Proceedings of the 45th IEEE International Conference on Acoustics, Speech and Signal Processing, May 2020.

MULTIVARIATE TROPICAL REGRESSION AND PIECEWISE-LINEAR SURFACE FITTING

Petros Maragos[1] *and Emmanouil Theodosis*[2]

[1]School of ECE, National Technical University of Athens, Athens 15773, Greece
[2]School of Engineering & Applied Sciences, Harvard University, Cambridge, MA 02138, USA
maragos@cs.ntua.gr , etheodosis@g.harvard.edu

ABSTRACT

In this paper we propose a novel approach for multivariate convex regression by using as approximation model a maximum of hyperplanes, which we represent as a multivariate max-plus tropical polynomial. Our approach uses concepts from tropical geometry and finds an optimal solution for the model parameters (that minimizes a data fitting error norm) by solving systems of max-plus equations using max-plus algebra and projections on weighted lattices. Our method has lower complexity than most other methods for fitting piecewise-linear (PWL) functions and we apply it to optimal PWL regression for fitting max-plus tropical surfaces to arbitrary data that constitute polyhedral shape approximations.

***Index Terms*—** multivariate convex regression, piecewise-linear surface fitting, tropical geometry, max-plus algebra

1. INTRODUCTION

In multidimensional signal modeling and machine learning a fundamental nonlinear regression problem is fitting *piecewise-linear (PWL)* functions to data. Approximations with PWL functions have proven analytically and computationally very useful in many fields of science and engineering, including splines, nonlinear circuits and systems modeling, machine learning, convex optimization, geometric programming, and statistics. Two major problems in this area 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.

The authors of [1, 2] 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. Tarela et al [3], by combining their previous work on representing PWL functions with lattice generalizations of Boolean polynomials with the general $f - \phi$ model for PWL functions of [4], developed a constructive way to generate min-max combinations of affine functions which provide a complete representation of continuous PWL functions in arbitrary dimensions. Wang [5] completed the construction of a canonical representation for arbitrary continuous PWL functions in n-dimensions by starting from the lattice presentation of [3], 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) and use them for data fitting. Starting from early least-squares solutions, some representative recent approaches to solve this *convex regression* problem include [6, 7, 8, 9]. 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. This iteration is viewed in [9] as a Gauss-Newton algorithm, similar to the K-means algorithm. The order K of the model can be increased until some error threshold is reached. 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 [7, 10].

In this paper we focus on multivariate *convex* PWL functions which, if represented as maximum of hyperplanes, can be identified with max-plus tropical polynomials. This allows us to use concepts and tools from the mathematical fields of tropical geometry and max-plus algebra to optimally solve a fundamental regression problem of approximating the shape of surfaces by fitting tropical polynomials to data, possibly in the presence of noise.

Tropical Geometry [11, 12] uses a max-plus and min-plus semiring arithmetic, which are also used in other fields including: max-plus control and optimization [13, 14, 15, 16, 17, 18]; finite automata [19]; convex analysis [20]; morphological image analysis [21, 22, 23]; speech recognition [24]; neural networks [25, 26, 27, 28].

In this paper, whose preliminary version is based on [29] and whose full theoretical details can be found in [30], we begin in Section 2 with some elementary concepts from weighted lattices, max-plus algebra, and tropical geometry. Then, in Section 3 we outline the optimal solution of max-plus equations using projections on weighted lattices, and apply it in Section 4 to optimal piecewise-linear regression for fitting max-plus tropical surfaces to arbitrary data that constitute polyhedral shape approximations.

2. BACKGROUND CONCEPTS

Notation: For max and min operations we use the well-established lattice-theoretic symbols of $\vee$ and $\wedge$. 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 also 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.

Weighted Lattices and Max-plus Algebra: Max and min operations (or more generally supremum and infimum) form the algebra of lattices. Max-plus arithmetic forms an idempotent semiring $(\mathbb{R}_{\max}, \vee, +)$ where $\mathbb{R}_{\max} = \mathbb{R} \cup \{-\infty\}$ and the real number addition and multiplication are replaced by the max and sum operations, respectively. We combine the max-plus and min-plus scalar arithmetic into an algebraic structure called complete lattice-ordered double monoid (*clodum*) which consists of the extended reals $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ equipped with the maximum ($\vee$), minimum ($\wedge$), and addition ($+$) operations. Then, we consider nonlinear vector spaces, called *complete weighted lattices (CWLs)* [31], where the traditional vector addition of linear spaces is replaced by vector supremum and its dual infimum, and the multiplication of vectors by scalars is replaced by addition of vectors with scalars from $\overline{\mathbb{R}}$. In this paper we work on the finite-dimensional CWL $\overline{\mathbb{R}}^n$, equipped with the lattice pointwise operations of 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 $\mathbf{x}, \mathbf{y} \in \overline{\mathbb{R}}^n$. Vector transformations from $\overline{\mathbb{R}}^n$ to $\overline{\mathbb{R}}^m$ that obey a max-plus (resp. min-plus) superposition are lattice dilations δ (resp. erosions ε) and can be represented as a max-plus product $\boxplus$ (resp. min-plus product $\boxplus'$) of a matrix $\mathbf{A} \in \overline{\mathbb{R}}^{m \times n}$ with an input vector $\mathbf{x} \in \overline{\mathbb{R}}^n$; these tropical matrix products are defined as:

$$[\mathbf{A} \boxplus \mathbf{B}]_{ij} \triangleq \bigvee_k a_{ik} + b_{kj} \ , \ [\mathbf{A} \boxplus' \mathbf{B}]_{ij} \triangleq \bigwedge_k a_{ik} +' b_{kj} \quad (1)$$

Thus, $\delta_{\mathbf{A}}(\mathbf{x}) \triangleq \mathbf{A} \boxplus \mathbf{x}$ is a dilation, and $\varepsilon_{\mathbf{A}}(\mathbf{x}) \triangleq \mathbf{A} \boxplus' \mathbf{x}$ is its dual erosion vector operator. More general clodums, using a max-$\star$ algebra where the operation $\star$ distributes over max, and corresponding CWLs are developed in [31].

Tropical Polynomial Curves and Surfaces: Consider the analytic expressions for a Euclidean line and parabola:

$$p_1(x) = ax + b, \quad p_2(x) = ax^2 + bx + c \quad (2)$$

'Tropicalization' (i.e. replacing sum with max and multiplication with addition) yields corresponding max-plus tropical polynomials:

$$p_1^{\max}(x) = \max(a{+}x, b), \quad p_2^{\max}(x) = \max(a{+}2x, b{+}x, c) \quad (3)$$

The graphs of all the above can be seen in Fig. 1. The equations for the min-plus case are identical as in (3) by replacing max with min.

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

$$f_1(x, y) = \min(5{+}x, 7{+}y, 9), \quad f_2(x, y) = \max(0{+}x, 2{+}y, 7), \quad (4)$$

whose graphs can be seen as surfaces in Fig. 1. As a next example, to the general Euclidean conic polynomial $ax^2{+}bxy{+}cy^2{+}dx{+}ey{+}f$ there corresponds the following two-variable max-plus tropical polynomial of degree 2:

$$p_{\mathrm{t-conic}}(x, y) = \max(a{+}2x, b{+}x{+}y, c{+}2y, d{+}x, e{+}y, f) \quad (5)$$

Its min-plus version is shown in Fig. 1.

In [30] we explain analytically how the tropical polynomials result from a dequantization [18] of real algebraic geometry [12].

3. SOLVING MAX-$\star$ EQUATIONS AND OPTIMIZATION

Consider the scalar clodum $(\overline{\mathbb{R}}, \vee, \wedge, +)$, a matrix $\mathbf{A} \in \overline{\mathbb{R}}^{m \times n}$ and a vector $\mathbf{b} \in \overline{\mathbb{R}}^m$. The set of solutions of the max-plus equation

$$\mathbf{A} \boxplus \mathbf{x} = \mathbf{b} \quad (6)$$

is either empty or forms an idempotent semigroup under vector $\vee$. 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 $\mathbf{A} \boxplus \mathbf{x} = \mathbf{b}$ 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. The optimum will be the solution of the following constrained minimization problem:

$$\text{Minimize } \|\mathbf{A} \boxplus \mathbf{x} - \mathbf{b}\|_p \ \text{ s.t. } \ \mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b} \quad (7)$$

where the norm $|| \cdot ||_p$ is any ℓ_p norm with $p = 1, 2, \ldots, \infty$. While the two above problems have been solved in [32] for the max-plus case and for $p = 1$ or $p = \infty$, in [31] a more general result has been found using adjunctions for the general case of an arbitrary clodum.

Theorem 1 *([32]) (a) If Eq. (6) has a solution, then its greatest solution is*

$$\hat{\mathbf{x}} = \varepsilon(\mathbf{b}) = \mathbf{A}^* \boxplus' \mathbf{b} = [\bigwedge_i b_i - a_{ij}], \quad \mathbf{A}^* \triangleq -\mathbf{A}^T \quad (8)$$

(b) The unique solution to problem (7) is (8) for $p = 1$ and $p = \infty$.
(c) 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, then

$$\tilde{\mathbf{x}} = \mu + \mathbf{A}^* \boxplus' \mathbf{b} \quad (9)$$

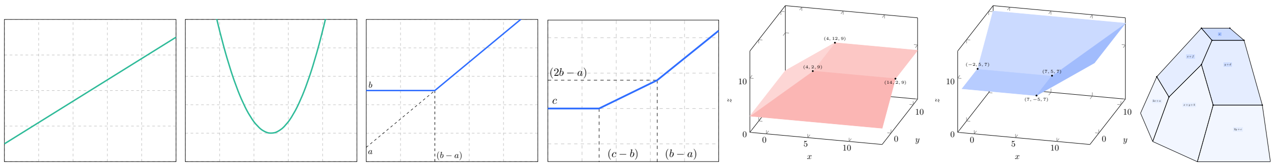
is the unique optimum solution of the unconstrained problem of minimizing $\|\mathbf{A} \boxplus \mathbf{x} - \mathbf{b}\|_\infty$.

The computational complexity to find both optimal solutions $\hat{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ is $O(mn)$ additions.

Consider the vector dilation $\delta(\mathbf{x}) = \mathbf{A} \boxplus \mathbf{x}$; then $\varepsilon(\mathbf{x}) = \mathbf{A}^* \boxplus' \mathbf{x}$ is its *adjoint* vector erosion. A main idea for solving (7) is to consider vectors $\mathbf{x}$ that are *subsolutions*, i.e. $\delta(\mathbf{x}) = A \boxplus \mathbf{x} \leq \mathbf{b}$, and find the greatest such subsolution $\hat{\mathbf{x}} = \varepsilon(\mathbf{b})$, which yields either the greatest exact solution of (6) or an optimum subsolution. This creates a lattice projection onto the max-plus span of the columns of $\mathbf{A}$ via the lattice opening $\delta(\varepsilon(\mathbf{b})) \leq \mathbf{b}$ that best approximates $\mathbf{b}$ from below. Details can be found in [31]. Projections on idempotent semimodules, which are weaker algebraic structures than weighted lattices, have been studied in [33, 13].

4. OPTIMAL FITTING TROPICAL POLYNOMIALS TO DATA AND SHAPE APPROXIMATION

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 equations of Sec. 3. 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 [34, 20]. Closely related ideas are based on lattice-theoretic slope transforms [35] that generalize this result to non-convex functions and approximate representations.



(a) Euclid line (b) Euclid parabola (c) Tropic line (d) Tropic parabola (e) Surface f_1 (f) Surface f_2 (g) Tropic conic

Fig. 1. (a)-(d) Euclidean and tropical 1D polynomials. (e)-(f) Tropical planes in (4). (g) Min-plus version of tropic conic in (5).

4.1. Optimal Fitting Tropical Lines and Planes

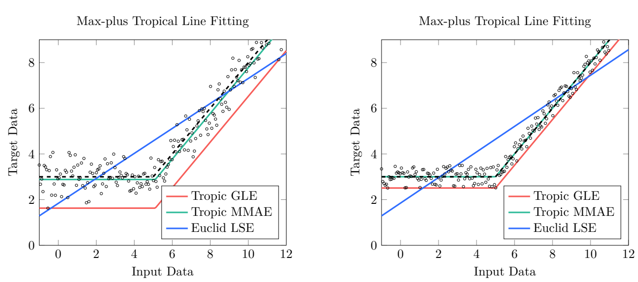
We examine a classic problem in machine learning, fitting a line to data by minimizing an error norm, in the light of tropical geometry. Given data $(x_i, f_i) \in \mathbb{R}^2$, $i = 1, ..., m$, if we wish to fit a Euclidean line $f(x) = ax + b$ by minimizing the ℓ_2 error norm, the optimal solution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

$$\hat{a}_{\text{LS}} = \frac{m\sum_i x_i f_i - (\sum_i x_i)(\sum_i f_i)}{m\sum_i (x_i)^2 - (\sum_i x_i)^2}, \quad \hat{b}_{\text{LS}} = \frac{1}{m}\sum_i (f_i - \hat{a}_{\text{LS}} x_i) \tag{10}$$

Suppose now we wish to fit a general tropical line $f(x) = \max(a + x, b)$ by minimizing some ℓ_p error norm. The equations and solution for finding the optimal parameters $\mathbf{w} = (a, b)$ become:

$$\underbrace{\begin{bmatrix} x_1 & 0 \\ \vdots & \vdots \\ x_m & 0 \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \Longrightarrow \underbrace{\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}}_{\hat{\mathbf{w}}} = \underbrace{\begin{bmatrix} \bigwedge_i f_i - x_i \\ \bigwedge_i f_i \end{bmatrix}}_{\mathbf{X}^* \boxplus' \mathbf{f}} \tag{11}$$

This vector $\hat{\mathbf{w}}$ yields (after max-plus 'multiplication' with $\mathbf{X}$) the *greatest lower estimate (GLE)* of the data $\mathbf{f}$. Thus, the above approach allows to optimally fit (w.r.t. *any* ℓ_p error norm) tropical lines to arbitrary data from below. In addition, 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 1, $\tilde{\mathbf{w}} = \hat{\mathbf{w}} + \mu$ where $\mu = \frac{1}{2}\|\mathbf{X} \boxplus \hat{\mathbf{w}} - \mathbf{f}\|_\infty$. Figure 2 shows an example.



(a) T-line with Gaussian Noise (b) T-line with Uniform Noise

Fig. 2. (a) Optimal fitting of a max-plus tropical line $y = \max(x - 2, 3)$ (shown in black dash 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 $\sim$ Unif$[-0.5, 0.5]$.

The above approach can be extended to fitting tropical planes

$$f(x, y) = \max(a + x, b + y, c) \tag{12}$$

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

$$\underbrace{\begin{bmatrix} x_1 & y_1 & 0 \\ \vdots & \vdots & \vdots \\ x_m & y_m & 0 \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \tag{13}$$

By Theorem 1 the optimal subsolution is

$$\begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = \hat{\mathbf{w}} = \mathbf{X}^* \boxplus' \mathbf{f} = \begin{bmatrix} \bigwedge_{i=1}^m f_i - x_i \\ \bigwedge_{i=1}^m f_i - y_i \\ \bigwedge_{i=1}^m f_i \end{bmatrix} \tag{14}$$

The MMAE solution is given by $\tilde{\mathbf{w}} = \hat{\mathbf{w}} + \mu$ where $\mu = \frac{1}{2}\|\mathbf{X} \boxplus \hat{\mathbf{w}} - \mathbf{f}\|_\infty$, but the data matrix $\mathbf{X}$ and vector $\mathbf{f}$ refer now to the plane.

4.2. Surface Regression by Fitting Tropical Polynomials

The above approach and solution can also be generalized to polynomials of higher degree and to multi-dimensional data. We wish to fit a n-dimensional max-plus tropical polynomial

$$f(\mathbf{x}) = \max(\mathbf{a}_1^T\mathbf{x} + b_1, \mathbf{a}_2^T\mathbf{x} + b_2, \ldots, \mathbf{a}_K^T\mathbf{x} + b_K) = \bigvee_{k=1}^K \mathbf{a}_k^T\mathbf{x} + b_k \tag{15}$$

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

$$\underbrace{\begin{bmatrix} \mathbf{a}_1^T\mathbf{x}_1 & \mathbf{a}_2^T\mathbf{x}_1 & \cdots & \mathbf{a}_K^T\mathbf{x}_1 \\ \mathbf{a}_1^T\mathbf{x}_2 & \mathbf{a}_2^T\mathbf{x}_2 & \cdots & \mathbf{a}_K^T\mathbf{x}_2 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{a}_1^T\mathbf{x}_m & \mathbf{a}_2^T\mathbf{x}_m & \cdots & \mathbf{a}_K^T\mathbf{x}_m \end{bmatrix}}_{\mathbf{X}} \boxplus \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}}_{\mathbf{f}} \tag{16}$$

We assume that the slope vectors $\mathbf{a}_k$ are given and we optimize for the parameters $\{b_k\}$. By Theorem 1, the optimal subsolution for minimum ℓ_p error is

$$\begin{bmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_K \end{bmatrix} = \hat{\mathbf{w}} = \mathbf{X}^* \boxplus' \mathbf{f} = \begin{bmatrix} \bigwedge_{i=1}^m f_i - \mathbf{a}_1^T\mathbf{x}_i \\ \vdots \\ \bigwedge_{i=1}^m f_i - \mathbf{a}_K^T\mathbf{x}_i \end{bmatrix} \tag{17}$$

Note that $\mathbf{X} \boxplus \hat{\mathbf{w}} \leq \mathbf{f}$. Further, by Theorem 1, the unconstrained solution that yields the minimum ℓ_∞ error is

$$\tilde{\mathbf{w}} = \mu + \hat{\mathbf{w}}, \quad \mu = \frac{1}{2}\|\mathbf{X} \boxplus \hat{\mathbf{w}} - \mathbf{f}\|_\infty \tag{18}$$

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 allow the $\mathbf{a}_k$ to 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, cluster the data gradients using K-means and use the centroids of the K clusters as our given slope vectors. Next we apply the above approaches for optimally solving two cases with 2D examples.

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

$$p(x,y) = \bigvee_{0\leq|k+\ell|\leq 2,\ k\cdot\ell\geq 0} b_{k\ell} + kx + \ell y \tag{19}$$

to given data $(x_i, y_i, f_i) \in \mathbb{R}^3$, $i = 1, ..., m$, where $f_i = p(x_i, y_i) +$ error by minimizing some ℓ_p error norm. From our general result in (18), the optimal unconstrained solution (for MMAE) is $\tilde{\mathbf{w}} = \mu + \hat{\mathbf{w}}$ where $\hat{\mathbf{w}} = \mathbf{X}^* \boxplus' \mathbf{f}$ and μ is half the ℓ_∞ error incurred by $\hat{\mathbf{w}}$. The MMAE solutions for the model are shown in Fig. 3 for fitting data from a noisy paraboloid.

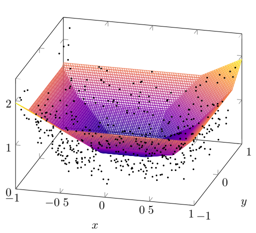
The 3D data tuples in Fig. 3 are 500 observations from the noisy paraboloid surface [6]

$$z = x^2 + y^2 + \epsilon \tag{20}$$

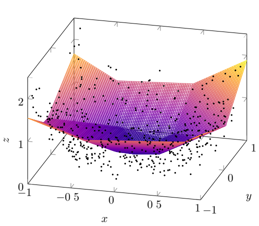
where $\epsilon \sim \mathcal{N}(0, 0.25^2)$ is zero-mean noise and the planar locations x_i, y_i of the data points were drawn as i.i.d. random variables $\sim$ Unif$[-1, 1]$. Now, the model we are fitting has rank K and is

$$f(x,y) = \max(a_1 x + b_1 y + c_1, ..., a_K x + b_K y + c_K), \tag{21}$$

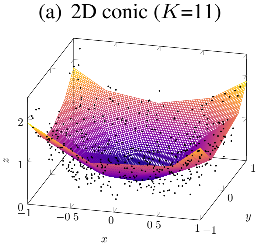
where (a_k, b_k) are computed using K-means on the numerical gradients of the 2D data, and c_k are computed using the tropical fitting algorithm. See Fig. 3. The errors are given in Table 1.



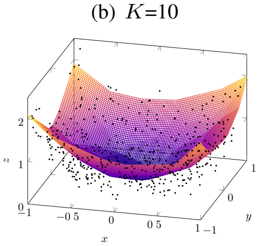
(a) 2D conic (K=11)



(b) K=10



(c) K=25



(d) K=100

Fig. 3. 2D Tropical fitting using the optimal unconstrained (MMAE) approach to data from (20). (a) Tropic conic with known integer slopes. (b)-(d) Slopes found via K-means on gradients.

Computational Complexity: The prevailing trend in recent methods to fitting m data points in $\mathbb{R}^{n+1}$ using a max of K n-dimensional hyperplanes $\mathbf{a}_k^T\mathbf{x} + b_k$, which we view as max-plus tropical polynomials, is a variety of iterative nonlinear least-squares algorithms. The number of model parameters is $K(n+1)$. The traditional least-squares estimator (LSE) [6] solves a quadratic program with constraints and has a total complexity of $O((n+1)^3 m^3)$. Clearly, this becomes intractable for large number of data points and, also, as the dimensionality increases. In [9, 7] the nonlinear 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 $O((n+1)^2 m\imath)$ where $\imath$ 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. The authors in [9, 7] propose running several instances of their algorithm, with different random initializations, in order to achieve a better fit to the data. The convex adaptive partitioning algorithm proposed in [6] is consistent and has a complexity of $O(n(n+1)^2 m \log(m) \log(\log(m)))$; its most demanding part is the least-squares fits, each of complexity $O((n+1)^2 m)$.

	GLE		MMAE	
K	error$_{\text{RMS}}$	$\|\|\text{error}\|\|_\infty$	error$_{\text{RMS}}$	$\|\|\text{error}\|\|_\infty$
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
100	0.5364	1.2828	0.3135	0.6414

Table 1. Minimum RMS error and minimum maximum absolute error for the optimal unconstrained tropical fitting of the function (20) using either a 2D tropical conic or a K-term optimal fit whose gradients are found via K-means.

In contrast, the complexity of our algorithm is dominated only by the K-means computation, which has a complexity of $O(Kmni)$, where i 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 $O(Km)$. Therefore, the overall complexity of our tropical regression algorithm is $O(Kmni)$. 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. In non-pathological cases, we can assume that the iK 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 [6], and greatly improving on the traditional LSE. In terms of performance, as long as the number of clusters is not too small (so that in each cluster its elements are adequately represented by the centroid), then the tropical algorithm will produce good PWL fits to the data.

5. CONCLUSIONS

Tropical geometry and max-plus algebra share a common idempotent semiring arithmetic, which also has a dual counterpart. Both can be extended and generalized using a weighted lattice algebra over corresponding nonlinear vector spaces. Using this framework, we developed optimal solutions (using adjunctions of vector dilations and erosions that are lattice projections) for solving general max-plus systems of equations and optimal fitting of tropical hyperplanes to data. This tropical regression provides piecewise-linear (PWL) approximations to surfaces at a linear complexity, which is significantly lower than least-square estimates for PWL shape regression of multi-dimensional data.

6. REFERENCES

[1] L. O. Chua and A.-C. Deng, "Canonical Piecewise-Linear Representation," *IEEE Trans. Circuits and Systems*, vol. 35, no. 1, pp. 101–111, Jan. 1988.

[2] 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.*, vol. 39, no. 3, pp. 222–236, Mar. 1992.

[3] J. M. Tarela and M. V. Martinez, "Region Configurations for Realizability of Lattice Piecewise-linear Models," *Math. Comput. Modelling*, vol. 30, pp. 17–27, 1999.

[4] J-N. Lin and R. Unbehauen, "Explicit Piecewise-Linear Models," *IEEE Trans. Circuits and Systems - I: Fund. Theories and Applic.*, vol. 41, no. 12, pp. 931–933, Dec. 1994.

[5] S. Wang, "General Constructive Representations for Continuous Piecewise-Linear Functions," *IEEE Trans. Circ. Syst.-I: Regular Papers*, vol. 51, no. 9, pp. 1889–1896, Sep. 2004.

[6] L. A. Hannah and D. B. Dunson, "Multivariate convex regression with adaptive partitioning," *arXiv preprint arXiv:1105.1924*, 2011.

[7] W. Hoburg, P. Kirschen, and P. Abbeel, "Data fitting with geometric-programming-compatible softmax functions," *Optim. Eng.*, vol. 17, pp. 897–918, 2016.

[8] J. Kim, L. Vandenberghe, 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.*, vol. 29, no. 11, pp. 1823–1827, Nov. 2010.

[9] A. Magnani and S. P. Boyd, "Convex piecewise-linear fitting," *Optim. Eng.*, vol. 10, pp. 1–17, 2009.

[10] 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*, vol. 30, no. 5, pp. 1–12, May 2019.

[11] D. Maclagan and B. Sturmfels, *Introduction to Tropical Geometry*, Amer. Math. Soc., 2015.

[12] O. Viro, "Dequantization of Real Algebraic Geometry on Logarithmic Paper," in *European Congress of Mathematics, Vol. I (Barcelona 2000)*, C. Casacuberta and et al., Eds., Basel, 2001, vol. 201 of *Progress in Mathematics*, pp. 135–146, Birkhauser.

[13] M. Akian, S. Gaubert, V. Nitica, and I. Singer, "Best approximation in max-plus semimodules," *Linear Alegbra and its Applications*, vol. 435, pp. 3261–3296, 2011.

[14] 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.

[15] P. Butkovič, *Max-linear Systems: Theory and Algorithms*, Springer, 2010.

[16] 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*, vol. 30, pp. 210–220, 1985.

[17] M. Gondran and M. Minoux, *Graphs, Dioids and Semirings: New Models and Algorithms*, Springer, 2008.

[18] G. L. Litvinov, "Maslov Dequantization, Idempotent and Tropical Mathematics: A Brief Introduction," *Journal of Mathematical Sciences*, vol. 140, no. 3, 2007.

[19] J.-E. Perin, "Tropical semirings," in *Idempotency*, J. Gunawardena, Ed., pp. 50–69. Cambridge Univ. Press, 1998.

[20] Y. Lucet, "What shape is your conjugate? a survey of computational convex analysis and its applications," *SIAM Review*, vol. 52, no. 3, pp. 505–542, 2010.

[21] H.J.A.M. Heijmans, *Morphological Image Operators*, Acad. Press, Boston, 1994.

[22] P. Maragos, "Morphological filtering for image enhancement and feature detection," in *Image and Video Processing Handbook*, A.C. Bovik, Ed., pp. 135–156. Elsevier Acad. Press, 2 edition, 2005.

[23] J. Serra, Ed., *Image Analysis and Mathematical Morphology*, vol. 2: Theoretical Advances, Acad. Press, 1988.

[24] M. Mohri, F. Pereira, and M. Ripley, "Weighted finite-state transducers in speech recognition," *Computer Speech and Language*, vol. 16, pp. 69–88, 2002.

[25] V. Charisopoulos and P. Maragos, "Morphological Perceptrons: Geometry and Training Algorithms," in *Proc. Int'l Symp. Mathematical Morphology (ISMM)*, J. Angulo and et al., Eds. 2017, vol. 10225 of *LNCS*, pp. 3–15, Springer, Cham.

[26] I. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," in *Proc. Int'l Conf. on Machine Learning*. 2013, vol. 28, pp. 1319–1327, PMLR.

[27] G. X. Ritter and G. Urcid, "Lattice algebra approach to single-neuron computation," *IEEE Trans. Neural Networks*, vol. 14, no. 2, pp. 282–295, Mar. 2003.

[28] L. Zhang, G. Naitzat, and L.-H. Lim, "Tropical geometry of deep neural networks," in *Proc. Int'l Conf. on Machine Learning*. 2018, vol. 80, pp. 5824–5832, PMLR.

[29] P. Maragos, "Tropical Geometry, Mathematical Morphology and Weighted Lattices," in *Proc. Int'l Symp. Mathematical Morphology (ISMM)*, B. Burgeth and et al., Eds. 2019, vol. 11564 of *LNCS*, pp. 3–15, Springer Nature.

[30] P. Maragos and E. Theodosis, "Tropical geometry and piecewise-linear approximation of curves and surfaces on weighted lattices," *arXiv:1912.03891*, 2019.

[31] P. Maragos, "Dynamical systems on weighted lattices: General theory," *Math. Control Signals Syst.*, vol. 29, no. 21, 2017.

[32] R. Cuninghame-Green, *Minimax Algebra*, Springer-Verlag, 1979.

[33] G. Cohen, S. Gaubert, and J.P. Quadrat, "Duality and separation theorems in idempotent semimodules," *Linear Alegbra and its Applications*, vol. 379, pp. 395–422, 2004.

[34] W. Fenchel, "On conjugate convex functions," *Canadian J. Math.*, vol. 1, pp. 73–77, 1949.

[35] H.J.A.M. Heijmans and P. Maragos, "Lattice calculus of the morphological slope transform," *Signal Processing*, vol. 59, no. 1, pp. 17–42, May 1997.