

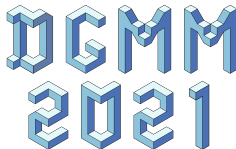
Sparse Approximate Solutions to Max-Plus Equations

Nikos Tsilivis ¹, Anastasios Tsiamis ², Petros Maragos ¹

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

²ESE Department, SEAS, University of Pennsylvania, USA

May 27, 2021



- Introduction
- Theory
- Application to Morphological neural networks pruning
- Conclusion

Motivation:

- Inverse problems: We observe a vector \mathbf{b} as linear measurements of an unknown quantity \mathbf{x} through a system \mathbf{A} . We want to recover the initial information.
Problem: Possibly infinite candidates that explain the data.

Motivation:

- Inverse problems: We observe a vector \mathbf{b} as linear measurements of an unknown quantity \mathbf{x} through a system \mathbf{A} . We want to recover the initial information.
Problem: Possibly infinite candidates that explain the data.
- Efficient representations: Consider a signal $\in \mathbb{R}^m$. Storing it with only k values, $k \ll m$? Idea: the signal may be really simple computed in a different basis! (e.g. DFT of cosines: only 2 non zero values). How to find the suitable basis? and how to compute the simple signal in this basis?

Sparsity in Linear Algebra

Motivation:

- Inverse problems: We observe a vector \mathbf{b} as linear measurements of an unknown quantity \mathbf{x} through a system \mathbf{A} . We want to recover the initial information. Problem: Possibly infinite candidates that explain the data.
- Efficient representations: Consider a signal $\in \mathbb{R}^m$. Storing it with only k values, $k \ll m$? Idea: the signal may be really simple computed in a different basis! (e.g. DFT of cosines: only 2 non zero values). How to find the suitable basis? and how to compute the simple signal in this basis?

Core of the problem:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ b_m \end{pmatrix}$$

Sparsity in Linear Algebra

Motivation:

- Inverse problems: We observe a vector \mathbf{b} as linear measurements of an unknown quantity \mathbf{x} through a system \mathbf{A} . We want to recover the initial information. Problem: Possibly infinite candidates that explain the data.
- Efficient representations: Consider a signal $\in \mathbb{R}^m$. Storing it with only k values, $k \ll m$? Idea: the signal may be really simple computed in a different basis! (e.g. DFT of cosines: only 2 non zero values). How to find the suitable basis? and how to compute the simple signal in this basis?

Core of the problem:

$$\begin{pmatrix} a_{12} & a_{1n} \\ a_{22} & a_{2n} \\ \cdot & \cdot \\ a_{m2} & a_{mn} \end{pmatrix} \begin{pmatrix} x_2 \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ b_m \end{pmatrix}$$

What is the solution with the least **non zero** elements? The most *sparse*?

In a variety of problems **non linear** representations arise (with maximum and additions):

In a variety of problems **non linear** representations arise (with maximum and additions):

Morphological signal and
image analysis.

In a variety of problems **non linear** representations arise (with maximum and additions):

Morphological signal and
image analysis.

Discrete event systems.

In a variety of problems **non linear** representations arise (with maximum and additions):

Morphological signal and
image analysis.

Discrete event systems.

Optimal control and
Dynamic programming.

Max-plus Algebra

In a variety of problems **non linear** representations arise (with maximum and additions):

Morphological signal and
image analysis.

Discrete event systems.

Optimal control and
Dynamic programming.

$$b_1 = \max(a_{11} + x_1, a_{12} + x_2, \dots, a_{1n} + x_n),$$

$$b_2 = \max(a_{21} + x_1, a_{22} + x_2, \dots, a_{2n} + x_n)$$

.

.

In a variety of problems **non linear** representations arise (with maximum and additions):

Morphological signal and
image analysis.

Discrete event systems.

Optimal control and
Dynamic programming.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix} \boxplus \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ b_m \end{pmatrix}$$

In a variety of problems **non linear** representations arise (with maximum and additions):

Morphological signal and image analysis.

Discrete event systems.

Optimal control and Dynamic programming.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix} \boxplus \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ b_m \end{pmatrix}$$

What is the solution with the least number of **non** $-\infty$ elements? The most sparse solution?

Contributions & Related Work

- Contributions:

- Related Work:

- Contributions:
 - Sparse approximate solutions as a discrete optimization problem.
- Related Work:

- Contributions:
 - Sparse approximate solutions as a discrete optimization problem.
 - Analysis of the *submodular* properties of the corresponding problems.
- Related Work:

- Contributions:
 - Sparse approximate solutions as a discrete optimization problem.
 - Analysis of the *submodular* properties of the corresponding problems.
 - Sparsity might be relevant to modern machine learning applications.
- Related Work:

- Contributions:
 - Sparse approximate solutions as a discrete optimization problem.
 - Analysis of the *submodular* properties of the corresponding problems.
 - Sparsity might be relevant to modern machine learning applications.
- Related Work:
 - Connection between max-plus equations & discrete optimization (min set cover) [Butkovič 2013].

- Contributions:
 - Sparse approximate solutions as a discrete optimization problem.
 - Analysis of the *submodular* properties of the corresponding problems.
 - Sparsity might be relevant to modern machine learning applications.
- Related Work:
 - Connection between max-plus equations & discrete optimization (min set cover) [Butkovič 2013].
 - [Gaubert et. al. 2011] Pruning in optimal control - related to sparse ℓ_1 approximate solutions.

- Contributions:
 - Sparse approximate solutions as a discrete optimization problem.
 - Analysis of the *submodular* properties of the corresponding problems.
 - Sparsity might be relevant to modern machine learning applications.
- Related Work:
 - Connection between max-plus equations & discrete optimization (min set cover) [Butkovič 2013].
 - [Gaubert et. al. 2011] Pruning in optimal control - related to sparse ℓ_1 approximate solutions.
 - [Tsiamis & Maragos 2019] introduced the concept of sparsity in max-plus algebra.

- Values from $\mathbb{R} \cup \{-\infty\}$.
- Max-plus and min-plus products:

$$[\mathbf{A} \boxplus \mathbf{x}]_i \triangleq \bigvee_{k=1}^n a_{ik} + x_k, \quad [\mathbf{A} \boxminus \mathbf{x}]_i \triangleq \bigwedge_{k=1}^n a_{ik} + x_k$$

- A max-plus equation $\mathbf{A} \boxplus \mathbf{x} = \mathbf{b}$ has a solution iff $\hat{\mathbf{x}} = -\mathbf{A}^T \boxminus \mathbf{b}$ (*principal solution*) satisfies it.
- $\mathbf{A} \boxplus \hat{\mathbf{x}} \leq \mathbf{b}$.

Definition (Submodular)

A set function $f : 2^U \rightarrow \mathbb{R}$ is called *submodular* if $\forall A \subseteq B \subseteq U, k \notin B$ holds:

$$f(A \cup \{k\}) - f(A) \geq f(B \cup \{k\}) - f(B).$$

Submodularity preliminaries (1/2)

Definition (Submodular)

A set function $f : 2^U \rightarrow \mathbb{R}$ is called *submodular* if $\forall A \subseteq B \subseteq U, k \notin B$ holds:

$$f(A \cup \{k\}) - f(A) \geq f(B \cup \{k\}) - f(B).$$

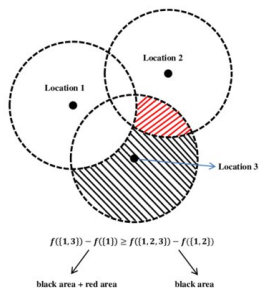


Figure: [Liu et al. 2019]

Submodularity preliminaries (2/2)

Generalization of submodularity

Definition (Submodularity ratio of an increasing, non-negative function [Das & Kempe 2018])

$$\gamma_{U,k}(f) \triangleq \min_{L \subseteq U, S: |S| \leq k, S \cap L = \emptyset} \frac{\sum_{x \in S} f(L \cup \{x\}) - f(L)}{f(L \cup S) - f(L)}$$

Submodularity preliminaries (2/2)

Generalization of submodularity

Definition (Submodularity ratio of an increasing, non-negative function [Das & Kempe 2018])

$$\gamma_{U,k}(f) \triangleq \min_{L \subseteq U, S: |S| \leq k, S \cap L = \emptyset} \frac{\sum_{x \in S} f(L \cup \{x\}) - f(L)}{f(L \cup S) - f(L)}$$

Proposition

An increasing function $f : 2^U \rightarrow \mathbb{R}$ is submodular if and only if $\gamma_{U,k}(f) \geq 1, \forall U, k$.

Definition (Sparsity)

We call a vector \mathbf{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: $\text{supp}(\mathbf{x}) = \{j \mid x_j \neq -\infty\}$.

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

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: $\text{supp}(\mathbf{x}) = \{j \mid x_j \neq -\infty\}$.

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

Theorem (Tsiamis & Maragos 2019)

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

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: $\text{supp}(\mathbf{x}) = \{j \mid x_j \neq -\infty\}$.

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

Theorem (Tsiamis & Maragos 2019)

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

essentially: Minimum Set Cover.

Problem formulation

$$\arg \min_{\mathbf{x}} |\text{supp}(\mathbf{x})|$$

$$\text{s.t. } \|\mathbf{b} - \mathbf{A} \boxplus \mathbf{x}\|_p^p \leq \epsilon,$$

$$\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m.$$

Problem formulation

$$\arg \min_{\mathbf{x}} |\text{supp}(\mathbf{x})|$$

$$\text{s.t. } \|\mathbf{b} - \mathbf{A} \boxplus \mathbf{x}\|_p^p \leq \epsilon,$$

$$\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^m.$$

Notes

- We restrict the ℓ_p , $p < \infty$, error to be small.
- We add an extra constraint $\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{b}$.
- Observe that for $\epsilon = 0$ reduces to an \mathcal{NP} -complete problem.
- The case $p = 1$ was examined in [Tsiamis & Maragos 2019].

Problem formulation

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

Problem formulation

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

Proposition (Informal)

We can fix the values of \mathbf{x} to be equal to those of $\hat{\mathbf{x}}$, and search only over the support set T .

Problem formulation

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

Proposition (Informal)

We can fix the values of \mathbf{x} to be equal to those of $\hat{\mathbf{x}}$, and search only over the support set T .

New Problem formulation

$$\begin{aligned} & \arg \min_T |T| \\ & \text{s.t. } E_p(T) \leq \epsilon \end{aligned}$$

Theorem

Error function E_p is decreasing and supermodular.

Theorem

Error function E_p is decreasing and supermodular.

Notes

- Proof leverages the submodularity ratio which clarifies the analysis.
- Problem becomes: Cardinality minimization problem subject to a supermodular equality constraint \Rightarrow Fast greedy algorithm works!

Greedy Algorithm

Algorithm 1: Approximate solution of problem (1)

Input: \mathbf{A} , \mathbf{b}

Compute $\hat{\mathbf{x}} = (-\mathbf{A})^\top \boxplus \mathbf{b}$

if $E_p(J) > \epsilon$ **then**

 | **return** Infeasible

Set $T_0 = \emptyset$, $k = 0$

while $E_p(T_k) > \epsilon$ **do**

 | $j = \arg \min_{s \in J \setminus T_k} E_p(T_k \cup \{s\})$

 | $T_{k+1} = T_k \cup \{j\}$

 | $k = k + 1$

end

$x_j = \hat{x}_j$, $j \in T_k$ and $x_j = -\infty$, otherwise

return \mathbf{x} , T_k

Time complexity: $\mathcal{O}(nm + n^2)$

Approximation ratio: $\mathcal{O}(\log(m\Delta^p))$, $\Delta = \bigvee_{i,j} (b_i - A_{ij} - \hat{x}_j)$.

Sparse ℓ_∞ Approximate Solutions (1/2)

Searching only for approximate sub-solutions is restrictive. Can we overcome this?

We shift our attention to the ℓ_∞ norm.

Sparse ℓ_∞ Approximate Solutions (1/2)

Searching only for approximate sub-solutions is restrictive. Can we overcome this?

We shift our attention to the ℓ_∞ norm.

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

Sparse ℓ_∞ Approximate Solutions (1/2)

Searching only for approximate sub-solutions is restrictive. Can we overcome this?

We shift our attention to the ℓ_∞ norm.

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

Proposition (Informal)

Problem (2) can also be written as a set-search problem.

Sparse ℓ_∞ Approximate Solutions (1/2)

Searching only for approximate sub-solutions is restrictive. Can we overcome this?

We shift our attention to the ℓ_∞ norm.

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

Proposition (Informal)

Problem (2) can also be written as a set-search problem.

New Problem formulation

$$\begin{aligned} \arg \min_T |T| \\ \text{s.t. } E_\infty(T) \leq \epsilon \end{aligned}$$

Sparse ℓ_∞ Approximate Solutions (1/2)

Searching only for approximate sub-solutions is restrictive. Can we overcome this?

We shift our attention to the ℓ_∞ norm.

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

Proposition (Informal)

Problem (2) can also be written as a set-search problem.

New Problem formulation

$$\begin{aligned} \arg \min_T |T| \\ \text{s.t. } E_\infty(T) \leq \epsilon \end{aligned}$$

Unfortunately ..

Sparse ℓ_∞ Approximate Solutions (1/2)

E_∞ is not even approximately supermodular ($\gamma(E_\infty) = 0$) \Rightarrow a greedy solution might have arbitrarily big support set!

Sparse ℓ_∞ Approximate Solutions (1/2)

E_∞ is not even approximately supermodular ($\gamma(E_\infty) = 0$) \Rightarrow a greedy solution might have arbitrarily big support set!

Proposal: Solve the ℓ_p problem greedily and add to the solution the half of its ℓ_∞ error.

Sparse ℓ_∞ Approximate Solutions (1/2)

E_∞ is not even approximately supermodular ($\gamma(E_\infty) = 0$) \Rightarrow a greedy solution might have arbitrarily big support set!

Proposal: Solve the ℓ_p problem greedily and add to the solution the half of its ℓ_∞ error.

Proposition

If $\mathbf{x}_{p,\epsilon}$ is a solution of ℓ_p Problem (1), then $\mathbf{x}^* = \mathbf{x}_{p,\epsilon} + \frac{\|\mathbf{b} - \mathbf{A}\mathbf{x}_{p,\epsilon}\|_\infty}{2}$ has the *smallest* ℓ_∞ error over **all** sparse vectors with the same support set T .

Sparse ℓ_∞ Approximate Solutions (1/2)

E_∞ is not even approximately supermodular ($\gamma(E_\infty) = 0$) \Rightarrow a greedy solution might have arbitrarily big support set!

Proposal: Solve the ℓ_p problem greedily and add to the solution the half of its ℓ_∞ error.

Proposition

If $\mathbf{x}_{p,\epsilon}$ is a solution of ℓ_p Problem (1), then $\mathbf{x}^* = \mathbf{x}_{p,\epsilon} + \frac{\|\mathbf{b} - \mathbf{A}\mathbf{x}_{p,\epsilon}\|_\infty}{2}$ has the *smallest* ℓ_∞ error over **all** sparse vectors with the same support set T .

Computational overhead: $\mathcal{O}(m|T|)$

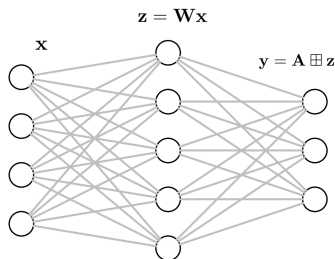
- Neural networks with layers that perform Morphological Operations, such as dilations and erosions [Ritter & Sussner 1996].

Morphological Neural Networks

- Neural networks with layers that perform Morphological Operations, such as dilations and erosions [Ritter & Sussner 1996].
- Recent studies have highlighted their ability to be pruned effectively and produce interpretable models [Charisopoulos & Maragos 2017, Zhang et al. 2019].

Morphological Neural Networks

- Neural networks with layers that perform Morphological Operations, such as dilations and erosions [Ritter & Sussner 1996].
- Recent studies have highlighted their ability to be pruned effectively and produce interpretable models [Charisopoulos & Maragos 2017, Zhang et al. 2019].



Neuron pruning as a sparsity problem

Consider a simple two layer network that performs a linear transformation followed by a dilation (*max-plus block* [Zhang et al. 2019]):

$$\mathbf{z} = \mathbf{W}\mathbf{x},$$

$$\mathbf{y} = \mathbf{A} \boxplus \mathbf{z}.$$

Neuron pruning as a sparsity problem

Consider a simple two layer network that performs a linear transformation followed by a dilation (*max-plus block* [Zhang et al. 2019]):

$$\begin{aligned}\mathbf{z} &= \mathbf{W}\mathbf{x}, \\ \mathbf{y} &= \mathbf{A} \boxplus \mathbf{z}.\end{aligned}$$

After training: fix \mathbf{y} , \mathbf{A} and search for a sparse solution over \mathbf{z} . Then, keep neurons that correspond to finite values.

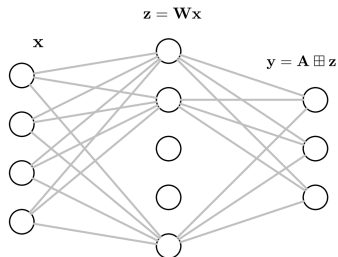
Neuron pruning as a sparsity problem

Consider a simple two layer network that performs a linear transformation followed by a dilation (*max-plus block* [Zhang et al. 2019]):

$$\mathbf{z} = \mathbf{W}\mathbf{x},$$

$$\mathbf{y} = \mathbf{A} \boxplus \mathbf{z}.$$

After training: fix \mathbf{y} , \mathbf{A} and search for a sparse solution over \mathbf{z} . Then, keep neurons that correspond to finite values.



Experiments on MNIST & FashionMNIST

- 2 networks of 64 and 128 neurons, trained for 20 epochs, with SGD.
- We are able to find the 10 most important neurons automatically and prune the rest of them (recording same accuracy).

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

Table: Test set accuracy before and after pruning.

- Sparsity in max-plus algebra.

Conclusion & Future work

- Sparsity in max-plus algebra.
- Analysis of the submodular structure of the problems \Rightarrow also relevant to **other** problems in max-plus algebra (unconstrained optimal ℓ_2 approximations).

Conclusion & Future work

- Sparsity in max-plus algebra.
- Analysis of the submodular structure of the problems \Rightarrow also relevant to **other** problems in max-plus algebra (unconstrained optimal ℓ_2 approximations).
- Sparse framework to assist advances in training and optimizing morphological models.

- Sparsity in max-plus algebra.
- Analysis of the submodular structure of the problems \Rightarrow also relevant to **other** problems in max-plus algebra (unconstrained optimal ℓ_2 approximations).
- Sparse framework to assist advances in training and optimizing morphological models.

Future work: A complete sparse representation theory in complete lattices with efficient algorithms for sparse approximate solutions to max-plus, max-min, smooth idempotent spaces and more.

- Sparsity in max-plus algebra.
- Analysis of the submodular structure of the problems \Rightarrow also relevant to **other** problems in max-plus algebra (unconstrained optimal ℓ_2 approximations).
- Sparse framework to assist advances in training and optimizing morphological models.

Future work: A complete sparse representation theory in complete lattices with efficient algorithms for sparse approximate solutions to max-plus, max-min, smooth idempotent spaces and more.

Thank you for your attention!