections on Weighted Lattices

The optimal subsolution of (48) can be viewed in the $\max\text{-plus}$ case as a nonlinear ‘projection’ of \mathbf{b} onto the column-space of \mathbf{A} [24]. To understand this, note first that any adjunction (δ, \mathcal{E}) automatically yields two lattice projections, an opening $\alpha = \delta\mathcal{E}$ and a closing $\beta = \mathcal{E}\delta$, such that

$$\alpha^2 = \alpha \leq \text{id} \leq \beta = \beta^2$$

where the composition of two operators is written as an operator product. We call them ‘projections’ because, in analogy to projection operators on linear spaces, they preserve the structure of the lattice space w.r.t. the partial ordering (due to their isotonicity) and they are idempotent.

Projections on idempotent semimodules⁴ have been studied in [22] for the general case and in more detail for the $\max\text{-plus}$ case in [3]. Let \mathcal{X} be a complete idempotent semimodule, and let \mathcal{S} be a subsemimodule of \mathcal{X} . Then a *canonical*

⁴ Idempotent semimodules are like vector spaces with idempotent vector ‘addition’ \vee whose vector and scalar arithmetic are defined over idempotent semirings. If in our definition of a weighted lattice, one focuses only on one vector ‘addition’, say the vector supremum, and its corresponding scalar ‘multiplication’, then the weaker algebraic structure becomes an idempotent semimodule over an idempotent semiring $(\mathcal{X}, \vee, \star)$. This has been studied in [22, 32, 54] where often closure under infinite suprema is assumed; in such cases, an ‘infimum’ operation can be also indirectly defined (since a complete sup-semilattice with a least element is a complete lattice), which makes the space a complete lattice, but this indirect infimum may be different than the direct (conventional) infimum of the original lattice (if it exists).

projector on \mathcal{S} is defined as the nonlinear map [22]

$$P_{\mathcal{S}} : \mathcal{X} \rightarrow \mathcal{X}, \quad P_{\mathcal{S}}(x) \triangleq \bigvee \{v \in \mathcal{S} : v \leq x\} \quad (51)$$

Its definition implies that $P_{\mathcal{S}}$ is a lattice opening, i.e. increasing, antiextensive, and idempotent. Further, there is a concept of ‘distance’ on such semimodules which allows to use a nonlinear projection theorem for best approximations. We shall outline these ideas only for the max-plus case, i.e. for $\mathcal{X} = \mathbb{R}^n$ viewed as complete semimodule over the complete max-plus semiring $\mathbb{R}_{\max} \cup \{\infty\}$. Specifically, let us consider the *Hilbert projective metric*

$$d_H(\mathbf{x}, \mathbf{y}) \triangleq -[(\mathbf{x} \setminus \mathbf{y}) + (\mathbf{y} \setminus \mathbf{x})], \quad \mathbf{x} \setminus \mathbf{y} := \max\{a \in \overline{\mathbb{R}} : \mathbf{x} + a \leq \mathbf{y}\} \quad (52)$$

between any vectors $\mathbf{x}, \mathbf{y} \in \overline{\mathbb{R}}^n$. Note that this is only a semimetric and for finite-valued vectors it assumes the simpler expression (called *range semimetric* in [23])

$$d_H(\mathbf{x}, \mathbf{y}) = \max_i (x_i - y_i) - \min_i (x_i - y_i), \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad (53)$$

Then, given a subsemimodule \mathcal{S} of $\overline{\mathbb{R}}^n$, it follows that for any vector $\mathbf{x} \in \overline{\mathbb{R}}^n$, $P_{\mathcal{S}}(\mathbf{x})$ is the best approximation (but not necessarily unique) of \mathbf{x} by elements of \mathcal{S} . Specifically [22, 3], the projection $P_{\mathcal{S}}(\mathbf{x})$ of \mathbf{x} onto \mathcal{S} is that element of \mathcal{S} within the shortest distance from \mathbf{x} than any other element of \mathcal{S} ; i.e.,

$$d_H(\mathbf{x}, P_{\mathcal{S}}(\mathbf{x})) = d_H(\mathbf{x}, \mathcal{S}) \quad (54)$$

where the distance between a vector \mathbf{x} and the subspace \mathcal{S} is defined by $d_H(\mathbf{x}, \mathcal{S}) := \inf\{d_H(\mathbf{x}, \mathbf{v}) : \mathbf{v} \in \mathcal{S}\}$. Note the analogy with Euclidean spaces \mathbb{R}^n where the linear projection of a point $\mathbf{x} \in \mathbb{R}^n$ to a linear subspace \mathcal{S} is given by the unique point $\mathbf{y} \in \mathcal{S}$ such that $\mathbf{x} - \mathbf{y}$ is orthogonal to \mathcal{S} .

Now, if we consider the optimization problem (48) and define the subsemimodule \mathcal{S} in (51) as the max-plus span of the columns of matrix \mathbf{A} , then the canonical projection of \mathbf{b} onto it equals

$$P_{\mathcal{S}}(\mathbf{b}) = \mathbf{A} \boxplus \hat{\mathbf{x}} = \mathbf{A} \boxplus \mathbf{A}^* \boxplus' \mathbf{b} \leq \mathbf{b} \quad (55)$$

which is a lattice opening $\delta(\mathcal{E}(\mathbf{b})) \leq \mathbf{b}$.

6.3 ℓ_{∞} Optimal Solution of Max-plus Equations

The solution (50) is the greatest subsolution of problem (48). Thus, in the max-plus case (see (30) for the definitions of max-plus and min-plus matrix products), $\hat{\mathbf{x}} = \mathbf{A}^* \boxplus' \mathbf{b}$ is the optimal solution of

$$\text{Minimize } \|\mathbf{A} \boxplus \mathbf{x} - \mathbf{b}\|_{\infty} \quad (56)$$

under the constraint $\mathbf{x} \leq \mathbf{b}$. The proof results since $\hat{\mathbf{x}}$ is the greatest solution of $\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}$, as shown by Cuninghame-Green [23]. It can also be directly seen from the adjunction

$$\mathbf{A} \boxplus \mathbf{x} = \delta_{\mathbf{A}}(\mathbf{x}) \leq \mathbf{b} \iff \mathbf{x} \leq \varepsilon_{\mathbf{A}}(\mathbf{b}) = \mathbf{A}^* \boxplus' \mathbf{b} \quad (57)$$

The following is actually a stronger result that is not biased to be a subsolution but provides the *unconstrained optimal solution* of (56).

Theorem 4 ([23]) *If $2\mu = \|\mathbf{A} \boxplus \hat{\mathbf{x}} - \mathbf{b}\|_{\infty} = \|\mathbf{A} \boxplus (\mathbf{A}^* \boxplus' \mathbf{b}) - \mathbf{b}\|_{\infty}$ is the ℓ_{∞} error corresponding to the greatest subsolution of $\mathbf{A} \boxplus \mathbf{x} = \mathbf{b}$, then*

$$\tilde{\mathbf{x}} = \mu + \mathbf{A}^* \boxplus' \mathbf{b} \quad (58)$$

is the unique optimum solution of (56).

The computational complexity to find both optimal solutions $\hat{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ is $O(mn)$ (additions in the max-plus case), where m is the number of data and n their dimensionality.

Unfortunately, the ℓ_{∞} optimality of $\tilde{\mathbf{x}}$ does not carry over in the case of a general clodum, as shown for the max-min clodum in [25].

7 Optimal Fitting Tropical Polynomials to Data and Shape Approximation

Herein we apply tropical geometry and max- \star algebra to a fundamental regression problem of approximating the shape of curves and surfaces by fitting tropical polynomials to data, sampled from their functional form possibly in the presence of noise.

7.1 Piecewise-Linear Function Representation and Data Fitting

Piecewise-Linear (PWL) functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ are defined as follows: (i) Their domain is divided into a finite number of polyhedral regions separated by linear $(n-1)$ -dimensional boundaries that are hyperplanes or subsets of hyperplanes; (ii) They are affine over each region and continuous on each boundary. Approximations with PWL functions have proven analytically and computationally very useful in many fields of science and engineering, including splines [26], nonlinear circuits and systems modeling [20], machine learning [8, 82], convex optimization [12], geometric programming [11, 50, 57], statistics [35], and recently tropical geometry [56, 86]. A conventional representation of PWL functions requires simplicial subdivision of their domain and interpolation of the PWL function on the subdivided domain; this is local, without a closed-formula, and requires many parameters for storage and processing. Thus, two major problems are *representation*, i.e. finding

a better class of functions with analytical expressions to represent them, and their *parameter estimation* for modeling a nonlinear system or fitting some data. Further, while these problems are well-explored in the 1D case, they remain relatively underdeveloped for multi-dimensional data.

Chua and his collaborators [49, 19, 48] have introduced the so-called *canonical representation* for continuous PWL functions, consisting of an affine function plus a weighted sum of absolute-value affine functions (defining linear partitions) and extensively studied its application for nonlinear circuit analysis and modeling. This has the advantages over the conventional representation that it is global, explicit, analytic, compact (smaller number of model functions and corresponding parameters), and computationally efficient (easy to store and program). However, it is complete only for 1D PWL functions. In higher dimensions it needs multi-level nestings of the absolute-value functions; the depth of this nesting depends on the geometry of the partitions of the domain and the order of intersections of the partition boundaries [47, 34, 48, 52, 46].

Tarela et al [81], by combining their previous work [80] on representing continuous PWL functions with lattice generalizations of Boolean polynomials of lines or hyperplanes, which extended similar work by [88], with the general $f - \phi$ model for PWL functions of [51], developed a constructive way to generate min-max (and their dual max-min) combinations of affine functions which provide a complete representation of continuous PWL functions in arbitrary dimensions. This is called the *lattice representation*. Another work for max-min representation of PWL functions is [71]. Wang [87] completed the construction of a canonical representation for arbitrary continuous PWL functions in n -dimensions by starting from the lattice presentation of [81], which is a min-max of affine functions, producing an equivalent representation as a difference of two convex functions, each being max-affine, and then converting each max-affine function to a canonical representation that involved n -level nestings of absolute-value functions.

A more recent approach is to focus on the class of *convex* PWL functions represented by a maximum of affine functions (i.e. hyperplanes), that are essentially max-plus topical polynomials as in (22), and use them for data fitting; we shall call this class *max-affine* functions. Starting from early least-squares solutions [41, 43], some representative recent approaches to solve this *convex regression* problem include [35, 36, 42, 50, 57]. In all these approaches, there is an iteration that alternates between partitioning the data domain and locally fitting affine functions (using least-squares or some linear optimization procedure) to update the local coefficients. For a known partition the convex PWL function is formed as the max of the local affine fits. Then, a PWL function generates a new partition which can be used to refit the affine functions and improve the estimate. As explained in [57], this iteration can be viewed as a Gauss-Newton algorithm to solve the above nonlinear least-squares problem, similar to the K -means algorithm. The order K of the model can be increased until some error threshold is reached. Interesting and promising generalizations of the above max-affine representation for convex functions include works that use softmax instead of max, via the *log-sum-exp* models for convex and log-log convex data [42, 16, 15]. Other iterative approaches for convex PWL data fitting

include [83]. For additional references, we refer the reader to the bibliography in the above works.

Next, we focus on convex PWL regression via the max-affine model, which has a tropical interpretation, and propose a direct *non-iterative* and *low-complexity* approach to estimate its parameters by using the optimal solutions of max-plus (or max- \star) equations of Sec. 6. We note that the max-affine representation is not limited to PWL functions only, because we can represent any convex function as a supremum of a (possibly infinite) number of affine functions via the Fenchel-Legendre transform [28, 75, 55]. Closely related ideas are based on morphological slope transforms [27, 39, 58, 59] that offer generalizations of this result to non-convex functions and approximate representations via adjunctions.

7.2 Optimal Fitting Tropical Lines and Planes

We first examine a classic problem in machine learning, fitting a line to data by minimizing an error norm, in the light of tropical geometry. Given data $(x_i, f_i) \in \mathbb{R}^2$, $i = 1, \dots, m$, if we wish to fit a Euclidean line $y = ax + b$ by minimizing the ℓ_2 error norm $\|\mathbf{f} - a\mathbf{x} - b\|_2$ where $\mathbf{f} = [f_i]$ and $\mathbf{x} = [x_i]$, the optimal solution (*least squares estimate* - LSE) for the parameters a, b is

$$\hat{a}_{\text{LS}} = \frac{m \sum_i x_i f_i - (\sum_i x_i)(\sum_i f_i)}{m \sum_i (x_i)^2 - (\sum_i x_i)^2}, \quad \hat{b}_{\text{LS}} = \frac{1}{m} \sum_i (f_i - \hat{a}_{\text{LS}} x_i) \quad (59)$$

Suppose now we wish to fit a general tropical line $p(x) = \max(a \star x, b)$ by minimizing some ℓ_p error norm. The equations to solve for finding the optimal parameter vector $\mathbf{w} = [a, b]^T$ become:

$$\underbrace{\begin{bmatrix} x_1 & e \\ \vdots & \vdots \\ x_m & e \end{bmatrix}}_{\mathbf{X}} \boxtimes \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \quad (60)$$

By Theorem 3, the optimal (min ℓ_p error) subsolution for any clodum arithmetic is

$$\hat{\mathbf{w}} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \mathbf{X}^T \square'_\zeta \mathbf{f} = \begin{bmatrix} \bigwedge_i \zeta(x_i, f_i) \\ \bigwedge_i \zeta(e, f_i) \end{bmatrix} \quad (61)$$

where ζ is the scalar adjoint erosion (38) of \star . This vector $\hat{\mathbf{w}}$ yields (after min- ζ ‘multiplication’ with \mathbf{X}^T) the *greatest lower estimate* (GLE) of the data \mathbf{f} . If \mathcal{K} is a clog, like in the max-plus and max-product case, then $\zeta(x_i, f_i) = x_i^- \star' f_i$. Next we write in detail the solution for the tropical line for the three special cases where the

scalar arithmetic is based either on the max-plus clog⁵, or the max-product clog, or the max-min clodum (the shapes of the corresponding lines are shown in Fig.8):

$$(\hat{a}, \hat{b}) = \begin{cases} \bigwedge_i f_i - x_i, \bigwedge_i f_i, & \text{max-plus } (\star = +) \\ \bigwedge_i f_i / x_i, \bigwedge_i f_i, & \text{max-times } (\star = \times) \\ \bigwedge_i \max([f_i \geq x_i], f_i), \bigwedge_i f_i, & \text{max-min } (\star = \wedge) \end{cases} \quad (62)$$

where $[\cdot]$ denotes Iverson's bracket in the max-min case. Thus, the above approach allows to optimally fit (w.r.t. any ℓ_p error norm) general tropical lines to arbitrary data from below. In addition, for the *max-plus* case we can obtain the best (unconstrained) approximation with a tropical line that yields the smallest ℓ_∞ error. This *minimum max absolute error (MMAE)* solution is, by Theorem 4,

$$\tilde{\mathbf{w}} = \hat{\mathbf{w}} + \mu, \quad \mu = \frac{1}{2} \|\mathbf{X} \boxplus \hat{\mathbf{w}} - \mathbf{f}\|_\infty = \frac{1}{2} \|\mathbf{X} \boxplus (\mathbf{X}^* \boxminus \mathbf{f}) - \mathbf{f}\|_\infty \quad (63)$$

Example 3 Suppose we have $m = 200$ data observations (x_i, f_i) from the tropical line $y = \max(x - 2, 3)$, where the 200 abscissae x_i were uniformly spaced in $[-1, 12]$ and their corresponding values $f_i = y_i + \varepsilon_i$ are contaminated with two different types of zero-mean noise i.i.d. random variables ε_i , Gaussian noise $\sim \mathcal{N}(0, 0.25)$ and uniform noise $\sim \text{Unif}[-0.5, 0.5]$. Figure 9 shows the two optimal solutions (62) and (63) for fitting a max-plus tropical line, superimposed with the least-squares Euclidean line fit. The parameter estimates and errors are in Table 2.

Line fit Method	$\ \text{error}\ _{\text{RMS}}$	$\ \text{error}\ _\infty$	\hat{a}	\hat{b}
Tropical GLE	0.598	0.988	-2.492	2.509
Tropical MMAE	0.288	0.494	-1.998	3.003
Euclidean LSE	0.968	2.135	0.560	1.849

Table 2 Errors and parameter estimates for optimal fitting of a max-plus tropical line $y = \max(x - 2, 3)$ both via a least-squares Euclidean line fit and via the tropical constrained (GLE) and unconstrained (MMAE) solutions, to data corrupted by uniform noise.

The above approach and tropical solution can also be extended to fitting planes. Specifically, we wish to fit a general max- \star tropical plane $p(x, y)$

$$p(x, y) = \max(a \star x, b \star y, c) \quad (64)$$

to given data $(x_i, y_i, f_i) \in \mathbb{R}^3$, $i = 1, \dots, m$, where $f_i = p(x_i, y_i) + \text{error}$, by minimizing some ℓ_p error norm. As in (60), the equations to solve for finding the optimal parameters $\mathbf{w} = [a, b, c]^T$ become:

⁵ To cover all cases of combining finite and infinite scalar numbers in the max-plus clog ($\mathbb{R}, \vee, \wedge, +, +'$), we should write the subtractions $f_i - x_i$ in (62) as $f_i +' (-x_i)$.

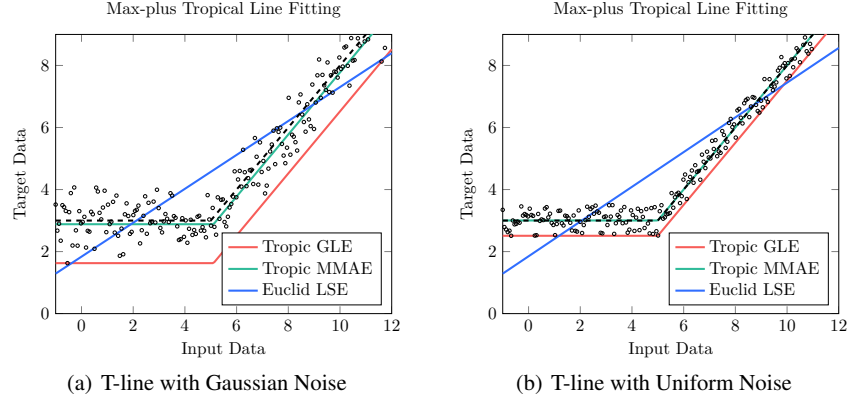


Fig. 9 (a) Optimal fitting via (62) or (63) of a max-plus tropical line $y = \max(x - 2, 3)$ (shown in black dashed curve) to data from the line corrupted by additive i.i.d. Gaussian noise $\sim \mathcal{N}(0, 0.25)$. Blue line: Euclidean line fitting via least squares. Red line: best subsolution (GLE). Green line: best unconstrained (MMAE) solution. (b) Same experiment as in (a) but with uniform noise $\sim \text{Unif}[-0.5, 0.5]$. Best viewed in color.

$$\underbrace{\begin{bmatrix} x_1 & y_1 & e \\ \vdots & \vdots & \vdots \\ x_m & y_m & e \end{bmatrix}}_{\mathbf{X}} \boxtimes \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \quad (65)$$

If we can accept subsolutions, which yield approximations of the given data from below, then by Theorem 3 the optimal subsolution for any clodum arithmetic is

$$\hat{\mathbf{w}} = \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = \mathbf{X}^T \square'_\zeta \mathbf{f} = \begin{bmatrix} \bigwedge_i \zeta(x_i, f_i) \\ \bigwedge_i \zeta(y_i, f_i) \\ \bigwedge_i \zeta(e, f_i) \end{bmatrix} \quad (66)$$

In the special case of *max-plus* arithmetic, then $\zeta(x_i, f_i) = f_i - x_i$ and the best subsolution (for $\min \ell_p$ error) becomes

$$\begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = \hat{\mathbf{w}} = \mathbf{X}^* \boxplus' \mathbf{f} = \begin{bmatrix} -x_1 & -x_2 & \cdots & -x_m \\ -y_1 & -y_2 & \cdots & -y_m \\ 0 & 0 & \cdots & 0 \end{bmatrix} \boxplus' \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} = \begin{bmatrix} \bigwedge_{i=1}^m f_i - x_i \\ \bigwedge_{i=1}^m f_i - y_i \\ \bigwedge_{i=1}^m f_i \end{bmatrix} \quad (67)$$

Furthermore, the minimum max absolute error (MMAE) solution is given by (63), but the data matrix \mathbf{X} and vector \mathbf{f} refer now to the plane case.

7.3 Shape Regression by Optimal Fitting Tropical Max-plus Polynomial Curves and Surfaces

For the *max-plus* case, the above approach and solution can also be generalized to polynomial curves of higher degree and to multi-dimensional data. We wish to fit a max-plus tropical polynomial

$$p(\mathbf{x}) = \max(\mathbf{a}_1^T \mathbf{x} + b_1, \mathbf{a}_2^T \mathbf{x} + b_2, \dots, \mathbf{a}_K^T \mathbf{x} + b_K) = \bigvee_{k=1}^K \mathbf{a}_k^T \mathbf{x} + b_k, \quad \mathbf{x} \in \mathbb{R}^n \quad (68)$$

to given data $(\mathbf{x}_i, f_i) \in \mathbb{R}^{n+1}$, $i = 1, \dots, m$, where $f_i = p(\mathbf{x}_i) + \text{error}$, by minimizing some ℓ_p error norm. The exact equations are

$$\underbrace{\begin{bmatrix} \mathbf{a}_1^T \mathbf{x}_1 & \mathbf{a}_2^T \mathbf{x}_1 & \cdots & \mathbf{a}_K^T \mathbf{x}_1 \\ \mathbf{a}_1^T \mathbf{x}_2 & \mathbf{a}_2^T \mathbf{x}_2 & \cdots & \mathbf{a}_K^T \mathbf{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_1^T \mathbf{x}_m & \mathbf{a}_2^T \mathbf{x}_m & \cdots & \mathbf{a}_K^T \mathbf{x}_m \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \quad (69)$$

We assume that the slope vectors \mathbf{a}_k are given and we optimize for the parameters $\{b_k\}$. By Theorem 3, the optimal subsolution for minimum ℓ_p error is

$$\begin{bmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_K \end{bmatrix} = \hat{\mathbf{w}} = \mathbf{X}^* \boxplus' \mathbf{f} = \begin{bmatrix} -\mathbf{a}_1^T \mathbf{x}_1 & -\mathbf{a}_1^T \mathbf{x}_2 & \cdots & -\mathbf{a}_1^T \mathbf{x}_m \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{a}_K^T \mathbf{x}_1 & -\mathbf{a}_K^T \mathbf{x}_2 & \cdots & -\mathbf{a}_K^T \mathbf{x}_m \end{bmatrix} \boxplus' \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} = \begin{bmatrix} \bigwedge_{i=1}^m f_i - \mathbf{a}_1^T \mathbf{x}_i \\ \vdots \\ \bigwedge_{i=1}^m f_i - \mathbf{a}_K^T \mathbf{x}_i \end{bmatrix} \quad (70)$$

Note that $\mathbf{X} \boxplus \hat{\mathbf{w}} \leq \mathbf{f}$. Further, by Theorem 4, the unconstrained solution that yields the minimum ℓ_∞ error is

$$\tilde{\mathbf{w}} = \mu + \hat{\mathbf{w}}, \quad \mu = \frac{1}{2} \|\mathbf{X} \boxplus \hat{\mathbf{w}} - \mathbf{f}\|_\infty \quad (71)$$

Our assumption for known slope vectors \mathbf{a}_k does not pose a significant constraint in many cases where the degree of the polynomial is relatively small, in which case we assume that \mathbf{a}_k are integer multiples of a slope step or simply that they assume all integer values up to the maximum degree. If this is not the case, another approach is to compute the derivatives (or gradients) of the given data, estimate the histogram of the derivative values, and use this for automatic selection of the slope parameters. Or simply to cluster the data gradients using K -means and use the centroids of the K clusters as our given slope vectors. In both approaches, setting $b_k = -\infty$ for some k , removes the corresponding line or hyperplane from the max-affine combination. Next we apply the above approaches for optimally solving two cases (1D and 2D) with examples.

7.3.1 Optimal Fitting 1D Tropical Max-plus Polynomial Curves

We wish to fit a max-plus tropical polynomial curve

$$p(x) = \max(b_{-r} - rx, \dots, b_{-1} - x, b_0, b_1 + x, \dots, b_r + rx) = \bigvee_{k=-r}^r b_k + kx \quad (72)$$

of relatively low degree r with $K = 2r + 1$ terms to given data $(x_i, f_i) \in \mathbb{R}^2$, $i = 1, \dots, m$, where $f_i = p(x_i) + \text{error}$, by minimizing some ℓ_p error norm. The tropical polynomial $p(x)$ is a maximum of straight lines with integer slopes $a_k = k \in \mathbb{R}$ and intercepts $b_k \in \mathbb{R}_{\max}$; a null intercept means that the corresponding line does not contribute to $p(x)$. The above PWL model is efficient if the function to approximate has both positive and nonnegative slopes and over the approximation interval its slopes do not exceed a relatively low polynomial degree r . If it has only nonnegative slopes, then we do not include the terms with negative slopes.

Based on the PWL model (72), the equations to solve for finding the optimal parameters $\{b_k\}$ become:

$$\underbrace{\begin{bmatrix} -rx_1 & (1-r)x_1 & \cdots & 0 & \cdots & (r-1)x_1 & rx_1 \\ -rx_2 & (1-r)x_2 & \cdots & 0 & \cdots & (r-1)x_2 & rx_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -rx_m & (1-r)x_m & \cdots & 0 & \cdots & (r-1)x_m & rx_m \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} b_{-r} \\ b_{1-r} \\ \vdots \\ b_r \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \quad (73)$$

Example 4 See Fig. 10 for a numerical example where the data to fit resulted from sampling the bottom half of the circular curve $x^2 + (y - 10)^2 = 7^2$ at $m = 5$ points with abscissae $(x_1, \dots, x_5) = (-5.5, -2, 1.5, 4, 6.5)$. The optimal fit was done using a tropical polynomial as in (72) with $r = 3$, i.e. a max of 7 lines, yielding a MMAE of $\|[p(x_i)] - [y_i]\|_\infty = 0.12$.

Example 5 As another 1D example, consider clean data (x_i, y_i) that are $m = 100$ points with abscissae x_i uniformly sampled within the interval $[-2, 2]$ and ordinates $y_i = f_i = f(x_i)$ where $f(x)$ is the convex function [42]

$$f(x) = \max(-6x - 6, \frac{x}{2}, \frac{x^5}{5} + \frac{x}{2}) \quad (74)$$

The tropical model we are fitting is of the form $p(x) = \max(a_1x + b_1, \dots, a_Kx + b_K)$ where the slopes a_k are computed using the Jenks natural breaks optimization (which is essentially a 1D K -means) algorithm applied to the numerical derivatives of the data and the intercepts b_k are computed using the 1D ($n = 1$) version of the tropical fitting algorithms (70) and (71). See Fig. 11 for the curve approximations and Table 3 for the corresponding errors.

For the case $K = 6$ the estimates for the slopes and the MMAE solution for intercepts yielded

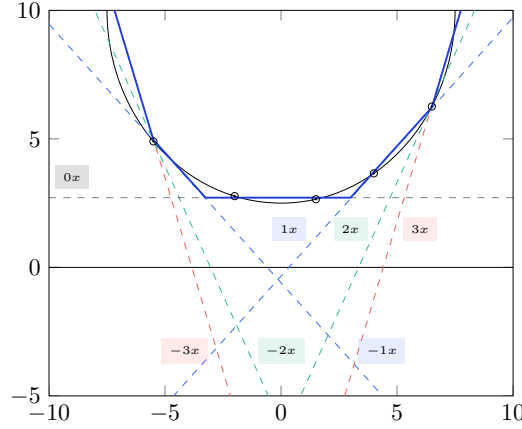


Fig. 10 Piecewise-linear curve approximation of a half circle (black solid line) by interpolating 5 samples with a tropical max-plus polynomial (blue solid line). The individual lines of the PWL function are shown with dashed lines.

K	GLE		MMAE	
	$\text{error}_{\text{RMS}}$	$\ \text{error}\ _{\infty}$	$\text{error}_{\text{RMS}}$	$\ \text{error}\ _{\infty}$
3	0.4101	0.9671	0.3535	0.4836
4	0.2048	0.5072	0.1799	0.2536
5	0.1230	0.7226	0.3004	0.3613
6	0.0801	0.1932	0.0625	0.0966

Table 3 Minimum RMS error and maximum absolute error for the optimal constrained (GLE) and unconstrained (MMAE) tropical fitting of the function (74).

$$\begin{aligned}
 (a_1, a_2, a_3, a_4, a_5, a_6) &= (-5.92, 0.64, 3.07, 6.43, 10.08, 14.11) \\
 (b_1, b_2, b_3, b_4, b_5, b_6) &= (-5.82, 0.03, -2.5, -7.3, -13.4, -20.83)
 \end{aligned} \tag{75}$$

For the case $K = 3$, although our method yields about double the RMS error of the method in [42], the latter is computationally more complex as explained later.

7.3.2 Optimal Fitting 2D Tropical Max-plus Polynomial Surfaces

As a 2D example with known slopes, let us fit the graph surface of a symmetric max-plus tropical conic polynomial

$$p(x, y) = \bigvee_{0 \leq |k+\ell| \leq 2, k\ell \geq 0} b_{k\ell} + kx + \ell y \tag{76}$$

to given data $(x_i, y_i, f_i) \in \mathbb{R}^3$, $i = 1, \dots, m$, where $f_i = p(x_i, y_i) + \text{error}$ by minimizing some ℓ_p error norm. The equations to solve for finding the optimal parameters $b_{k\ell}$, adjusted as in (72) so that they include both negative and positive slopes, become:

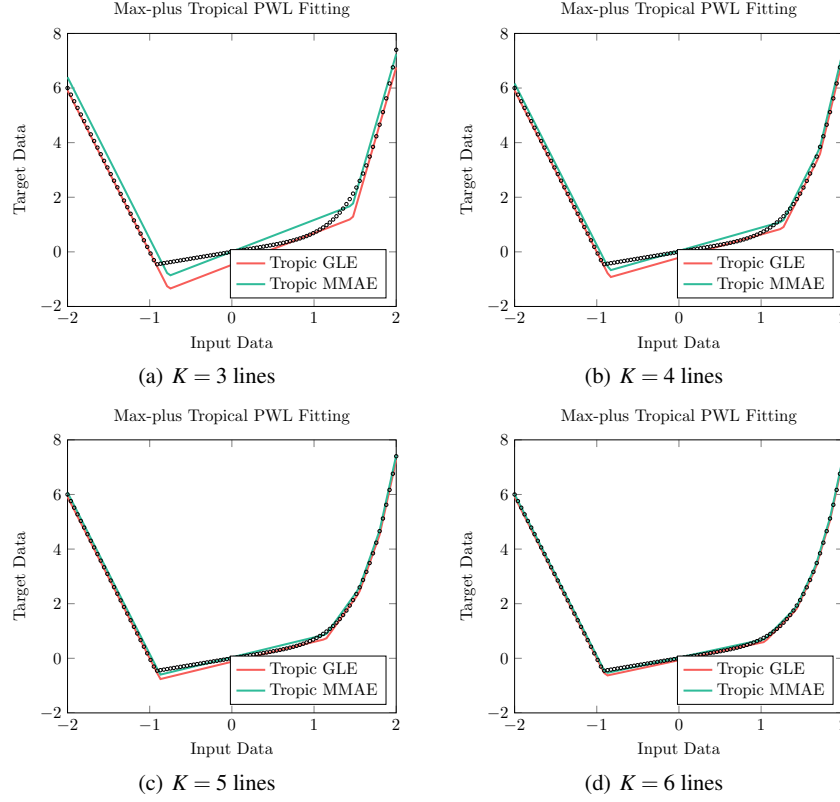


Fig. 11 Fitting max-plus tropical polynomials to the function (74). Best viewed in color.

$$\underbrace{\begin{bmatrix} -2y_1 & -2x_1 & -x_1 - y_1 & -y_1 & -x_1 & 0 & x_1 & y_1 & x_1 + y_1 & 2x_1 & 2y_1 \\ -2y_2 & -2x_2 & -x_2 - y_2 & -y_2 & -x_2 & 0 & x_2 & y_2 & x_2 + y_2 & 2x_2 & 2y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -2y_m & -2x_m & -x_m - y_m & -y_m & -x_m & 0 & x_m & y_m & x_m + y_m & 2x_m & 2y_m \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} b_{0,-2} \\ b_{-2,0} \\ b_{-1,-1} \\ b_{0,-1} \\ b_{-1,0} \\ b_{0,0} \\ b_{1,0} \\ b_{0,1} \\ b_{1,1} \\ b_{2,0} \\ b_{0,2} \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \quad (77)$$

By Theorems 3 and 4, the optimal subsolution (GLE) $\hat{\mathbf{w}}$ for minimum ℓ_p error and the optimal unconstrained solution $\tilde{\mathbf{w}}$ (for MMAE) equal

$$\hat{\mathbf{w}} = \mathbf{X}^* \boxplus' \mathbf{f}, \quad \tilde{\mathbf{w}} = \mu + \hat{\mathbf{w}} \quad (78)$$

where μ is half the ℓ_∞ error incurred by $\hat{\mathbf{w}}$. The GLE and MMAE solutions for the model are shown in Fig. 12 for fitting data from a noisy paraboloid surface.

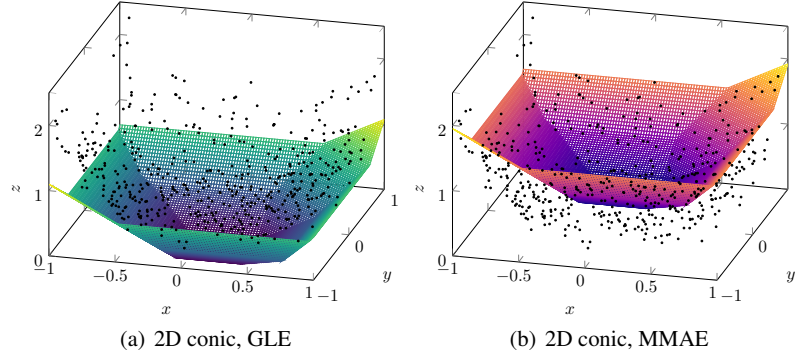


Fig. 12 Piecewise-linear surface approximation of a noisy paraboloid with a 2D tropical max-plus conic polynomial.

Example 6 The data (x_i, y_i, f_i) in Fig. 12 are 500 observations [35] from the noisy paraboloid surface $z = x^2 + y^2$ corrupted by a zero-mean random noise $\varepsilon \sim \mathcal{N}(0, 0.25^2)$. Thus, $f_i = f(x_i, y_i)$ where

$$f(x, y) = z + \varepsilon = x^2 + y^2 + \varepsilon \quad (79)$$

and the planar locations x_i, y_i of the data points were drawn as i.i.d. random variables $\sim \text{Unif}[-1, 1]$. Now, the general model we are fitting has degree K and is

$$p(x, y) = \max(a_1x + b_1y + c_1, \dots, a_Kx + b_Ky + c_K), \quad (80)$$

where the slopes (a_k, b_k) are computed using K -means on the numerical gradients of the 2D data, and the intercepts c_k are computed using the tropical fitting algorithm. See Fig. 13 for the resulting approximations and Table 4 for the error norms.

Computational Complexity:

The prevailing trend in recent methods to fitting m data points in \mathbb{R}^{n+1} using as model n -dimensional max-affine functions (i.e. max of K hyperplanes $\mathbf{a}_k^T + b_k$), which we view as max-plus tropical polynomials with real slope and intercept parameters, is a variety of iterative nonlinear least-squares algorithms. The number of model parameters is $K(n+1)$. The traditional least-squares estimator (LSE) [41, 43] solves a quadratic program with constraints and has a total complexity of $O((n+1)^3 m^3)$. Clearly, this becomes practically intractable for large number of data points and, also, as the dimensionality increases. In [57, 42] the nonlinear

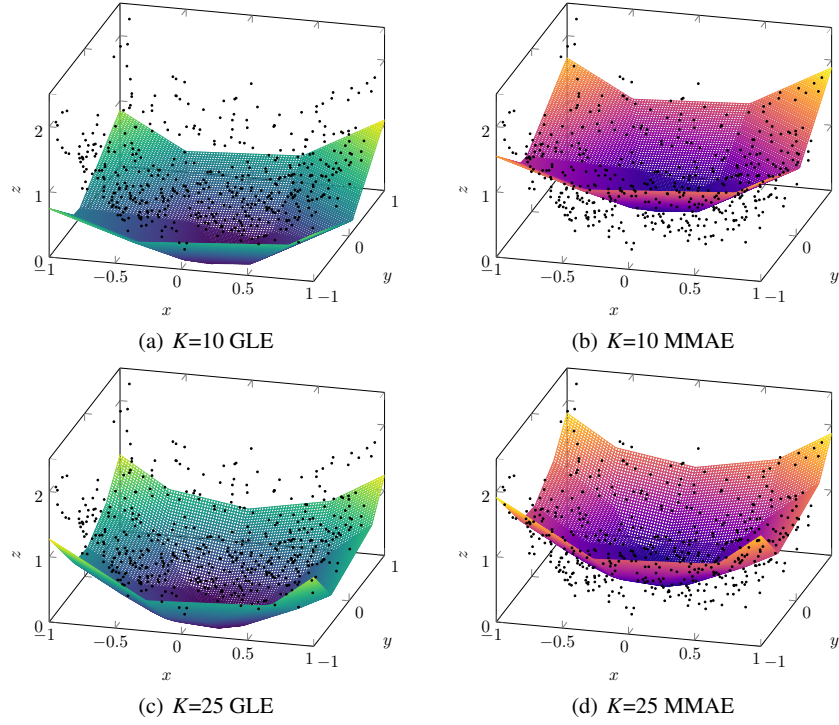


Fig. 13 2D Tropical fitting using the optimal constrained (GLE) and unconstrained (MMAE) approach to data from (79). Best viewed in color.

K	GLE		MMAE	
	error _{RMS}	$\ \text{error}\ _{\infty}$	error _{RMS}	$\ \text{error}\ _{\infty}$
11 (conic)	0.6307	1.7049	0.4167	0.8524
10	0.6659	1.6022	0.3641	0.8011
25	0.5674	1.2779	0.3016	0.6389
50	0.5489	1.3068	0.3159	0.6534
75	0.5433	1.2950	0.3150	0.6475
100	0.5364	1.2828	0.3135	0.6414
250	0.5273	1.2786	0.3172	0.6393

Table 4 Minimum RMS error and maximum absolute error for the optimal constrained (GLE) and unconstrained (MMAE) tropical fitting of the function (79) using either a 2D tropical conic or K -term optimal fit whose gradients are found via K -means.

least-squares problems is solved iteratively where each iteration involves some partitioning of the data into K clusters and least-squares fitting of hyperplanes over the different K clusters. This has a complexity of $O((n+1)^2 m i_C)$ where i_C is the number of iterations until convergence; however, this least-squares partition algorithm does not always converge, and even in cases of convergence the fit to the data may be poor. To overcome this obstacle, the authors in [57, 42] propose running sev-

eral instances of their algorithm, with different random initializations, in order to achieve a better fit to the data. The convex adaptive partitioning (CAP) algorithm proposed in [35] has a complexity at $O(n(n+1)^2 m \log(m) \log(\log(m)))$, where its most demanding part is linear regression since each least-squares fit has complexity $O((n+1)^2 m)$. Although the CAP algorithm seems to have a slightly larger complexity than that of [57], it provides a *consistent* estimator.

In contrast, the complexity of our algorithm is dominated only by the K -means computation, which has a complexity of $O(Kmni_K)$, where i_K is the number of K -means iterations. After the K centroids \mathbf{a}_k have been computed, our algorithm simply does a single pass over the data for the tropical regression to find the b_k , with total complexity $O(Kmn)$. Therefore, the overall complexity of our tropical regression algorithm (both via the GLE and the MMAE criteria) is $O(Kmni_K)$. In general, assuming that the data have some clustering structure, the required number of K -means iterations to find the slopes is small and thus our algorithm can be considered ‘linear’ in practice. As such, in non-pathological cases, we can assume that the product Ki_K is significantly smaller than m and can be treated as a constant, resulting in an overall complexity of $O(mn)$, thus improving on the CAP algorithm bound [35], and greatly improving on the traditional LSE. In terms of performance, as long as the number of clusters is not too small (and thus there are many elements in the cluster that are not adequately represented by the centroid), then the tropical algorithm will produce good PWL fits to the data (as is evident for the 1D case in Fig. 11 for $K = 6$ and the 2D case in Fig. 13 for $K = 25$.)

8 Conclusions

(Max-plus) Tropical Geometry and (Weighted) Mathematical Morphology share a common idempotent semiring arithmetic, which also has a dual counterpart. Both can be extended and generalized using max- \star algebra over Complete Weighted Lattices (CWLs) which are nonlinear vector spaces. The CWL framework allows for optimal solutions (using adjunctions and lattice projections) for general max- \star systems of equations, which are applied to optimal fitting of tropical lines or hyperplanes to data. This tropical regression provides convex piecewise-linear (PWL) approximations to curves and surfaces with max-affine functions at a linear complexity with respect to the number of data and their dimension, which is significantly lower than the complexity of least-square estimates for PWL shape regression.

Acknowledgements The authors wish to thank V. Charisopoulos for insightful discussions on tropical geometry and neural networks.

References

1. A. Achache. Galois Connexion of a Fuzzy Subset. *Fuzzy Sets and Systems*, 8:215–218, 1982.
2. M. Akian, S. Gaubert, and A. Guterman. Tropical Polyhedra Are Equivalent To Mean Payoff Games. *Int'l J. Algebra and Computation*, 22(1), 2012.
3. M. Akian, S. Gaubert, V. Nitica, and I. Singer. Best approximation in max-plus semimodules. *Linear Algebra and its Applications*, 435:3261–3296, 2011.
4. F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity: An Algebra for Discrete Event Systems*. J. Wiley & Sons, 1992, web ed. 2001.
5. G. J.F. Banon and J. Barrera. Decomposition of mappings between complete lattices by mathematical morphology, Part I. General lattices. *Signal Processing*, 30:299–327, 1993.
6. R. Bellman and W. Karush. On a New Functional Transform in Analysis: The Maximum Transform. *Bull. Amer. Math. Soc.*, 67:501–503, 1961.
7. G. Birkhoff. *Lattice Theory*. Amer. Math. Soc., Providence, Rhode Island, 3 edition, 1967.
8. C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
9. T. S. Blyth. *Lattices and Ordered Algebraic Structures*. Springer-Verlag, 2005.
10. T. S. Blyth and M. F. Janowitz. *Residuation Theory*. Pergamon Press, Oxford, 1972.
11. S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi. A tutorial on geometric programming. *Optim. Eng.*, 8:67–127, 2007.
12. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, 2004.
13. R. W. Brockett and P. Maragos. Evolution equations for continuous-scale morphological filtering. *IEEE Trans. Signal Processing*, 42(12):3377–3386, Dec. 1994.
14. P. Butkovič. *Max-linear Systems: Theory and Algorithms*. Springer, 2010.
15. G. C. Calafiore, S. Gaubert, and C. Possieri. A Universal Approximation Result for Difference of Log-Sum-Exp Neural Networks. *arXiv preprint arXiv:1905.08503*, 2019.
16. 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 Networks and Learning Systems*, 30(5):1–12, May 2019.
17. V. Charisopoulos and P. Maragos. Morphological Perceptrons: Geometry and Training Algorithms. In J. Angulo and et al., editors, *Proc. Int'l Symp. Mathematical Morphology (ISMM)*, volume 10225 of *LNCS*, pages 3–15. Springer, Cham, 2017.
18. V. Charisopoulos and P. Maragos. A tropical approach to neural networks with piecewise linear activations. *arXiv preprint arXiv:1805.08749*, 2018.
19. L. O. Chua and A.-C. Deng. Canonical Piecewise-Linear Representation. *IEEE Trans. Circuits and Systems*, 35(1):101–111, Jan. 1988.
20. L. O. Chua, C. A. Desoer, and E. S. Kuh. *Linear and Nonlinear Circuits*. McGraw-Hill, 1987.
21. G. Cohen, D. Dubois, J.P. Quadrat, and M. Viot. A linear system theoretic view of discrete event processes and its use for performance evaluation in manufacturing. *IEEE Trans. Automatic Control*, 30:210–220, 1985.
22. G. Cohen, S. Gaubert, and J.P. Quadrat. Duality and separation theorems in idempotent semimodules. *Linear Algebra and its Applications*, 379:395–422, 2004.
23. R. Cuninghame-Green. *Minimax Algebra*. Springer-Verlag, 1979.
24. R. A. Cuninghame-Green. Projections in Minimax Algebra. *Math. Programming*, 10:111–123, 1976.
25. R. A. Cuninghame-Green and K. Cechlarova. Residuation in fuzzy algebra and some applications. *Fuzzy Sets and Systems*, 71:227–239, 1995.
26. C. de Boor. *A Practical Guide to Splines*. Springer, 1978.
27. L. Dorst and R. van den Boomgaard. Morphological Signal Processing and the Slope Transform. *Signal Processing*, 38(1):79–98, 1994.
28. W. Fenchel. On conjugate convex functions. *Canadian J. Math.*, 1:73–77, 1949.
29. S. Gaubert and R. D. Katz. Minimal half-spaces and external representation of tropical polyhedra. *J. Algebr. Comb.*, 33:325–348, 2011.
30. S. Gaubert and Max Plus. Methods and Applications of $(\max, +)$ Linear Algebra. In *Proc. 14th Annual Symp. on Theor. Aspects of Comp. Science (STACS'97)*, number 1200 in *LNCS*, page 261–282. Springer, 1997.

31. G. Gierz, K. H. Hoffman, K. Keimel, J. D. Lawson, M. Mislove, and D. S. Scott. *A Compendium of Continuous Lattices*. Springer-Verlag, 1980.
32. M. Gondran and M. Minoux. *Graphs, Dioids and Semirings: New Models and Algorithms*. Springer, 2008.
33. I. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. In *Proc. Int'l Conf. on Machine Learning*, volume 28, pages 1319–1327. PMLR, 2013.
34. C. Güzelis and I. C. Gökner. A Canonical Representation for Piecewise-Affine Maps and Its Applications to Circuit Analysis. *IEEE Trans. Circuits and Systems*, 38(11):1342–1354, Nov. 1991.
35. L. A. Hannah and D. B. Dunson. Multivariate convex regression with adaptive partitioning. *arXiv preprint arXiv:1105.1924*, 2011.
36. L. A. Hannah and D. B. Dunson. Ensemble methods for convex regression with applications to geometric programming based circuit design. In *Proc. Int'l Conf. on Machine Learning*, 2012.
37. 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 Univ. Press, 2006.
38. H.J.A.M. Heijmans. *Morphological Image Operators*. Acad. Press, Boston, 1994.
39. H.J.A.M. Heijmans and P. Maragos. Lattice calculus of the morphological slope transform. *Signal Processing*, 59:17–42, 1997.
40. H.J.A.M. Heijmans and C. Ronse. The Algebraic Basis of Mathematical Morphology. Part I: Dilations and Erosions. *Computer Vision, Graphics, and Image Processing*, 50:245–295, 1990.
41. C. Hildreth. Point estimates of ordinates of concave functions. *J. Amer. Stat. Assoc.*, 49(267):598–619, 1954.
42. W. Hoberg, P. Kirschen, and P. Abbeel. Data fitting with geometric-programming-compatible softmax functions. *Optim. Eng.*, 17:897–918, 2016.
43. C. A. Holloway. On the estimation of convex functions. *Oper. Res.*, 27(2):401–407, 1979.
44. T. Hori and A. Nakamura. *Speech Recognition Algorithms Using Weighted Finite-State Transducers*. Morgan & Claypool, 2013.
45. J.-E. Perin. Tropical semirings. In J. Gunawardena, editor, *Idempotency*, pages 50–69. Cambridge Univ. Press, 1998.
46. P. Julian. The Complete Canonical Piecewise-Linear Representation: Functional Form Minimal Degenerate Intersections. *IEEE Trans. Circuits and Systems - I: Fund. Theories and Applic.*, 50(3):387–396, Mar. 2003.
47. C. Kahlert and L. O. Chua. A Generalized Canonical Piecewise-Linear Representation. *IEEE Trans. Circuits and Systems*, 37(3):373–383, Mar. 1990.
48. C. Kahlert and L. O. Chua. The Complete Canonical Piecewise-Linear Representation - Part I: The Geometry of the Domain Space. *IEEE Trans. Circuits and Systems - I: Fund. Theories and Applic.*, 39(3):222–236, Mar. 1992.
49. S. M. Kang and L. O. Chua. A Global Representation of Multidimensional Piecewise-Linear Functions with Linear Partitions. *IEEE Trans. Circuits and Systems*, 25(11):938–940, Nov. 1978.
50. J. Kim, L. Vandenbergh, and C.K.K. Yang. Convex Piecewise-Linear Modeling Method for Circuit Optimization via Geometric Programming. *IEEE Trans. Computer-Aided Design of Integr. Circuits Syst.*, 29(11):1823–1827, Nov. 2010.
51. J.-N. Lin and R. Unbehauen. Explicit Piecewise-Linear Models. *IEEE Trans. Circuits and Systems - I: Fund. Theories and Applic.*, 41(12):931–933, Dec. 1994.
52. J.-N. Lin, H.-Q. Xu, and R. Unbehauen. A Generalization of Canonical Piecewise-Linear Functions. *IEEE Trans. Circuits and Systems - I: Fund. Theories and Applic.*, 41(4):345–347, Apr. 1994.
53. G. L. Litvinov. Maslov Dequantization, Idempotent and Tropical Mathematics: A Brief Introduction. *Journal of Mathematical Sciences*, 140(3), 2007.
54. G. L. Litvinov, V. P. Maslov, and G. B. Shpiz. Idempotent functional analysis: An algebraic approach. *Math. Notes*, 69(5):696–729, 2001.

55. Y. Lucet. What shape is your conjugate? a survey of computational convex analysis and its applications. *SIAM Review*, 52(3):505–542, 2010.
56. D. Maclagan and B. Sturmfels. *Introduction to Tropical Geometry*. Amer. Math. Soc., 2015.
57. A. Magnani and S. P. Boyd. Convex piecewise-linear fitting. *Optim. Eng.*, 10:1–17, 2009.
58. P. Maragos. Morphological Systems: Slope Transforms and Max-Min Difference and Differential Equations. *Signal Processing*, 38(1):57–77, Jul. 1994.
59. P. Maragos. Slope Transforms: Theory and Application to Nonlinear Signal Processing. *IEEE Trans. Signal Processing*, 43(4):864–877, Apr. 1995.
60. P. Maragos. Lattice image processing: A unification of morphological and fuzzy algebraic systems. *J. Math. Imaging and Vision*, 22:333–353, 2005.
61. P. Maragos. Morphological filtering for image enhancement and feature detection. In A.C. Bovik, editor, *Image and Video Processing Handbook*, pages 135–156. Elsevier Acad. Press, 2 edition, 2005.
62. P. Maragos. Representations for morphological image operators and analogies with linear operators. In P.W. Hawkes, editor, *Advances in Imaging and Electron Physics*, volume 177, pages 45–187. Acad. Press: Elsevier Inc., 2013.
63. P. Maragos. Dynamical systems on weighted lattices: General theory. *Math. Control Signals Syst.*, 29(21), 2017.
64. P. Maragos. Tropical Geometry, Mathematical Morphology and Weighted Lattices. In B. Burgeth and et al., editors, *Proc. Int’l Symp. Mathematical Morphology (ISMM)*, volume 11564 of *LNCS*, pages 3–15. Springer Nature, 2019.
65. V. P. Maslov. On a new superposition principle for optimization problems. *Uspekhi Mat. Nauk [Russian Math. Surveys]*, 42(3):39–48, 1987.
66. W. M. McEneaney. *Max-Plus Methods for Nonlinear Control and Estimation*. Birkhauser, Boston, 2006.
67. F. Meyer. *Topographic Tools for Filtering and Segmentation 1 & 2*. Wiley, 2019.
68. G. Mikhalkin. Enumerative Tropical Algebraic Geometry in \mathbb{R}^2 . *J. Amer. Math. Soc.*, 18(2):313–377, 2005.
69. M. Mohri, F. Pereira, and M. Ripley. Weighted finite-state transducers in speech recognition. *Computer Speech and Language*, 16:69–88, 2002.
70. J.-J. Moreau. Inf-convolution, sous-additivité, convexité des fonctions numériques. *J. Math Pures at Appl.*, 49:109–154, 1970.
71. S. Ovchinnikov. Max-Min Representation of Piecewise Linear Functions. *Beitr. Algebra Geom.*, 43(1):297–302, 2002.
72. 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 Recognition*, 33:945–960, 2000.
73. G. X. Ritter and G. Urcid. Lattice algebra approach to single-neuron computation. *IEEE Trans. Neural Networks*, 14(2):282–295, Mar. 2003.
74. G. X. Ritter and J. N. Wilson. *Handbook of Computer Vision Algorithms in Image Algebra*. CRC Press, 2 edition, 2001.
75. R. T. Rockafellar. *Convex Analysis*. Princeton Univ. Press, Princeton, 1970.
76. J. Serra. *Image Analysis and Mathematical Morphology*. Acad. Press, 1982.
77. J. Serra, editor. *Image Analysis and Mathematical Morphology*, volume 2: Theoretical Advances. Acad. Press, 1988.
78. I. Simon. On semigroups of matrices over the tropical semiring. *Theoretical Informatics and Applications*, 28(3-4):277–294, 1994.
79. S. R. Sternberg. Grayscale morphology. *Computer Vision, Graphics, and Image Processing*, 35:333–355, 1986.
80. J. M. Tarela, E. Alonso, and M. V. Martinez. A Representation Method for PWL Functions Oriented to Parallel Processing. *Math. Comput. Modelling*, 13(10):75–83, 1990.
81. J. M. Tarela and M. V. Martinez. Region Configurations for Realizability of Lattice Piecewise-linear Models. *Math. Comput. Modelling*, 30:17–27, 1999.
82. S. Theodoridis. *Machine Learning: A Bayesian and Optimization Perspective*. Elsevier Acad. Press, 2015.

83. A. Toriello and J. P. Vielma. Fitting piecewise linear continuous functions. *Europ. J. Operational Research*, 219:86–95, 2012.
84. A. Tsiamis and P. Maragos. Sparsity in Max-plus Algebra. *Discrete Events Dynamic Systems*, 29:163–189, May 2019.
85. T.J.J. van den Boom and B. De Schutter. Modeling and control of switching max-plus-linear systems with random and deterministic switching. *Discrete Event Dyn. Syst.*, 22:293–33, 2012.
86. O. Viro. Dequantization of Real Algebraic Geometry on Logarithmic Paper. In C. Casacuberta and et al., editors, *European Congress of Mathematics, Vol. I (Barcelona 2000)*, volume 201 of *Progress in Mathematics*, pages 135–146, Basel, 2001. Birkhauser.
87. S. Wang. General Constructive Representations for Continuous Piecewise-Linear Functions. *IEEE Trans. Circ. Syst.-I: Regular Papers*, 51(9):1889–1896, Sep. 2004.
88. R. H. Wilkinson. A Method For Generating Functions of Several Variables Using Analog Diode Logic. *IEEE Trans. Electronic Computers*, pages 112–129, 1963.
89. P.-F. Yang and P. Maragos. Min-Max Classifiers: Learnability, Design And Application. *Pattern Recognition*, 28(6):879–899, 1995.
90. L. Zhang, G. Naitzat, and L.-H. Lim. Tropical geometry of deep neural networks. In *Proc. Int’l Conf. on Machine Learning*, volume 80, pages 5824–5832. PMLR, 2018.
91. Y. Zhang, S. Blusseau, S. Velasco-Forero, I. Bloch, and J. Angulo. Max-Plus Operators Applied to Filter Selection and Model Pruning in Neural Networks. In B. Burgeth and et al., editors, *Proc. Int’l Symp. Mathematical Morphology (ISMM)*, volume 11564 of *LNCS*, pages 310–322. Springer Nature, 2019.
92. U. Zimmermann. *Linear and Combinatorial Optimization in Ordered Algebraic Structures*. North-Holland Publ., 1981.