Sparsity in Max-plus Algebra And Applications in Multivariate Convex Regression

Nikos Tsilivis [1], Anastasios Tsiamis [2], Petros Maragos [1]

[1]School of ECE, National Technical University of Athens, Greece

[2]ESE Department, SEAS, University of Pennsylvania, USA

June 6-11, 2021

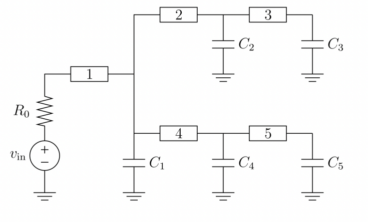
ICASSP2021 TORONTO Canada June 6-11, 2021 Metro Toronto Convention Centre

Overview

Convex regression - why?

Problem: Learn a convex function from data.

Optimization



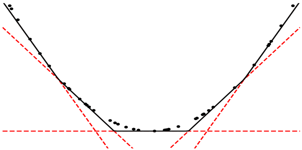
Machine Learning, Signal Processing & Finance

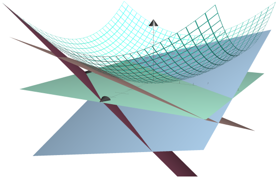
Convex regression

Data $(\mathbf{x}_i, y_i)$, where $y_i = f(\mathbf{x}_i)$, f an **unknown** convex function. How to estimate f?

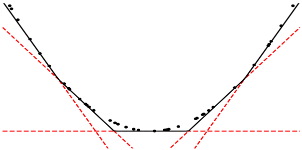
Data $(\mathbf{x}_i, y_i)$, where $y_i = f(\mathbf{x}_i)$, f an **unknown** convex function. How to estimate f?

Idea from Convex Analysis [Rockafellar 1970, Maragos 1994]: Any convex function $\approx$ maximum of hyperplanes $\mathbf{a}_j^\intercal \mathbf{x} + \mathbf{b}_j$.

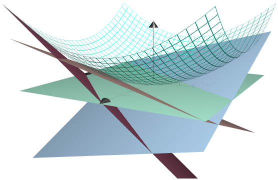




Data $(\mathbf{x}_i, y_i)$, where $y_i = f(\mathbf{x}_i)$, f an **unknown** convex function. How to estimate f?
Idea from Convex Analysis [Rockafellar 1970, Maragos 1994]: Any convex function $\approx$ maximum of hyperplanes $\mathbf{a}_j^\intercal \mathbf{x} + \mathbf{b}_j$.

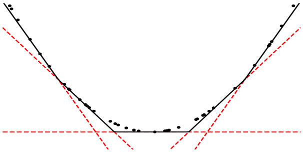


$$\max(\mathbf{a}_1^\intercal \mathbf{x}_1 + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_1 + b_K) = y_1$$
$$\max(\mathbf{a}_1^\intercal \mathbf{x}_2 + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_2 + b_K) = y_2$$
$$\ldots$$

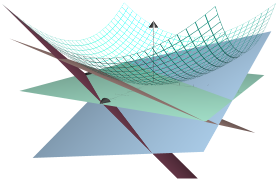


Data $(\mathbf{x}_i, y_i)$, where $y_i = f(\mathbf{x}_i)$, f an **unknown** convex function. How to estimate f?

Idea from Convex Analysis [Rockafellar 1970, Maragos 1994]: Any convex function $\approx$ maximum of hyperplanes $\mathbf{a}_j^\intercal \mathbf{x} + \mathbf{b}_j$.



$$\max(\mathbf{a}_1^\intercal \mathbf{x}_1 + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_1 + b_K) = y_1$$
$$\max(\mathbf{a}_1^\intercal \mathbf{x}_2 + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_2 + b_K) = y_2$$
$$\ldots$$



Question 1: How to *optimally* estimate $(\mathbf{a}_j, b_j)_{j=1}^K$ from these equations?

Question 2: How to keep # of hyperplanes as *small* as possible?

This is not a linear problem!

This is not a linear problem!

- But in a suitable algebra, it can be written in a linear like form.

This is not a linear problem!

- But in a suitable algebra, it can be written in a linear like form.
- Max-Plus Algebra: "Linear Algebra, but $+$ becomes $\max$, and $\times$ becomes $+$".

This is not a linear problem!

- But in a suitable algebra, it can be written in a linear like form.
- Max-Plus Algebra: "Linear Algebra, but + becomes **max**, and $\times$ becomes +".

$$\underbrace{\begin{pmatrix} \mathbf{a}_1^\intercal \mathbf{x}_1 & \mathbf{a}_2^\intercal \mathbf{x}_1 & \dots & \mathbf{a}_K^\intercal \mathbf{x}_1 \\ . & . & . & . \\ . & . & . & . \\ \mathbf{a}_1^\intercal \mathbf{x}_m & \mathbf{a}_2^\intercal \mathbf{x}_m & \dots & \mathbf{a}_K^\intercal \mathbf{x}_m \end{pmatrix}}_{\mathrm{A}} \boxplus \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ b_K \end{pmatrix}}_{\mathrm{b}} = \underbrace{\begin{pmatrix} y_1 \\ . \\ . \\ y_m \end{pmatrix}}_{\mathrm{y}},$$

where $[\mathbf{C} \boxplus \mathbf{D}]_{ij} = \max_{k=1}^{n}(c_{ik} + d_{kj})$.

This is not a linear problem!

- But in a suitable algebra, it can be written in a linear like form.
- Max-Plus Algebra: "Linear Algebra, but + becomes **max**, and × becomes +".

$$\underbrace{\begin{pmatrix} \mathbf{a}_1^\intercal \mathbf{x}_1 & \mathbf{a}_2^\intercal \mathbf{x}_1 & \dots & \mathbf{a}_K^\intercal \mathbf{x}_1 \\ . & . & . & . \\ . & . & . & . \\ \mathbf{a}_1^\intercal \mathbf{x}_m & \mathbf{a}_2^\intercal \mathbf{x}_m & \dots & \mathbf{a}_K^\intercal \mathbf{x}_m \end{pmatrix}}_{\mathrm{A}} \boxplus \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ b_K \end{pmatrix}}_{\mathrm{b}} = \underbrace{\begin{pmatrix} y_1 \\ . \\ . \\ y_m \end{pmatrix}}_{\mathrm{y}},$$

where $[\mathbf{C} \boxplus \mathbf{D}]_{ij} = \max_{k=1}^{n}(c_{ik} + d_{kj})$.

- If we assume known slopes, problem reduces to solving a "linear" max-plus matrix equation $\mathbf{A} \boxplus \mathbf{b} = \mathbf{y}$.

This is not a linear problem!

- But in a suitable algebra, it can be written in a linear like form.
- Max-Plus Algebra: "Linear Algebra, but $+$ becomes **max**, and $\times$ becomes $+$".

$$\underbrace{\begin{pmatrix} \mathbf{a}_1^\intercal \mathbf{x}_1 & \mathbf{a}_2^\intercal \mathbf{x}_1 & \dots & \mathbf{a}_K^\intercal \mathbf{x}_1 \\ . & . & . & . \\ . & . & . & . \\ \mathbf{a}_1^\intercal \mathbf{x}_m & \mathbf{a}_2^\intercal \mathbf{x}_m & \dots & \mathbf{a}_K^\intercal \mathbf{x}_m \end{pmatrix}}_{\text{A}} \boxplus \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ b_K \end{pmatrix}}_{\text{b}} = \underbrace{\begin{pmatrix} y_1 \\ . \\ . \\ y_m \end{pmatrix}}_{\text{y}},$$

where $[\mathbf{C} \boxplus \mathbf{D}]_{ij} = \max_{k=1}^{n}(c_{ik} + d_{kj})$.

- If we assume known slopes, problem reduces to solving a "linear" max-plus matrix equation $\mathbf{A} \boxplus \mathbf{b} = \mathbf{y}$.

Question 1: How to *optimally* estimate $(\mathbf{a}_j, b_j)_{j=1}^{K}$ from these equations? ✓

Observation

If $b_j = -\infty$, then the maximum will never attain its value on the hyperplane $\mathbf{a}_j^\intercal \mathbf{x} + b_j \Rightarrow$ The whole region can be discarded.

Sparsity Motivation

Observation

If $b_j = -\infty$, then the maximum will never attain its value on the hyperplane $\mathbf{a}_j^\intercal \mathbf{x} + b_j \Rightarrow$ The whole region can be discarded.

$$\text{e.g. } \max(5x+3, 20x-\infty, x) = \max(5x+3, x), \ \forall x$$

Sparsity Motivation

Observation

If $b_j = -\infty$, then the maximum will never attain its value on the hyperplane $\mathbf{a}_j^\intercal \mathbf{x} + b_j \Rightarrow$ The whole region can be discarded.

$$\text{e.g. } \max(5x + 3, 20x - \infty, x) = \max(5x + 3, x), \ \forall x$$

Simpler & more robust approximation!

Observation

If $b_j = -\infty$, then the maximum will never attain its value on the hyperplane $\mathbf{a}_j^\intercal \mathbf{x} + b_j \Rightarrow$ The whole region can be discarded.

$$\text{e.g. } \max(5x + 3, 20x - \infty, x) = \max(5x + 3, x), \ \forall x$$

Simpler & more robust approximation!

Idea

Force as many b_j as possible to be $-\infty$!

Sparsity Motivation

Observation

If $b_j = -\infty$, then the maximum will never attain its value on the hyperplane $\mathbf{a}_j^T \mathbf{x} + b_j \Rightarrow$ The whole region can be discarded.

$$\text{e.g. } \max(5x + 3, 20x - \infty, x) = \max(5x + 3, x), \ \forall x$$

Simpler & more robust approximation!

Idea

Force as many b_j as possible to be $-\infty$!

Question 2: How to keep # of hyperplanes as *small* as possible? ✓

Step 1. Enforce **sparsity** constraints - Optimization problem:

$$\arg\min_{\mathrm{b}} |\mathrm{supp}(\mathbf{b})|$$

$$\text{s.t. } \mathbf{A} \boxplus \mathbf{b} = \mathbf{y}, \qquad \text{where } \mathrm{supp}(\mathbf{b}) \triangleq \{j \mid b_j \neq -\infty\}.$$

Step 1. Enforce **sparsity** constraints - Optimization problem:

$$\arg\min_{\mathbf{b}} |\mathrm{supp}(\mathbf{b})|$$
$$\text{s.t. } \mathbf{A} \boxplus \mathbf{b} = \mathbf{y}, \qquad \text{where } \mathrm{supp}(\mathbf{b}) \triangleq \{j \mid b_j \neq -\infty\}.$$

Step 2. Account for **noise** - Optimization problem:

$$\arg\min_{\mathbf{b}} |\mathrm{supp}(\mathbf{b})|$$
$$\text{s.t. } \mathrm{dist}(\mathbf{A} \boxplus \mathbf{b}, \mathbf{y}) \leq \epsilon.$$

Prior work & Contributions

- Prior work:

- Contributions:

Prior work & Contributions

- Prior work:
  - [Tsiamis & Maragos 2019] introduced sparsity in Max-plus algebra.

- Contributions:

Prior work & Contributions

- Prior work:
  - [Tsiamis & Maragos 2019] introduced sparsity in Max-plus algebra.
  - [Magnani & Boyd 2009, Kim et. al. 2010, Hannah & Dunson 2011] solve convex regression by alternating between partitioning the input and locally fitting affine functions.
- Contributions:

Prior work & Contributions

- Prior work:
  - [Tsiamis & Maragos 2019] introduced sparsity in Max-plus algebra.
  - [Magnani & Boyd 2009, Kim et. al. 2010, Hannah & Dunson 2011] solve convex regression by alternating between partitioning the input and locally fitting affine functions.
  - [Maragos & Theodosis 2020] initiated the study of the problem through the lens of Max-plus algebra $\Rightarrow$ complexity benefits.
- Contributions:

Prior work & Contributions

- Prior work:
  - [Tsiamis & Maragos 2019] introduced sparsity in Max-plus algebra.
  - [Magnani & Boyd 2009, Kim et. al. 2010, Hannah & Dunson 2011] solve convex regression by alternating between partitioning the input and locally fitting affine functions.
  - [Maragos & Theodosis 2020] initiated the study of the problem through the lens of Max-plus algebra $\Rightarrow$ complexity benefits.
- Contributions:
  - Algorithms for sparse approximate solutions to max-plus matrix equations for any ℓ_p norm - Submodularity properties of the optimization problems.

Prior work & Contributions

- Prior work:
  - [Tsiamis & Maragos 2019] introduced sparsity in Max-plus algebra.
  - [Magnani & Boyd 2009, Kim et. al. 2010, Hannah & Dunson 2011] solve convex regression by alternating between partitioning the input and locally fitting affine functions.
  - [Maragos & Theodosis 2020] initiated the study of the problem through the lens of Max-plus algebra $\Rightarrow$ complexity benefits.
- Contributions:
  - Algorithms for sparse approximate solutions to max-plus matrix equations for any ℓ_p norm - Submodularity properties of the optimization problems.
  - Application into the context of convex regression - robust approximations with an approximately *minimum* number of affine regions.

Definition (Sparsity)

We call a vector **x** *sparse* if it contains many $-\infty$ elements.

Definition (Sparsity)

We call a vector $\mathbf{x}$ *sparse* if it contains many $-\infty$ elements.

Definition (Support set)

The *support set* of a vector is the set of indices of its values that are not equal to $-\infty$, that is: $\mathrm{supp}(\mathbf{x}) = \{j \mid x_j \neq -\infty\}$.

e.g. $|\mathrm{supp}(1, 4, -\infty, -2, 0, 0)| = 5$

Definition (Sparsity)

We call a vector $\mathbf{x}$ *sparse* if it contains many $-\infty$ elements.

Definition (Support set)

The *support set* of a vector is the set of indices of its values that are not equal to $-\infty$, that is: $\mathrm{supp}(\mathbf{x}) = \{j \mid x_j \neq -\infty\}$.

e.g. $|\mathrm{supp}(1, 4, -\infty, -2, 0, 0)| = 5$

Theorem (Tsiamis & Maragos 2019)

Computing the sparsest solution of $\mathbf{A} \boxplus \mathbf{b} = \mathbf{y}$ is an NP-complete problem.

Sparse Approximate Solutions (1/3)

Problem formulation

$$\begin{aligned}
&\arg\min_{\mathrm{b}} |\mathrm{supp}(\mathbf{b})| \\
&\quad \text{s.t. } \|\mathbf{y} - \mathbf{A} \boxplus \mathbf{b}\|_p^p \leq \epsilon, \\
&\qquad \mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m.
\end{aligned}$$

Problem formulation

$$\begin{aligned} \arg\min_{\mathrm{b}} & |\mathrm{supp}(\mathbf{b})| \\ \text{s.t. } & \|\mathbf{y} - \mathbf{A} \boxplus \mathbf{b}\|_p^p \leq \epsilon, \\ & \mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m. \end{aligned}$$

Notes

- We restrict the ℓ_p, $p < \infty$, error to be small
- We add an extra constraint $\mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}$: Technical reasons & enforce approximation from below!

Problem formulation

$$\begin{aligned} &\arg\min_{\mathbf{b}} |\mathrm{supp}(\mathbf{b})| \\ &\text{s.t. } \|\mathbf{y} - \mathbf{A} \boxplus \mathbf{b}\|_p^p \leq \epsilon, \\ &\mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m. \end{aligned} \tag{1}$$

Problem formulation

$$\begin{aligned} &\arg\min_{\mathrm{b}} |\mathrm{supp}(\mathbf{b})| \\ &\text{s.t. } \|\mathbf{y} - \mathbf{A} \boxplus \mathbf{b}\|_p^p \leq \epsilon, \\ &\mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m. \end{aligned} \tag{1}$$

Theorem

Problem (1) *can be approximately solved in* $\mathcal{O}(nm + n^2)$ *time with a greedy algorithm.*

Problem formulation

$$
\begin{aligned}
&\arg\min_{\mathrm{b}} |\mathrm{supp}(\mathbf{b})| \\
&\text{s.t. } \|\mathbf{y} - \mathbf{A} \boxplus \mathbf{b}\|_p^p \leq \epsilon, \\
&\mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m.
\end{aligned}
\tag{1}
$$

Theorem

Problem (1) *can be approximately solved in* $\mathcal{O}(nm + n^2)$ *time with a greedy algorithm.*

Technical details

Tools from Submodular Optimization ⇒ approximation ratio guarantees.

Sparse Approximate Solutions (3/3)

Approximating noise corrupted data from below might be problematic! Is it possible to remove the $\mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}$ contraint?

Sparse Approximate Solutions (3/3)

Approximating noise corrupted data from below might be problematic! Is it possible to remove the $\mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}$ contraint?

$$\begin{aligned} &\arg\min_{\mathrm{b}} |\mathrm{supp}(\mathbf{b})| \\ &\text{s.t. } \|\mathbf{y} - \mathbf{A} \boxplus \mathbf{b}\|_\infty \leq \epsilon, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m. \end{aligned} \tag{2}$$

Sparse Approximate Solutions (3/3)

Approximating noise corrupted data from below might be problematic! Is it possible to remove the $\mathbf{A} \boxplus \mathbf{b} \leq \mathbf{y}$ contraint?

$$\begin{aligned} &\arg\min_{\mathrm{b}} |\mathrm{supp}(\mathbf{b})| \\ &\quad \text{s.t. } \|\mathbf{y} - \mathbf{A} \boxplus \mathbf{b}\|_\infty \leq \epsilon, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m. \end{aligned} \tag{2}$$

Proposition

We can find a **locally optimal** *solution of Problem* (2) *by solving Problem* (1).

Application to Multivariate Convex Regression (1/2)

Input: Data $(\mathbf{x}_i, y_i) \in \mathbb{R}^{n+1}$, $i = 1, 2, \ldots, m$.

Application to Multivariate Convex Regression (1/2)

Input: Data $(\mathbf{x}_i, y_i) \in \mathbb{R}^{n+1}$, $i = 1, 2, \ldots, m$.
Model: $\max(\mathbf{a}_1^\intercal \mathbf{x}_i + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_i + b_K) = y_i, \; i = 1, 2, \ldots, m.$

Application to Multivariate Convex Regression (1/2)

Input: Data $(\mathbf{x}_i, y_i) \in \mathbb{R}^{n+1}$, $i = 1, 2, \ldots, m$.
Model: $\max(\mathbf{a}_1^\intercal \mathbf{x}_i + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_i + b_K) = y_i$, $i = 1, 2, \ldots, m$.

Step 1: Estimate K slopes $\mathbf{a}_k$:

- Fixed values from an n-dimensional interval, or
- Numerical gradients of data.

Application to Multivariate Convex Regression (1/2)

Input: Data $(\mathbf{x}_i, y_i) \in \mathbb{R}^{n+1}$, $i = 1, 2, \ldots, m$.
Model: $\max(\mathbf{a}_1^\intercal \mathbf{x}_i + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_i + b_K) = y_i,\ \ i = 1, 2, \ldots, m$.

Step 1: Estimate K slopes $\mathbf{a}_k$:

- Fixed values from an n-dimensional interval, or
- Numerical gradients of data.

Step 2: Solve Problem (1) (method called *Sparse Greatest Lower Estimate - SGLE*) or Problem (2) (called *Sparse Minimum Max Absolute Error - SMMAE*), and calculate intercepts b_k.

Application to Multivariate Convex Regression (1/2)

Input: Data $(\mathbf{x}_i, y_i) \in \mathbb{R}^{n+1}$, $i = 1, 2, \ldots, m$.
Model: $\max(\mathbf{a}_1^\intercal \mathbf{x}_i + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_i + b_K) = y_i, \;\; i = 1, 2, \ldots, m.$

Step 1: Estimate K slopes $\mathbf{a}_k$:

- Fixed values from an n-dimensional interval, or
- Numerical gradients of data.

Step 2: Solve Problem (1) (method called *Sparse Greatest Lower Estimate - SGLE*) or Problem (2) (called *Sparse Minimum Max Absolute Error - SMMAE*), and calculate intercepts b_k.

Complexity: $\mathcal{O}(K^2 + K(n+1)m)$ or $\mathcal{O}(K^2 + K(n+2)m)$, respectively.

Application to Multivariate Convex Regression (1/2)

Input: Data $(\mathbf{x}_i, y_i) \in \mathbb{R}^{n+1}$, $i = 1, 2, \ldots, m$.
Model: $\max(\mathbf{a}_1^\intercal \mathbf{x}_i + b_1, \ldots, \mathbf{a}_K^\intercal \mathbf{x}_i + b_K) = y_i, \; i = 1, 2, \ldots, m$.

Step 1: Estimate K slopes $\mathbf{a}_k$:

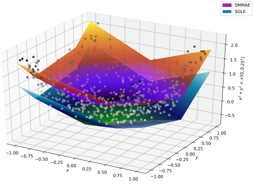
- Fixed values from an n-dimensional interval, or
- Numerical gradients of data.

Step 2: Solve Problem (1) (method called *Sparse Greatest Lower Estimate - SGLE*) or Problem (2) (called *Sparse Minimum Max Absolute Error - SMMAE*), and calculate intercepts b_k.

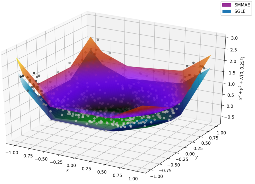
Complexity: $\mathcal{O}(K^2 + K(n+1)m)$ or $\mathcal{O}(K^2 + K(n+2)m)$, respectively.

Output: A PWL convex approximation of the data with the approximately *minimum* number of affine regions needed for achieving the desired level of data fidelity.

Data from noisy paraboloid $z = x^2 + y^2 + \mathcal{N}(0,1)$.



(a) Approximation with 5 affine regions.



(b) Approximation with 16 affine regions.

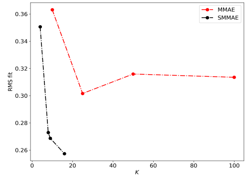


Figure: Comparison of SMMAE method (black line) with [Maragos & Theodosis 2020] (red line).

Conclusion & Future work

- Fundamental problem of convex regression through the lens of Max-plus algebra.

Conclusion & Future work

- Fundamental problem of convex regression through the lens of Max-plus algebra.
- Theoretical progress in algorithms for sparsity in this nonlinear space.

Conclusion & Future work

- Fundamental problem of convex regression through the lens of Max-plus algebra.
- Theoretical progress in algorithms for sparsity in this nonlinear space.
- Application into the problem of convex regression yields more robust estimators.

Conclusion & Future work

- Fundamental problem of convex regression through the lens of Max-plus algebra.
- Theoretical progress in algorithms for sparsity in this nonlinear space.
- Application into the problem of convex regression yields more robust estimators.

Future work: Statistical properties of the estimators & Sparsity in more abstract nonlinear spaces with an underlying lattice structure.

Conclusion & Future work

- Fundamental problem of convex regression through the lens of Max-plus algebra.
- Theoretical progress in algorithms for sparsity in this nonlinear space.
- Application into the problem of convex regression yields more robust estimators.

Future work: Statistical properties of the estimators & Sparsity in more abstract nonlinear spaces with an underlying lattice structure.

Thank you for your attention!

References

L. A. Hannah and D. B. Dunson. "Multivariate convex regression with adaptive partitioning". In: (2011). arXiv: 1105.1924.

J. Kim, L. Vandenberghe, and C.K.K. Yang. "Convex Piecewise-Linear Modeling Method for Circuit Optimization via Geometric Programming". In: *IEEE Trans. Computer-Aided Design of Integr. Circuits Syst.* 29.11 (Nov. 2010), pp. 1823–1827.

A. Magnani and S. P. Boyd. "Convex piecewise-linear fitting". In: *Optim. Eng.* 10 (2009), pp. 1–17.

P. Maragos. "Morphological Systems: Slope Transforms and Max-Min Difference and Differential Equations". In: *Signal Processing* 38.1 (July 1994), pp. 57–77.

P. Maragos and E. Theodosis. "Multivariate Tropical Regression and Piecewise-Linear Surface Fitting". In: *Proc. IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. 2020, pp. 3822–3826.

R. T. Rockafellar. *Convex Analysis*. Princeton Univ. Press, 1970.

A. Tsiamis and P. Maragos. "Sparsity in Max-plus Algebra". In: *Discrete Events Dynamic Systems* 29 (2019), pp. 163–189.