TROPICAL LINEAR REGRESSION AND LOW-RANK APPROXIMATION — A FIRST STEP IN TROPICAL DATA ANALYSIS

YANG QI

Abstract. This note is partially based on the talk given by the author at Institute of Mathematics Polish Academy of Sciences (IMPAN) in the AGATES workshop "AGATES: Tensors from the physics viewpoint". Here we only focus on the part of tropical linear regression. The author would like to thank Jarosław Buczyński, Weronika Buczyńska, Francesco Galuppi, and Joachim Jelisiejew for organizing the thematic semester "Tensors: geometry, complexity and quantum entanglement" at IMPAN and University of Warsaw, and thank Wojciech Bruzda, Jarosław Buczyński, and Karol Życzkowski for organizing the workshop "Tensors from the physics viewpoint".

This work is partially supported by the Thematic Research Programme "Tensors: geometry, complexity and quantum entanglement", University of Warsaw, Excellence Initiative – Research University and the Simons Foundation Award No. 663281 granted to the Institute of Mathematics of the Polish Academy of Sciences for the years 2021–2023.

1. Introduction

1.1. **The tropical linear regression problem.** The max-plus semifield $\mathbb{R}_{\max}$ is the set of real numbers, completed by $-\infty$ and equipped with the addition $(a, b) \mapsto a \oplus b := \max(a, b)$ and the multiplication $(a, b) \mapsto a \odot b := a + b$. A *tropical hyperplane* in the n-dimensional tropical vector space $\mathbb{R}^n_{\max}$ is a set of vectors of the form

$$\mathcal{H}_a = \{x \in \mathbb{R}^n_{\max} \mid \max_{1\le i\le n}(a_i + x_i) \text{ is achieved at least twice}\} \ . \tag{1}$$

Such a hyperplane is parametrized by the vector $a = (a_1, \dots, a_n) \in \mathbb{R}^n_{\max}$ so that not every a_i is $-\infty$. Tropical hyperplanes are among the most basic objects in tropical geometry. They are images by the valuation of hyperplanes over non-archimedean fields, and so, they are the simplest examples of tropical linear spaces and tropical hypersurfaces.

We denote by $\perp$ the vector in $\mathbb{R}^n_{\max}$ whose entries are all $-\infty$, and let $\mathbb{P}(\mathbb{R}^n_{\max})$ be the *tropical projective space*, i.e., the quotient $\mathbb{R}^n_{\max}\backslash\{\perp\}$ by the equivalence relation $\sim$ which identifies tropically proportional vectors. We shall abuse notation and denote by the same symbol a vector and its equivalence class in $\mathbb{P}(\mathbb{R}^n_{\max})$. The Hilbert projective metric is defined by

$$\|x\|_H := \max_{i\in[n]} x_i - \min_{i\in[n]} x_i$$

for any $x \in \mathbb{P}(\mathbb{R}^n_{\max})$. In this note, we address the following tropical analogue of the linear regression problem: Given a set of finitely many points $\mathcal{V} \subset \mathbb{R}^n_{\max}$, we look for a best fitting tropical hyperplane of these points. More precisely, the *tropical linear regression problem* is formulated as

$$\min_{a\in\mathbb{P}(\mathbb{R}^n_{\max})} \max_{v\in\mathcal{V}} \min_{x\in\mathcal{H}_a\cap\mathbb{R}^n} \|v - x\|_H. \tag{2}$$

The *Hilbert's projective distance* for vectors $x, y \in \mathbb{R}^n_{\max}$, where at least one of them is not $\perp$, is given by $\|x - y\|_H$, and we set $d(\perp, \perp) := 0$. The *support* of a vector $x \in \mathbb{R}^n_{\max}$ is defined by $\operatorname{supp} x := \{i \in [n] \mid x_i \neq -\infty\}$. Each subset $I \subseteq [n]$ yields a *part* P_I of $\mathbb{R}^n_{\max}$, consisting of vectors with support I. Observe that $d(x, y)$ is finite if and only if x and y belong to the same part P_I.

Moreover, if $I \neq \emptyset$,

$$d(x,y) = \max_{i\in I}(x_i - y_i) - \min_{i\in I}(x_i - y_i) \ .$$

1.2. **Tropical cones.** A subset $\mathcal{C}$ of $\mathbb{R}^n_{\max}$ is a *tropical (convex) cone* or equivalently a *tropical submodule* of $\mathbb{R}^n_{\max}$ if it satisfies that $x, y \in \mathcal{C}$ and $\lambda \in \mathbb{R}_{\max}$ implies $\lambda + x \in \mathcal{C}$ and $\max(x,y) \in \mathcal{C}$. For any given subset $\mathcal{V}$ of $\mathbb{R}^n_{\max}$, we denote by $\mathrm{Sp}(\mathcal{V})$ the tropical submodule of $\mathbb{R}^n_{\max}$ generated by $\mathcal{V}$. A tropical polyhedral cone $\mathcal{C}$ is a tropical cone which is finitely generated. Equivalently, a tropical polyhedral cone can be defined externally by

$$\max_{j\in[n]}(A_{ij} + x_j) \leq \max_{j\in[n]}(B_{ij} + x_j), \qquad i \in [m]. \tag{3}$$

A scalar μ is a *tropical eigenvalue* of a matrix $M \in \mathbb{R}^{n\times n}_{\max}$ if there exists a vector $u \in \mathbb{R}^n_{\max}$, so that $u \neq \perp$ and $M \odot u = \mu \odot u$ in the tropical sense. The eigenvalue is known to be unique when the digraph of M is strongly connected, then it coincides with the maximum weight-to-length ratio of the circuits of the digraph of M. We denote it by $\lambda(M)$.

1.3. **Mean payoff games.** We consider zero-sum deterministic games, with perfect information, defined as follows. There are two players, "Max" and "Min" (the maximizer and the minimizer), who will move a token on a weighted digraph. We assume this digraph is finite and bipartite: the node set is the disjoint union of two non-empty sets $S^{\max}$ and $S^{\min}$, and the arc set $\mathcal{A}$ is included in $(S^{\max} \times S^{\min}) \cup (S^{\min} \times S^{\max})$. The set of states of the game is the set of nodes of the digraph. We associate a real weight w_{rs} to each arc (r,s).

The two players alternate their actions. When the token is in node $i \in S^{\min}$, Player Min must choose an arc (i,j) in the digraph, meaning he moves the token to node j, and pays w_{ij} to player Max. When Player Min has no possible action, that is, when there are no arcs of the form (i,j) in the digraph, the game terminates, and Player Max receives $+\infty$. Similarly, when the token is in node $j \in S^{\max}$, Player Max must choose an arc (j,i) in the digraph. Then he moves the token from node j to node i, and receives w_{ji} from Player Min. When Player Max has no possible action, that is there are no arcs of the form (j,i) in the digraph, the game terminates, and Player Max receives $-\infty$. If the game does not terminate before time k, the *history* of the game is described by a sequence of nodes $\bar{\imath} = i_0, j_1, i_1, \ldots, j_k, i_k$, belonging alternatively to $S^{\min}$ and $S^{\max}$, and the total payment received by Player Max is given by

$$R^k_{\bar{\imath}} = w_{i_0 j_1} + w_{j_1 i_1} + w_{i_1 j_2} + \cdots + w_{j_k i_k}.$$

If the game terminates by k, we set $R^k_{\bar{\imath}} = \pm\infty$ depending on the player who had no available action. The following assumption requires Player Min to have at least one available action in every state:

Assumption 1. For all $i \in S^{\min}$, there exists $j \in S^{\max}$ such that (i,j) is an arc of the digraph of the zero-sum deterministic game.

In this way, we always have $R^k_{\bar{\imath}} \in \mathbb{R} \cup \{-\infty\}$. We shall also consider the dual assumption.

Assumption 2. For all $j \in S^{\max}$, there exists $i \in S^{\min}$ such that (j,i) is an arc of the digraph of the zero-sum deterministic game.

A strategy of a player is a map which associates to the history of the game an action of this player. Assuming that Player Min plays according to strategy σ, and that Player Max plays according to strategy τ, we shall write $R^k_{\bar{\imath}} = R^k_{\bar{\imath}}(\sigma,\tau)$ to indicate the dependence on these strategies. It follows that the game in horizon k starting from node $\bar{\imath}$ has a value $v^k_{\bar{\imath}}$ and that Players Min and Max have optimal strategies σ^* and τ^* respectively, i.e.,

$$R^k_{\bar{\imath}}(\sigma,\tau^*) \leq v^k_{\bar{\imath}} = R^k_{\bar{\imath}}(\sigma^*,\tau^*) \leq R^k_{\bar{\imath}}(\sigma^*,\tau)$$

for all strategies σ, τ. Moreover, the *value vector* $v^k := (v_i^k)_{i \in S^{\min}}$ is determined by

$$v^k = T(v^{k-1}), \qquad v^0 = 0$$

where $T : \mathbb{R}^n_{\max} \to \mathbb{R}^n_{\max}$ is the *Shapley operator*, defined, for $i \in S^{\min}$, by

$$T_i(x) = \min_{j,\ (i,j) \in \mathcal{A}} (w_{ij} + \max_{l,\ (j,l) \in \mathcal{A}} (w_{jl} + x_l)). \tag{4}$$

Assumption 1 and Assumption 2 guarantee that the above minimum and maximum are never taken over an empty set, i.e., when both assumptions hold, T sends $\mathbb{R}^n$ to $\mathbb{R}^n$.

The mean payoff vector is defined by

$$\chi(T) := \lim_{k \to \infty} T^k(0)/k = \lim_{k \to \infty} v^k/k \ .$$

By [Koh80], the limit does exist, and that $\chi(T) \in \mathbb{R}^n$ when Assumption 1 and Assumption 2 hold. The problem determining if $\chi_\imath(T) \geq 0$ belongs to the class NP $\cap$ coNP [ZP96], and no polynomial time algorithm is known.

1.4. **Perron-Frobenius tools.** The *spectral radius* of a Shapley operator T is defined as

$$\rho(T) = \sup\{\lambda \in \mathbb{R}_{\max} \mid \exists u \in \mathbb{R}^n_{\max}, u \neq \bot,\ T(u) = \lambda + u\} \ . \tag{5}$$

Variants of this spectral radius are given by the Collatz-Wielandt number cw defined by

$$\mathrm{cw}(T) = \inf\{\lambda \in \mathbb{R} \mid \exists u \in \mathbb{R}^n,\ T(u) \leq \lambda + u\} \ , \tag{6}$$

and by the dual Collatz-Wielandt number

$$\mathrm{cw}'(T) = \sup\{\lambda \in \mathbb{R}_{\max} \mid \exists u \in \mathbb{R}^n_{\max}, u \neq \bot,\ T(u) \geq \lambda + u\} \ . \tag{7}$$

For all $x \in \mathbb{R}^n_{\max}$, we define $\mathrm{top}\, x := \max_{i \in [n]} x_i$ and

$$\overline{\chi}(T) := \lim_k \mathrm{top}(T^k(0))/k = \inf_{k \geq 1} \mathrm{top}(T^k(0))/k \ .$$

The following result, which follows from [AGG12], provides several spectral characterizations of this upper mean payoff.

Theorem 1. *Let $T : \mathbb{R}^n_{\max} \to \mathbb{R}^n_{\max}$ be a Shapley operator. Then*

$$\mathrm{cw}'(T) = \rho(T) = \overline{\chi}(T) = \mathrm{cw}(T) \ , \tag{8}$$

and the suprema in (5) and (7) are always achieved.

Moreover, if the restriction of T to $\mathbb{R}^n$ is piecewise affine, and if $\rho(T) \neq -\infty$, then the infimum in (6) is also achieved.

2. Inner radius of a tropical polyhedron defined by generators

For any subset $\mathcal{W}$ of $\mathbb{R}^n_{\max}$, we define the *inner radius* of $\mathcal{W}$, denoted in-rad$(\mathcal{W})$, as

$$\sup\{r \geq 0 \mid \text{there is a Hilbert ball } B(x,r) \text{ for some } x \in \mathbb{R}^n \text{ such that } B(x,r) \subseteq \mathrm{Sp}(\mathcal{W})\}.$$

More generally, for all non-empty subsets $I \subset [n]$, we define the *relative inner radius* of $\mathcal{W}$, denoted by in-rad$_I(\mathcal{W})$, as the supremum of the radii of the Hilbert balls centered at a point in the part P_I of $\mathbb{R}^n_{\max}$ and included in $\mathrm{Sp}(\mathcal{W})$. Recall that a tropical polyhedral cone is given by

$$\mathrm{Col}(V) = \{V \odot x \mid x \in \mathbb{R}^p_{\max}\}$$

for some $V \in \mathbb{R}^{n \times p}_{\max}$. We shall make the following assumption.

Assumption 3. The matrix V has no identically $-\infty$ rows and no identically $-\infty$ columns.

Set $E = \{(i,k) \in [n] \times [p] \mid V_{ik} \neq -\infty\}$. Then the Shapley operator $T : \mathbb{R}^n_{\max} \to \mathbb{R}^n_{\max}$ is given by

$$T_i(x) = \inf_{k \in [p], (i,k) \in E} \Big[-V_{ik} + \max_{j \in [n], j \neq i} (V_{jk} + x_j) \Big] \ . \tag{9}$$

For $a \in \mathbb{R}^n$, the tropical hyperplane $\mathcal{H}_a$ divides $\mathbb{R}^n_{\max}$ into n sectors $(S_i(a))_{i \in [n]}$, defined by

$$S_i(a) := \{x \in \mathbb{R}^n_{\max} \mid \forall j \in [n],\ x_i + a_i \geq x_j + a_j\} \ . \tag{10}$$

The vector $-a$, which is unique up to an additive constant, is called the *apex* of $\mathcal{H}_a$. The following theorem builds the connection between the inner radius of $\mathrm{Col}(V)$ and the spectral radius of T.

Theorem 2. [AGQS21, Theorem 6] *Let T be the Shapley operator associated to the matrix $V \in \mathbb{R}^{n\times p}_{\max}$ defined in* (9)*. Then $\rho(T) \leq 0$. Moreover,*

$$-\rho(T) = \text{in-rad}(\mathrm{Col}(V)).$$

If $\rho(T)$ is finite, a maximal Hilbert ball included in $\mathrm{Col}(V) \cap \mathbb{R}^n$ is given by $B(-a, -\rho(T))$ where a is any vector in $\mathbb{R}^n$ such that $T(a) \leq \rho(T) + a$.

As a corollary, we recover the following statement which fits our intuition, i.e., a finitely generated polytope has empty interior if and only if it is contained in a hyperplane.

Corollary 3 (Compare with Theorem 4.2 of [DSS05])**.** *The set $\mathrm{Col}(V) \cap \mathbb{R}^n$ is of empty interior if and only if $\mathrm{Col}(V)$ is included in a tropical hyperplane.*

3. The strong duality theorem for tropical linear regression

Let $\mathcal{V} = \{v^{(1)}, \cdots, v^{(p)}\} \subset \mathbb{P}(\mathbb{R}^n_{\max})$ be a finite subset of the tropical projective space, and $V \in \mathbb{R}^{n\times p}_{\max}$ be the matrix whose columns are given by some representatives of $v^{(1)}, \cdots, v^{(p)}$. Recall that a one-sided Hausdorff distance from a set $A \subseteq \mathbb{P}(\mathbb{R}^n_{\max})$ to a set $B \subseteq \mathbb{P}(\mathbb{R}^n_{\max})$ with respect to the Hilbert projective metric is given by

$$\mathrm{dist}_H(A, B) := \sup_{a \in A} \mathrm{dist}_H(a, B) \ , \quad \text{with } \mathrm{dist}_H(a, B) := \inf_{b \in B} d(a, b) \ . \tag{11}$$

Then the *tropical linear regression problem* reads as

$$\inf_{a \in \mathbb{P}(\mathbb{R}^n_{\max})} \mathrm{dist}_H(\mathcal{V}, \mathcal{H}_a) \ . \tag{12}$$

The following theorem presents a strong duality result between a best tropical hyperplane approximation of a set $\mathcal{V}$ of points and the largest inner balls contained in $\mathrm{Sp}(\mathcal{V})$, namely the inner radius geometrically characterize the distance from $\mathcal{V}$ to its best fitting tropical hyperplane.

Theorem 4 (Strong duality)**.** [AGQS21, Theorem 20] *We have*

$$\min_{b \in \mathbb{P}(\mathbb{R}^n_{\max})} \mathrm{dist}_H(\mathcal{V}, \mathcal{H}_b) = r^{\mathrm{in}}_{\mathcal{V}} = \sup\{r \geq 0 \mid \exists a \in \mathbb{R}^n, B(a, r) \subseteq \mathrm{Sp}(\mathcal{V})\}. \tag{13}$$

The minimum is achieved by any vector $b \in \mathbb{P}(\mathbb{R}^n_{\max})$ such that $T(b) \geq \rho(T) + b$. Moreover, if $r^{\mathrm{in}}_{\mathcal{V}}$ is finite, the supremum is achieved by a ball $B(-c, r^{\mathrm{in}}_{\mathcal{V}})$ where $c \in \mathbb{R}^n$ is any vector such that $T(c) \leq \rho(T) + c$.

Given a hyperplane $\mathcal{H}_b$, we call *witness point* of $\mathcal{H}_b$ any point p in $\mathcal{V}$ such that the distance from p to the hyperplane $\mathcal{H}_b$ equals the distance from the set $\mathcal{V}$ to this hyperplane. By Theorem 2 and Theorem 4, we obtain the following close relation between spectral radius of T and the tropical linear regression.

Theorem 5 (Optimality certificates)**.** [AGQS21, Theorem 22] *For $a \in \mathbb{R}^n$, then the following assertions are equivalent:*

(1) *$T(a) = \rho(T) + a$;*

(2) *The hyperplane $\mathcal{H}_a$ admits a witness point in each sector, meaning that* $\forall i \in [n], \exists k \in [p]$, $v^{(k)} \in S_i(a)$ *and* $\operatorname{dist}_H(v^{(k)}, \mathcal{H}_a) = \operatorname{dist}_H(\mathcal{V}, \mathcal{H}_a)$.

Moreover, if these assertions hold, then $\rho(T) = -\operatorname{dist}_H(\mathcal{V}, \mathcal{H}_a)$, $\mathcal{H}_a$ *is an optimal solution of the tropical linear regression problem, and* $B(-a, \operatorname{dist}_H(\mathcal{V}, \mathcal{H}_a))$ *is a Hilbert ball of maximal radius included in* $\operatorname{Sp}(\mathcal{V})$.

Due to Theorem 5, it would be interesting and useful to investigate the situations where T has a finite eigenvector with respect to its spectral radius. Below is a special case where we can guarantee this property of T.

Proposition 6. [AGQS21, Proposition 23] *Suppose that all the vectors $v \in \mathcal{V}$ have finite entries. Then, the operator T has a finite eigenvector a.*

Theorem 5 also implies the following corollary which reduces the tropical linear regression problem to finding the spectral radius of T.

Corollary 7. *The tropical linear regression problem is polynomial time Turing-equivalent to mean payoff game.*

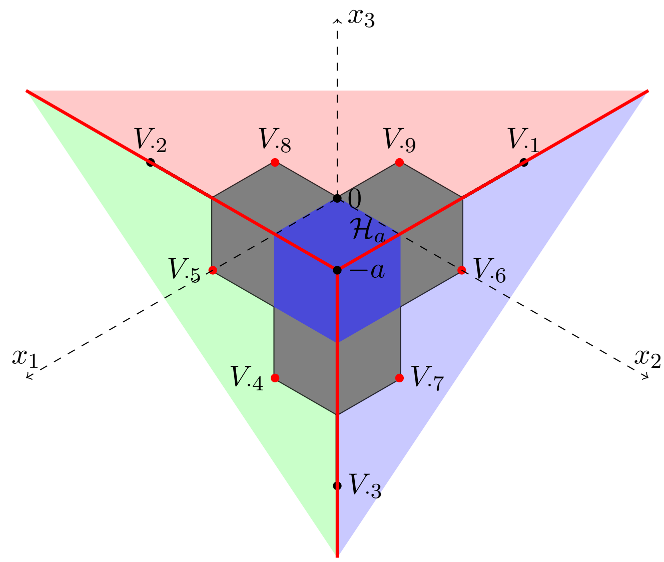

FIGURE (1) The inner ball of a column space $\operatorname{Col}(V)$ and the linear regression of the columns of V.

4. Tropical linear regression with sign or type patterns

4.1. **Tropical linear regression with signs.** Given $I, J \subset [n]$ such that $I, J \neq \emptyset$, $I \cup J = [n]$ and $I \cap J = \emptyset$ and $a \in \mathbb{P}(\mathbb{R}^n_{\max})$, we define the *signed tropical hyperplane* of type (I, J) by

$$\mathcal{H}_a^{IJ} := \{x \in \mathbb{R}^n_{\max} \mid \max_{i \in I}(a_i + x_i) = \max_{j \in J}(a_j + x_j)\}. \tag{14}$$

Given a finite set $\mathcal{V} \subset \mathbb{R}^n_{\max}$, of cardinality $|\mathcal{V}| = p$, the *signed tropical linear regression problem* of type (I, J) consists in finding the best approximation of $\mathcal{V}$ by a signed hyperplane of type (I, J):

$$\min_{a \in \mathbb{P}(\mathbb{R}^n_{\max})} \operatorname{dist}_H(\mathcal{V}, \mathcal{H}_a^{IJ}) \ . \tag{15}$$

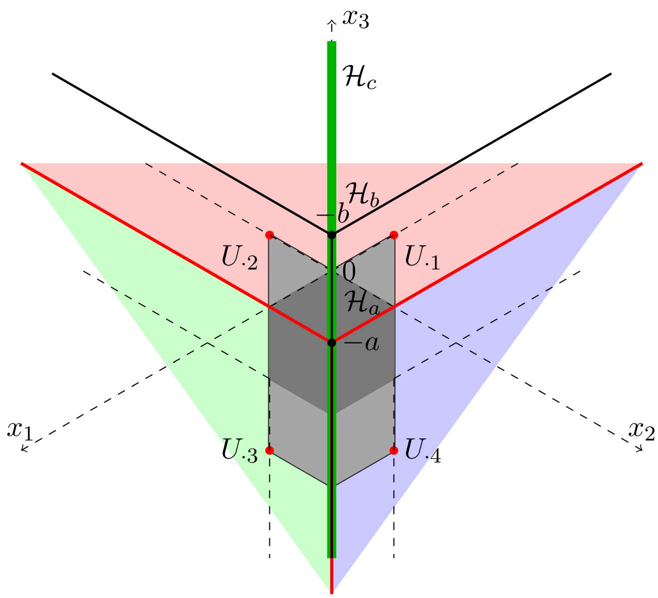

FIGURE (2) A column space $\mathrm{Col}(U)$ (light and dark gray regions) with multiple hyperplanes that are optimal solutions of the tropical linear regression problem, and multiple inner balls of maximal radius, but a unique optimal hyperplane with witness points in each sector, corresponding to the finite eigenvector $a = (0,0,1)^\top$ of T and to the inner ball in dark gray.

Let M be a closed tropical cone of $\mathbb{R}^n_{\max}$ and $x \in \mathbb{R}^n_{\max}$. The projection $P_M(x)$ of the point x onto M [CGQ04] is defined by

$$P_M(x) := \max\{z \in M \mid z \leq x\} \ . \tag{16}$$

The theorem below guarantees that the Hilbert distance from x to M is achieved by the projection $P_M(x)$ of x onto M.

Theorem 8. [CGQ04, Theorem 18] *Given a closed tropical semimodule $M \subset \mathbb{R}^n_{\max}$ and $x \in \mathbb{R}^n_{\max}$, we have:*

$$\mathrm{dist}_H(x, M) = d(x, P_M(x)) \ .$$

In the sequel, we suppose that the following Assumption 4 holds.

Assumption 4. We suppose that for each $l \in [n]$, there exists $v \in \mathcal{V}$, such that $v_l \neq -\infty$.

We now introduce the operator $T^{IJ} : \mathbb{R}^n_{\max} \mapsto \mathbb{R}^n_{\max}$, defined by

$$T^{IJ}_l(x) := \begin{cases} \inf_{v \in \mathcal{V}, v_l \neq -\infty}\{-v_l + \max_{j \in J}(v_j + x_j)\}, \text{ if } \ l \in I \ , \\ \inf_{v \in \mathcal{V}, v_l \neq -\infty}\{-v_l + \max_{i \in I}(v_i + x_i)\}, \text{ if } \ l \in J \ . \end{cases} \tag{17}$$

Similar to Theorem 4, we now derive a strong duality theorem for signed tropical regression.

Theorem 9. [AGQS21, Theorem 34] *We have*

$$\min_{a \in \mathbb{P}(\mathbb{R}^n_{\max})} \mathrm{dist}_H(\mathcal{V}, \mathcal{H}^{IJ}_a) = -\rho(T^{IJ}) = \sup\{r \geq 0 \mid \exists w \in \mathbb{R}^n, B_{IJ}(w, r) \subset \mathrm{Sp}(\mathcal{V})\} \ . \tag{18}$$

The minimum is achieved by any vector $b \in \mathbb{P}(\mathbb{R}^n_{\max})$ such that $T^{IJ}(b) \geq \rho(T^{IJ}) + b$. Moreover, if $\rho(T^{IJ})$ is finite, the supremum is achieved by a ball $B(c, \rho(T^{IJ}))$ where $c \in \mathbb{R}^n$ can be deduced from any vector u such that $T^{IJ}(u) \leq \rho(T^{IJ}) + u$.

Remark 1. When the set $I = \{i\}$ is of cardinality one, the regression problem for the signed hyperplane (14) has the following special form:

$$\min_{a \in \mathbb{R}^n} \max_{v \in \mathcal{V}} |v_i - (\max_{j \neq i} a_j - a_i + v_j)| \ . \tag{19}$$

This can be solved in a direct way [MCT21], avoiding the recourse to mean payoff games. Indeed, (19) reduces to the following "one-sided" tropical linear regression problem. Given sample points $(x^{(k)}, y^{(k)})$ in $\mathbb{R}^n \times \mathbb{R}^m$, for $k \in [p]$, compute

$$\min_{A} \max_{k \in [p]} \|y^{(k)} - Ax^{(k)}\|_\infty \ , \tag{20}$$

where the minimum is taken over tropical matrices A of size $m \times n$, and the product $Ax^{(k)}$ is understood tropically. Up to a straightforward duality, this problem was solved in [But10, Theorem 3.5.2], the result being attributed there to Cuninghame-Green [CG79]. Alternatively, this solution may be recovered by combining [CF00, Coro. 1] with the explicit formula of the tropical projection [CGQ04, Th. 5]. More precisely, define the matrix $\bar{A} \in \mathbb{R}^{m \times n}$ by $\bar{A}_{ij} := \min_{k \in [p]} y_i^{(k)} - x_j^{(k)}$, so that $\bar{A}$ is the maximal matrix such that $Ax^{(k)} \leq y^{(k)}$ for all $k \in [p]$. Let $\delta := \max_{k \in p} \|y^{(k)} - \bar{A}x^{(k)}\|_\infty$, and $A^{\mathrm{opt}}_{ij} = \bar{A}_{ij} + \delta/2$. Then, A^{opt} is the greatest optimal solution. It can be computed in $O(mnp)$ arithmetic operations. By specializing this formula, one can solve (19) in $O(np)$ arithmetic operations. We refer the reader to [MCT21] for more information, and for the solution of further problems of this category.

4.2. **Tropical linear regression with type information.** Suppose the set of points $\mathcal{V}$ is the disjoint union $\mathcal{V} = \cup_{i \in [n]} \mathcal{V}_i$, where each $\mathcal{V}_i$ is non-empty. We shall say that the points of $\mathcal{V}_i$ are of *type* $i \in [n]$. Note that the set of types is the same as the set of indices of vectors. For each type $i \in [n]$, we consider the signed hyperplane:

$$\mathcal{H}_a^i := \mathcal{H}_a^{\{i\}\{i\}^c} = \{x \in \mathbb{R}^n_{\max} \mid a_i + x_i = \max_{j \neq i}(a_j + x_j)\} \ .$$

The *typed tropical linear regression* problem associated to the partition $\mathcal{V} = \cup_{i \in [n]} \mathcal{V}_i$ is defined as

$$\min_{a \in \mathbb{P}(\mathbb{R}^n_{\max})} \max_{i \in [n]} \mathrm{dist}_H(\mathcal{V}_i, \mathcal{H}_a^i) \ . \tag{21}$$

Assume in the sequel that Assumption 4 holds. For each type $i \in [n]$, consider the Shapley operator $T^{\mathrm{ty},i} : \mathbb{R}^n_{\max} \to \mathbb{R}^n_{\max}$, given by (17) where the type considered is $(I, J) = (\{i\}, \{i\}^c)$:

$$T_l^{\mathrm{ty},i}(x) := \begin{cases} \inf_{v \in \mathcal{V}_i, v_i \neq -\infty} \{-v_i + \max_{j \neq i}(v_j + x_j)\}, \text{ if } \ l = i \ , \\ \inf_{v \in \mathcal{V}_i, v_l \neq -\infty} \{-v_l + v_i\} + x_i, \text{ if } \ l \neq i \ . \end{cases} \tag{22}$$

Consider now the Shapley operator $T^{\mathrm{ty}} : \mathbb{R}^n_{\max} \to \mathbb{R}^n_{\max}$ given by the infimum of the operators $T^{\mathrm{ty},i}, i \in [n]$:

$$T_l^{\mathrm{ty}}(x) := \min_{i \in [n]} T_l^{\mathrm{ty},i}(x) \ . \tag{23}$$

Similar to Theorem 5, we obtain the equivalence between the typed tropical linear regression problem and the problem of calculating the spectral radius of T^{ty}.

Theorem 10. [AGQS21, Theorem 36] *We have*

$$\min_{a \in \mathbb{P}(\mathbb{R}^n_{\max})} \max_{i \in [n]} \mathrm{dist}_H(\mathcal{V}_i, \mathcal{H}_a^i) = -\rho(T^{\mathrm{ty}}) \ .$$

Moreover, the minimum is achieved by any vector $a \in \mathbb{P}(\mathbb{R}^n_{\max})$ *such that* $T^{\mathrm{ty}}(a) \geq \rho(T^{\mathrm{ty}}) + a$.

Remark 2. Typed tropical linear regression should be compared with the tropical SVM problem introduced in [GJ08]. In the tropical SVM setting, we have a partition of the set of points in n *color classes*, $\mathcal{V}_{c_1}, \dots, \mathcal{V}_{c_n}$, and we are looking for a tropical hyperplane $\mathcal{H}_a$, and for a permutation σ of $\{1, \dots, n\}$ such that for all $i \in [n]$, all the points of color c_i are in the same *sector* $S_{\sigma(i)}(a)$. In other words, we want the tropical hyperplane to separate the n color classes. This is not possible in general, so one needs to consider metric versions, modelling the minimization of classification errors [TWY20]. A possible metric formulation, in the spirit of the present approach, would be to consider

$$\min_{\sigma \in \mathfrak{S}_n} \min_{a \in \mathbb{R}^n} \max_{i \in [n]} \operatorname{dist}_H(\mathcal{V}_i, S_{\sigma(i)}(a)) \qquad \text{(Metric Tropical SVM)} \tag{24}$$

where $\mathfrak{S}_n$ denotes the symmetric group on n letters. By comparison with (21), we see that we have in addition a minimization over the symmetric group, but the subproblem with a fixed permutation σ arising in the SVM problem is simpler than the analogous problem of typed tropical linear regression, since the sector $S_{\sigma(i)}$ is convex, whereas the set $\mathcal{H}_a^i$ arising in (21) is not a convex one.

5. Algorithmic aspects

Considering the strong duality result, Theorem 4, and the result on the existence of witness points Theorem 5, the key algorithmic issues are:

(i) to compute the upper mean payoff, $\rho(T)$ (which is the opposite of the value of the tropical linear regression problem);

(ii) to decide whether there is a finite eigenvector $u \in \mathbb{R}^n$ such that $T(u) = \rho(T) + u$, and to compute such an eigenvector (when this is so, $-u$ is the center an an optimal ball included in $\operatorname{Sp}(\mathcal{V})$ and the apex of an optimal regression hyperplane);

(iii) to find a sub-eigenvector $b \in (\mathbb{R}_{\max})^n \setminus \{\bot\}$, satisfying $T(b) \geq \rho(T) + b$ (then, $\mathcal{H}_b$ is an optimal regression hyperplane);

(iv) to find a super-eigenvector $c \in \mathbb{R}^n$ satisfying $T(c) \leq \rho(T) + c$ (then, $-c$ is the center of an optimal ball included in $\operatorname{Sp}(\mathcal{V})$.

Problems (i)–(iv) will be solved simultaneously as soon as we know an invariant half-line of T. More generally, an algorithm which returns an optimal policy σ of Player Min, i.e., a policy such that $\chi(T) = \chi(T^\sigma)$, can be used to produce a finite vector $c \in \mathbb{R}^n$ such that $T(c) \leq \overline{\chi}(T) + c$, by reduction to a tropical eigenvalue problem. Moreover, any algorithm which returns an optimal policy τ of Player Max, i.e., a policy such that $\chi(T) = \chi({}^\tau T)$, can be used to produce a vector $b \in \mathbb{R}^n_{\max} \setminus \{\bot\}$, satisfying $T(b) \geq \rho(T) + b$.

We refer the reader to [Cha09] for a comparative discussion of mean payoff game algorithms. The main known algorithms include the pumping algorithm of [GKK88], value iteration [ZP96], and different algorithms based on the idea of policy iteration [BV07, Sch08, DG06]. In particular, the algorithm of [DG06] returns an invariant half-line. The policy iterations algorithms [BV07, DG06] were reported in [Cha09] to have the best experimental behavior, although policy iteration is generally exponential [Fri09].

For the application to tropical linear regression, we often know in advance that the operator T has a finite eigenvector. Then, one can use another algorithm, projective Krasnoselkii-Mann value iteration [GS20], which is straightforward to implement and still effective. Starting from a vector $v^0 = (0, \cdots, 0)^\top$, this algorithm computes the following sequence:

$$\tilde{v}^{k+1} = T(v^k) - (\max_{i \in [n]} T(v^k)_i)e, \tag{25}$$

$$v^{k+1} = (1-\gamma)v^k + \gamma \tilde{v}^{k+1}. \tag{26}$$

where $e = (1, \cdots, 1)^\top \in \mathbb{R}^n$, and $\gamma \in (0,1)$ is fixed, $1-\gamma$ being interpreted as a *damping parameter*.

A direct computation gives us the number of arithmetic operations we need for the projective Krasnoselkii-Mann algorithm.

Proposition 11. [AGQS21, Proposition 37] *The operator T can be evaluated in $O(|E|)$ arithmetic operations.*

We set:

$$W := \max_{v \in \mathcal{V}} \|v\|_H \ .$$

The complexity of the approximate optimality certificate is given below.

Corollary 12. *Suppose that $\mathcal{V} \subset \mathbb{R}^n$ is of cardinality p. Then, an ϵ-approximation of the inner radius of* $\mathrm{Col}(V)$*, as well as vectors $v, z \in \mathbb{R}^n$ satisfying $B_H(v, \text{in-rad}(\mathcal{V}) - \epsilon) \subset \mathcal{V}$ and $\mathrm{dist}_H(\mathcal{V}, \mathcal{H}_z) \leq \text{in-rad}(\mathcal{V}) + \epsilon$ can be obtained in $O(npW/\epsilon)$ arithmetic operations.*

6. Open problems

Several open problems related to the present work arise when changing either the class of metrics or of tropical spaces.

For instance, we may replace the Hilbert metric by the L_p-projective metric, i.e., the metric obtained by modding out the L_p normed space $\mathbb{R}^n$ by the action of additive constants, or by replacing the Hausdorff distance in (11) by a L_p type distance, for $p \in [1, \infty)$. Approaches based on mixed linear programming, or on local descent, have been proposed in [YZZ19, PYZ20, Hoo17] in some specific cases.

Another generalization consists in replacing hyperplanes by tropical linear spaces of a codimension not necessarily 1. Recall that the *tropical Grassmannian* $\mathrm{Gr}_{k,n}^{\mathrm{trop}}$ can be defined as the image by a non-archimedean valuation of the Grassmannian $\mathrm{Gr}_{k,n}(\mathbb{K})$ over an (algebraically closed) non-archimedean field, under the Plücker embedding, see [SS04, FR15]. In this way, an element of $\mathrm{Gr}_{k,n}^{\mathrm{trop}}$ is represented by its *tropical Plücker coordinates* $p = (p_I) \in \mathbb{R}_{\max}^{\binom{n}{k}}$. This vector yields a *tropical linear space* $L(p)$, defined by

$$L(p) = \bigcap_I \{x \in \mathbb{R}^n_{\max} \mid \max_{i \in I}(p_{I \setminus \{i\}}) + x_j) \text{ is achieved at least twice}\} \ ,$$

where the minimum is taken over all subsets of $[n]$ of cardinality $k+1$. When $k = n-1$, $V(p)$ is a tropical hyperplane. Hence, a general version of tropical linear regression problem can be written as

$$\min_{p \in \mathrm{Gr}_{k,n}^{\mathrm{trop}}} \max_{v \in \mathcal{V}} \min_{x \in L(v)} \|v - x\|_H \ . \tag{27}$$

We solved here this problem when $k = n-1$. When $k = 1$, $L(p)$ is reduced to a single point, and it is not difficult to see that (27) reduces to a linear program. We leave it as an open question to solve this problem when $1 < k < n-1$. The same problem may be considered when p is a valuated matroid, or when it is inside the image of the Stiefel map [FR15], meaning that p is given by the maximal tropical minors of a matrix. A version of the latter problem (with a L_1-type error) is considered in [YZZ19]. One may also replace the linear space $L(p)$ by the *column space* of a tropical matrix A, which boils down to finding a best approximation by a tropical polyhedral cone with a fixed number of vertices, see [Hoo17, PYZ20].

Another open question is to characterize the Boolean patterns of V which guarantee that the operator T has a finite eigenvector.

Acknowledgment

This work is partially supported by the Thematic Research Programme "Tensors: geometry, complexity and quantum entanglement", University of Warsaw, Excellence Initiative – Research University and the Simons Foundation Award No. 663281 granted to the Institute of Mathematics of the Polish Academy of Sciences for the years 2021–2023.

References

[AGG12] M. Akian, S. Gaubert, and A. Guterman. Tropical polyhedra are equivalent to mean payoff games. *International Journal of Algebra and Computation*, 22(1):125001 (43 pages), 2012.

[AGQS21] Marianne Akian, Stéphane Gaubert, Yang Qi, and Omar Saadi. Tropical linear regression and mean payoff games: or, how to measure the distance to equilibria. Preprint, 2021.

[But10] P. Butkovič. *Max-linear Systems: Theory and Algorithms*. Springer Monogr. Math. Springer, London, 2010.

[BV07] H. Bjorklund and S. Vorobyov. A combinatorial strongly subexponential strategy improvement algorithm for mean payoff games. *Discrete Appl. Math.*, 155:210229, 2007.

[CF00] V. Chepoi and B. Fichet. ℓ_∞-approximation via subdominants. *Journal of Mathematical Psychology*, 44:600–616, 2000.

[CG79] R. A. Cuninghame-Green. *Minimax algebra*, volume 166 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 1979.

[CGQ04] G. Cohen, S. Gaubert, and J. P. Quadrat. Duality and separation theorems in idempotent semimodules. *Linear Algebra and Appl.*, 379:395–422, 2004.

[Cha09] J. Chaloupka. Parallel algorithms for mean-payoff games: an experimental evaluation. In *Algorithms—ESA 2009*, volume 5757 of *Lecture Notes in Comput. Sci.*, pages 599–610. Springer, Berlin, 2009.

[DG06] V. Dhingra and S. Gaubert. How to solve large scale deterministic games with mean payoff by policy iteration. In *Proceedings of the 1st international conference on Performance evaluation methodolgies and tools (VALUETOOLS)*, volume 180, Pisa, Italy, 2006. article No. 12.

[DSS05] M. Develin, F. Santos, and B. Sturmfels. On the rank of a tropical matrix. In *Combinatorial and computational geometry*, volume 52 of *Math. Sci. Res. Inst. Publ.*, pages 213–242. Cambridge Univ. Press, Cambridge, 2005.

[FR15] A. Fink and F. Rincón. Stiefel tropical linear spaces. *Journal of Combinatorial Theory, Series A*, 135:291 – 331, 2015.

[Fri09] O. Friedmann. An exponential lower bound for the parity game strategy improvement algorithm as we know it. In *Proceedings of the 24th Annual IEEE Symposium on Logic in Computer Science (LICS 2009)*, pages 145–156. IEEE Computer Society, 2009.

[GJ08] B. Gärtner and M. Jaggi. Tropical support vector machines. Technical report, ACSTR-362502-01, 2008.

[GKK88] V. A. Gurvich, A. V. Karzanov, and L. G. Khachiyan. Cyclic games and finding minimax mean cycles in digraphs. *Zh. Vychisl. Mat. i Mat. Fiz.*, 28(9):1407–1417, 1439, 1988.

[GS20] S. Gaubert and N. Stott. A convergent hierarchy of non-linear eigenproblems to compute the joint spectral radius of nonnegative matrices. *Mathematical Control and Related Fields*, 10(3):573–590, 2020.

[Hoo17] J. Hook. Linear regression over the max-plus semiring: algorithms and applications, 2017. arXiv:1712.03499.

[Koh80] E. Kohlberg. Invariant half-lines of nonexpansive piecewise-linear transformations. *Math. Oper. Res.*, 5(3):366–372, 1980.

[MCT21] P. Maragos, V. Charisopoulos, and E. Theodosis. Tropical geometry and machine learning. *Proceedings of the IEEE*, 109(5):728–755, May 2021.

[PYZ20] R. Page, R. Yoshida, and L. Zhang. Tropical principal component analysis on the space of phylogenetic trees. *Bioinformatics*, 36(17):4590–4598, 06 2020.

[Sch08] S. Schewe. An optimal strategy improvement algorithm for solving parity and payoff games. In M. Kaminski and S. Martini, editors, *Computer Science Logic*, pages 369–384, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[SS04] D. Speyer and B. Sturmfels. The tropical Grassmannian. *Adv. Geom.*, 4(3):389–411, 2004.

[TWY20] X. Tang, H. Wang, and R. Yoshida. Tropical support vector machine and its applications to phylogenomics, 2020. arXiv:2003.00677.

[YZZ19] R. Yoshida, L. Zhang, and X. Zhang. Tropical principal component analysis and its application to phylogenetics. *Bulletin of Mathematical Biology*, 81:568–597, 2019.

[ZP96] U. Zwick and M. Paterson. The complexity of mean payoff games on graphs. *Theoret. Comput. Sci.*, 158(1-2):343–359, 1996.

INRIA and CMAP, École polytechnique, IP Paris, CNRS
Email address: Yang.Qi@inria.fr