

# TROPICAL SUPPORT VECTOR MACHINES AND MEAN-PAYOFF GAMES

XAVIER ALLAMIGEON, STÉPHANE GAUBERT, SAMUEL BOÏTÉ, THÉO MOLFESSIS

**ABSTRACT.** We introduce a pseudo-polynomial algorithm for tropical support vector machines (SVMs), using mean-payoff games for binary and multiclass classification. We provide a simple numerical criterion for tropical separability. In the separable case, we construct an optimal margin tropical hyperplane. In the non-separable case, we use the same algorithm to provide a suitable classifier ; it also measures the distance to tropical separability and allows to geometrically transform the data to reduce this distance to zero. We then extend our algorithm and results to the multiclass case. Finally, we study a feature map method to make the data separable in higher dimensions. We propose a heuristic method for sparsely adding relevant features to better separate the data.

## 1. INTRODUCTION

**Motivation and context.** *Support vector machines* (SVMs), introduced by Boser et al. (1992) [1], are used in data science to determine separating hyperplanes for binary classification tasks.

Tropical geometry recently found applications in machine learning: tropical linear regression (Akian et al., 2021) [2], analysis of deep neural networks and parametric statistical models (Maragos et al., 2021) [3] and principal component analysis (Page et al., 2019) [4].

First considered by Gärtner and Jaggi (2006) [5], *tropical support vector machines* classify data points with a tropical hyperplane and using the tropical metric. Tropical SVMs find natural applications in phylogenomics (Tang et al., 2020) [6].

Initially formulated as linear programs [5], tropical hyperplanes can also be obtained by tropicalizing a classical hyperplane using *Maslov dequantization* [7] or Viro’s method [8]. Yet, redundant features can cause difficulties in the performance of  $L^2$ -norm classical SVMs (Zhang and Zhou, 2010) [9]. Moreover, tropicalization relies on exponentialization and is subject to arithmetic errors, while not directly maximizing the margin in the initial space.

Tropical linear regression has been shown by Akian et al. (2021) [2] as equivalent to solving mean-payoff games. We consider this parallel for tropical support vector machines.

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

$$(1.1) \quad \mathcal{H}_a = \left\{ x \in \mathbb{R}_{\max}^d, \quad \max_{1 \leq i \leq d} (x_i + a_i) \text{ is achieved at least twice} \right\}.$$

Such a hyperplane is parameterized by the vector  $a = (a_1, \dots, a_d) \in \mathbb{R}_{\max}^d$ , which is required to be non-identically  $-\infty$ . It hence partitions the space depending on which coordinate reaches its maximum, i.e. between  $d$  different sectors. By definition, tropical hyperplanes are invariant by translation along the constant vector  $\mathbf{1}_d := (1, \dots, 1)$ .

In this paper, we address the following tropical analogue of support vector machines. Given  $n \in \{1, \dots, d\}$  classes of  $d$ -dimensional data points  $X^1, \dots, X^n$ , we look for a maximum-margin separating tropical hyperplane to build a classifier. The notion of margin depends on the metric: the canonical choice in tropical geometry is the (additive version of) Hilbert’s projective metric. Its restriction to  $\mathbb{R}^d$  is induced by the so-called *Hilbert’s seminorm* or *Hopf oscillation*

$$\|x\|_H := \max_{1 \leq i \leq d} x_i - \min_{1 \leq i \leq d} x_i.$$

It is a projective metric, in the sense that the distance between two points is zero if and only if these two points differ by an additive constant.

When dealing with hard-margin support vector machines (SVMs), we also have to ensure that every point lands in the right sector. Each class  $k$  belongs to some sectors  $I^k \subset \{1, \dots, d\}$ , with each sector assigned to only one class. We then define the *tropical parameterized hyperplane* of configuration  $\sigma := \{I^1, \dots, I^n\}$ , where  $I^k$  and  $I^\ell$  are disjoint for  $k \neq \ell$ , as:

$$\mathcal{H} := \{x \in \mathbb{R}_{\max}^d, \quad \exists k \neq \ell, \quad (x - a) \text{ reaches its max coordinate in } I^k \text{ and } I^\ell\}.$$

We note  $\mathcal{S}^k \subset \mathbb{R}_{\max}^d$  the *sector* classified as class  $k$ , where the maximum coordinate is reached in  $I^k$ , and  $\mathcal{S}^0$  the union of unclassified sectors, corresponding to coordinates  $I^0$ .  $\mathcal{H}$  is said to *separate* point clouds  $(X^k)_{1 \leq k \leq p}$  with a margin of  $\nu$  when for every point  $x^k$  of class  $k$ :

- (1)  $x^k$  is on the right sector of the hyperplane, i.e its maximum coordinate is reached in  $I^k$
- (2)  $x^k$  is at distance at least  $\nu$  from any other sector  $\mathcal{S}^\ell$ ,  $\ell \in \{1, \dots, n\} \cup \{0\}$ ,  $\ell \neq k$ :

$$d_H(\mathcal{S}^\ell, x^k) := \max(x^k - a) - \max_{i \in I^\ell} (x^k - a)_i \geq \nu.$$

We say that the margin is *exact* when at least one of these distance inequalities is reached, and that it is *optimal* when no other tropical hyperplane can separate our points with a better margin. Notably, here, multiclass optimality only requires maximizing the minimum of the distances  $d(\mathcal{S}^\ell, x^k)$ .

In Figure 1.1a, we separate two classes of 3-dimensional points from a toy tropical dataset using a tropical hyperplane. The two lower sectors are assigned to one class, the upper sector corresponds to the other class. Its margin is maximal and is represented in pale yellow. The figure is represented in the projective space  $\mathbb{P}(\mathbb{R}_{\max}^3)$ , i.e. the quotient of the set of non-identically  $-\infty$  vectors of  $\mathbb{R}_{\max}^3$  by the equivalence relation which identifies tropically proportional values.

Since our separator is translation-invariant by the constant vector, in order to separate any  $d$ -dimensional dataset, we plunge it into the  $d$  dimensional subspace  $\mathbb{R}_{\max}^{d+1}$  defined by equation  $x_1 + \dots + x_{d+1} = 0$ . This amounts to artificially adding a new coordinate to each point, equal to the opposite of the sum of the initial coordinates. In the rare case where we are dealing with an intrinsically tropical dataset, for which this translation invariance makes sense, it is better to apply the tropical classifier on the initial features.

**Contribution.** We formulate the tropical classification problem as a mean-payoff game. We apply a Krasnoselskii-Mann iteration scheme to derive an  $\varepsilon$ -approximation of the tropical classifier in pseudo-polynomial time, and we develop geometric guarantees, e.g. on the optimality of the margin in the separable case. Finally, we propose a feature map for fitting tropical polynomials while retaining margin information. Thanks to their evaluation speed and their piecewise linear hypersurfaces, we enable more complex classification tasks. Figure 1.1 shows how our classifiers look like for 3D data. Our algorithm is implemented in *Python* in the following repository: <https://gitlab.inria.fr/tropical/tropical-svm>.

## 2. SHAPLEY OPERATORS AND MEAN-PAYOFF GAMES

In this section, we have a finite set of points  $X = (x_1, \dots, x_p) \in \mathbb{R}_{\max}^{d \times p}$ . We define the *tropical span* of  $X$  as the set of tropical linear combinations of these points:

$$\text{Span}(X) := \left\{ \max_{1 \leq i \leq p} (x_i + \lambda_i), \quad \lambda \in \mathbb{R}^p \right\}.$$

As this is also the smallest tropically convex set containing these points, separating several point clouds amounts to separating their tropical spans. Given a tropically convex, compact and nonempty subset  $\mathcal{V}$  of  $\mathbb{R}_{\max}^d$ , we define the projection of  $x$  in  $\mathbb{R}_{\max}^d$  by:

$$P_{\mathcal{V}}(x) := \max\{y \in \mathcal{V}, \quad y \leq x\}.$$

When  $\mathcal{V}$  is the tropical span of  $X$ , we will also note this projection as  $P_X$ . Tropical projections can be expressed as a mean-payoff game: for  $i \in \{1, \dots, p\}$  and  $x$  in  $\mathbb{R}_{\max}^d$ , [10]

$$P_X(x)_i = \max_{1 \leq j \leq p} \left\{ X_{ij} + \min_{1 \leq k \leq d} (-X_{kj} + x_k) \right\}.$$

Here, we recognize the payoff of a zero-sum deterministic game with perfect information. There are two players, “Max” and “Min” (the maximizer and the minimizer), who alternate their actions. Starting on node  $i$ , Player Max chooses to move to node  $j$ , and receives  $X_{ij}$  from Player Min. Similarly, Player Min in turn chooses a node  $k$  and has to pay  $-X_{kj}$  to Player Max.

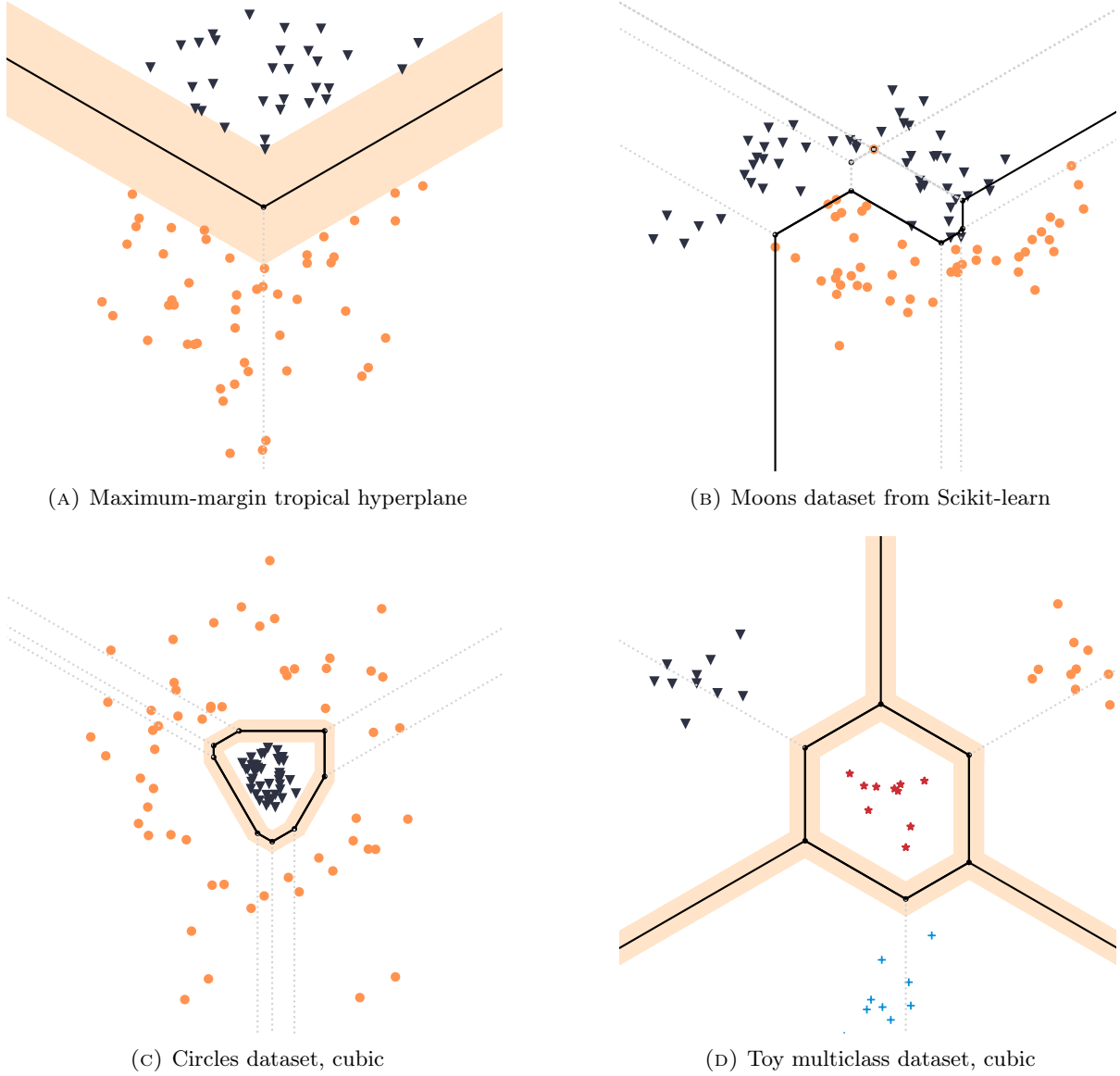


FIGURE 1.1. Tropical piecewise linear classifiers

We now recall some tools from Perron-Frobenius theory, related to mean-payoff games. We refer the reader to [11] for more information.

We call *Shapley operator* a map  $T : \mathbb{R}_{\max}^d \rightarrow \mathbb{R}_{\max}^d$  that is *order-preserving* and *additively homogenous*, i.e.  $T(\alpha + x) = \alpha + T(x)$  for all  $\alpha$  in  $\mathbb{R}_{\max}^d$  and  $x$  in  $\mathbb{R}_{\max}^d$ .  $T$  is said to be *diagonal free* when  $T_i(x)$  is independent of  $x_i$  for all  $i$  in  $\{1, \dots, n\}$ , i.e. when for all  $x, y$  in  $\mathbb{R}_{\max}^d$  which coincide on all coordinates except the  $i$ -th, we have  $T_i(x) = T_i(y)$ .

Tropical projections are Shapley operators. Given a Shapley operator  $T$ , we also define

$$\mathcal{S}(T) := \{x \in \mathbb{R}_{\max}^d, \quad x \leq T(x)\}.$$

More specifically, any tropically convex set  $\mathcal{V}$  is equal to  $\mathcal{S}(P_{\mathcal{V}})$ . Indeed, as for any vector  $x$  in  $\mathbb{R}_{\max}^d$ ,  $P_{\mathcal{V}}(x) \leq x$ , having  $x$  in  $\mathcal{S}(P_{\mathcal{V}})$  implies  $x = P_{\mathcal{V}}(x)$ , meaning  $x$  is in  $\mathcal{V}$ .

The *spectral radius* of  $T$  is, by definition: [12]

$$\rho(T) := \sup \{ \lambda \in \mathbb{R}_{\max}, \quad \exists u \in \mathbb{R}_{\max}^d \setminus \{\perp\}, \quad T(u) = \lambda + u \}.$$

$\rho(T)$  is also equal to the *upper Collatz–Wielandt number* of  $T$ : [12]

$$(2.1) \quad \rho(T) = \inf \{ m \in \mathbb{R}, \quad \exists u \in \mathbb{R}_{\max}^d \setminus \{\perp\}, \quad T(u) \leq m + u \}.$$

Operator  $P$  can be slightly tweaked to make it *diagonal-free* – but the resulting operator is not a projector anymore. In the modified game, the opponent is prevented from replying eye-for-an-eye:

$$[P_X^{\text{DF}}(x)]_i := \max_{1 \leq j \leq p} \left\{ X_{ij} + \min_{k \neq i} (-X_{kj} + x_k) \right\}.$$

Noticeably,  $\mathcal{S}(P_X^{\text{DF}}) = \mathcal{S}(P_X)$ .  $P$  and  $P^{\text{DF}}$  can be evaluated in time  $\mathcal{O}(dp)$  as they involve computing the two greatest values of  $(-X_{kj} + x_k)_{1 \leq k \leq d}$  for  $j$  in  $\{1, \dots, p\}$ .

To conclude, the point clouds we seek to separate are represented by the tropical projection or its diagonal-free counterpart.

### 3. BINARY TROPICAL SUPPORT VECTOR MACHINES

**3.1. Hard-Margin classifier.** In this section, we introduce a practical criterion for the existence of a separating tropical hyperplane, and if it exists, we construct one with optimal margin. We also place ourselves in the binary setting, where we seek to separate two convex hulls  $\mathcal{V}^+$  and  $\mathcal{V}^-$ . We assume we have two Shapley operators  $T^+$  and  $T^-$  such that  $\mathcal{S}(T^\pm) = \mathcal{V}^\pm$ , which is for instance the case when separating finite point clouds: corresponding operators are projections  $P_{\mathcal{V}^\pm}$ . We define:

$$T = \min(T^+, T^-),$$

which is also a Shapley operator. We denote by  $\perp$  the vector of  $\mathbb{R}_{\max}^d$  identically equal to  $-\infty$ .

Note that  $\mathcal{S}(T)$  is equal to the intersection of  $\mathcal{S}(T^+)$  and  $\mathcal{S}(T^-)$ , the convex hulls we are seeking to separate. From [12], we know that  $\mathcal{V}^+$  and  $\mathcal{V}^-$  being disjoint is equivalent to  $\rho(T) < 0$ .

Given  $a$  the eigenvector corresponding to  $\rho(T)$ , we define the sectors:

$$I^\pm := \{i \in \{1, \dots, d\}, \quad T^\pm(a)_i > \rho(T) + a_i\},$$

and the corresponding configuration  $\sigma := \{I^+, I^-\}$ .

We first state the following result, which will be demonstrated in the multiclass case in Lemma 5:

**Lemma 1.** *If a tropical hyperplane of apex  $\alpha$  separates our two data point clouds with a margin of at least  $-\mu > 0$ , then  $T(\alpha) \leq \alpha + \mu$ .*

We now characterize the maximal margin separating hyperplane using  $T$ :

**Proposition 2.** *The tropical hyperplane of apex  $a$  and configuration  $\sigma$ , given the sectors defined above, separates  $\mathcal{V}^+$  and  $\mathcal{V}^-$  with a margin of at least  $-\rho(T)$ , if and only if  $\rho(T) < 0$ . Moreover, this margin is optimal among all tropical hyperplanes when separating finite point clouds.*

*Proof.* For instance, let  $i \notin I^+$ . As  $T^+$  is non-expansive, for  $v^+ \in \mathcal{V}^+$ :

$$v_i^+ \leq T^+(v^+)_i = (T^+(v^+) - T^+(a))_i + T^+(a)_i,$$

hence

$$v_i^+ \leq \max(v^+ - a) + T^+(a)_i.$$

As  $i \notin I^+$ ,  $T^+(a)_i \leq \rho(T) + a_i$ , and

$$v_i^+ - a_i \leq \max(v^+ - a) + \rho(T).$$

In particular,  $v_i^+ - a_i < \max(v^+ - a)$  and any element of  $\mathcal{V}^+$  belongs to a sector of  $I^+$  with respect to  $\mathcal{H}_a$ , from which the sectors are well-defined. Finally,

$$d_H(\mathcal{S}^-, v^+) = \max(v^+ - a) - \max(v^+ - a)_{I^-} \geq -\rho(T),$$

(and the same results hold for the distance to  $\mathcal{S}^0$ ), hence the margin is at least  $-\rho(T)$  (which is always positive in the separable case, as shown in [12], making our sufficient condition for separability also necessary).

Moreover, when  $T^\pm$  are tropical projections, for  $i \in I^-$ , we can find  $x^+ \in \mathcal{V}^+$  such that  $x_i^+ - \max(v - a) = T^+(a)_i$ , giving us  $x_i^+ - a_i - \max(x^+ - a) = \rho(T)$ . Maximizing over all  $i \in I^-$  gives that  $x^+$  is at distance  $-\rho(T)$  of  $\mathcal{S}^-$ , hence the margin.

Finally, combining Equation 2.1 and Lemma 1,  $-\rho(T)$  is the optimal margin in the binary setting.  $\square$

*Remark 3.* To save on calculations, especially in large dimensions, we do not evaluate the coordinates associated with unassigned sectors  $I^0$ .

We can also prove similarly to above that if  $i$  is a positive sector, there is a witness point  $x$  from positive class at distance at most  $-\rho(T)$  from sector  $i$ . In a context of learning, this shows that excluding sectors of  $I^\pm$  that do not contain training data points can be unsafe as we have no guarantee that no testing points can fall into such sectors, or that a noise, even smaller than  $-\rho(T)$ , might move points there. By contrast, we do have the guarantee that excluding sectors  $i$  that do not belong to  $I^+$  or  $I^-$  is safe, as no training data point lies at distance less than  $-\rho(T)$  from such sectors.

Ignoring some sectors in our classifier is equivalent to setting corresponding apex coordinates to  $+\infty$ , which only increases the margin. For example, for  $x^+ \in X^+$ , and  $i$  a sector not assigned to the positive class,

$$d_H(\mathcal{S}^i, x^+) = \max(x^+ - a) - (x_i^+ - a_i)$$

is an increasing quantity in  $a_i$ , as the max is not reached in the  $i$  coordinate. Lastly, this also applies to the multiclass case presented in Section 4.

To build a practical algorithm, we still need to compute the spectral radius of  $T$  and its corresponding eigenvector. As it amounts to solving mean-payoff games, we can apply the Krasnoselskii-Mann iteration scheme described in Algorithm 1.

---

**Algorithm 1** Krasnoselskii-Mann iteration scheme for finding Shapley eigenpairs

---

- 1: **input:** A Shapley operator  $T$  and a relative convergence criterion  $\eta > 0$ .
  - 2:  $x := \mathbf{1}_d \in \mathbb{R}_{\max}^d$
  - 3: **repeat**
  - 4:    $z := (x + T(x)) / 2$
  - 5:    $x := z - \max(z) \cdot \mathbf{1}_d$
  - 6: **until** relative change of  $x$  is lower than  $\eta$
  - 7:  $a := x - \bar{x}$ ;  $\lambda := 2 \max(z)$
  - 8: **return**  $(a, \lambda)$
- 

**3.2. Geometric approach for soft-margin SVM.** We seek to classify convex hulls  $\mathcal{V}^+$  and  $\mathcal{V}^-$ . Assuming that the data overlap, we want to have a sense of their non-separability. Using the operator  $T$  we previously defined,  $\rho(T)$  is the (positive) inner radius of  $\mathcal{S}(T)$ , i.e. the supremum of the radii of the Hilbert balls contained in the intersection between  $\mathcal{V}^+$  and  $\mathcal{V}^-$ , as shown by Allamigeon et al. (2018) [12] and as seen in Figure 3.1.

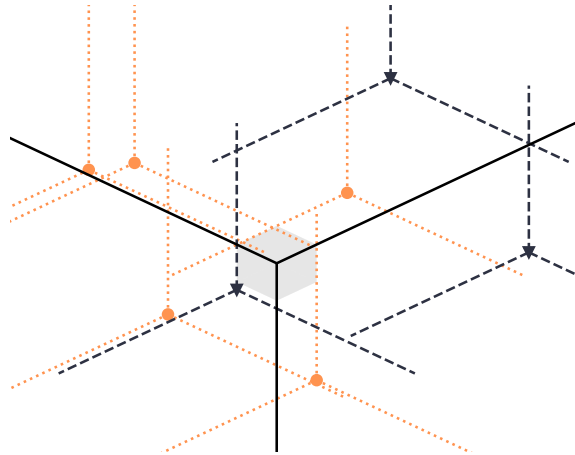


FIGURE 3.1. The spectral radius is the inner radius of the intersection between tropical spans

The next Proposition shows us that this measure can be interpreted geometrically, in the case where  $\mathcal{V}^+$  and  $\mathcal{V}^-$  are generated by point clouds  $X^+$  and  $X^-$ :

quote a reference  
& talk about  
efficiency

**Proposition 4.** *By moving some points from  $X^+$  and  $X^-$  by a distance of at most  $\rho(T)$ , we can nullify the interior of the intersection of the tropical spans.*

*More specifically, we note  $a \in \mathbb{R}_{\max}^d$  the eigenvector corresponding to  $\rho(T)$ , and  $\mathcal{H}_a$  the (unparameterized) tropical hyperplane as defined by equation 1.1. We project all points of  $X^+$  and  $X^-$  whose distance to  $\mathcal{H}_a$  is less than  $\rho(T)$ , onto  $\mathcal{H}_a$ . Then the intersection of new convex hulls is of empty interior.*

*Proof.* The distance between any point  $x$  in  $\mathbb{R}_{\max}^d$  and the tropical hyperplane of apex  $a$  is:

$$d_H(x, \mathcal{H}_a) := \max(x - a) - \max_2(x - a),$$

where  $\max_2$  stands for the second maximal value of any vector.

Considering  $X$  the cloud consisting of all points (regardless of sign), and for  $x_j \in X$ , we note  $s_j$  its sector and  $d_j$  the second argmax of  $(x_j - a)$ . For each point  $x_j = X_{\cdot j} \in X$  at distance less than  $\rho(T)$  of  $\mathcal{H}_a$ , the transformation described above consists in setting

$$W_{kj} := \begin{cases} X_{kj} & \text{if } k \neq s_j \\ X_{s_j j} - d_H(x_j, \mathcal{H}_a) & \text{at } s_j \end{cases}$$

so that  $w_j$  is projected on the hyperplane. As  $T^\pm$  is non-expansive and diagonal-free, Let us remark that for  $x^\pm \in V^\pm$ :

$$x_i^\pm \leq T^\pm(x^\pm)_i = (T^\pm(x^\pm) - T^\pm(a))_i + T^\pm(a)_i,$$

hence

$$(3.1) \quad x_i^\pm \leq \max(x^\pm - a)_{\neq i} + T^\pm(a)_i.$$

Let  $1 \leq i \leq d$ . If  $x_j \in X$  is not in the  $i$ -th sector, then for  $k$  different from the sector of  $x_j$ , by definition:

$$(w_j - a)_k \leq (x_j - a)_{d_j} \leq (w_j - a)_{s_j},$$

hence

$$W_{ij} - \max_{k \neq i} (W_{kj} - a_k) \leq a_i.$$

Otherwise,  $x_j$  is in the  $i$ -th sector and:

$$\max_{k \neq i} (W_{kj} - a_k) = X_{d_j j} - a_{d_j},$$

thus

$$W_{ij} - \max_{\neq i} (w_j - a) = (w_j - a)_{s_j} - (x_j - a)_{d_j} + a_{s_j} \geq a_{s_j} = a_i,$$

with equality iff  $d_H(x_j, \mathcal{H}_a) \leq \rho(T)$ .

Suppose by symmetry that  $T(a)_i = T^+(a)_i = \rho(T) + a_i$ . We also have  $T^-(a)_i \geq \rho(T) + a_i$ . Then, using the proof of Theorem 22 in [2], we know that there is  $j^+, j^-$  in  $\{1, \dots, p\}$  such that  $x_{j^+} \in X^+$  and  $x_{j^-} \in X^-$  are in sector  $i$ , with  $x_{j^+}$  being at distance  $\rho(T)$  from  $\mathcal{H}_a$  and  $x_{j^-}$  at distance greater than  $\rho(T)$ . Therefore,  $W_{ij^+} - \max_{\neq i} (w_{j^+} - a) = a_i$  and  $W_{ij^-} - \max_{\neq i} (w_{j^-} - a) \geq a_i$ . Moreover, for any  $j$  such that  $x_j \in X^+$  is in sector  $i$ , equation 3.1 gives  $d_H(x_j, \mathcal{H}_a) \leq \rho(T)$ .

Let  $Q^\pm$  be the *diagonal-free* projections over transformed point clouds, and  $Q = \inf(Q^+, Q^-)$ . We have just shown that  $Q(a)_i = Q^+(a)_i = a_i$ , and finally  $Q(a) = a$ .  $\square$

Thus, the spectral radius can be interpreted as a bound on the distance under which some points have to be moved to separate the interiors of convex hulls.

As  $a$  is the center of the inner ball of  $\mathcal{V}^+ \cap \mathcal{V}^-$ , we also have a purely geometric heuristic for determining a hyperplane in the non-separable case: we simply take the hyperplane  $\mathcal{H}$  of apex  $a$  and then assign the sectors based on dominant population, for instance.

#### 4. MULTICLASS CLASSIFICATION

In this section, we consider convex hulls of  $n$  point classes, noted  $(\mathcal{V}^k)_{1 \leq k \leq n}$  and described by Shapley operators  $T^k$ . We give a necessary and sufficient condition for these classes to be tropically separable. We then characterize the margin as another spectral radius in this case.

We now consider the Shapley operator

$$T := \max_{1 \leq k < l \leq n} \min(T^k, T^l),$$

which can be rewritten, for any  $x$  in  $\mathbb{R}_{\max}^d$  and  $i$  in  $\{1, \dots, d\}$ :

$$(4.1) \quad T(x)_i := \max_2 \{T^k(x)_i, 1 \leq k \leq n\}.$$

Morally, we can think of  $\mathcal{S}(T)$  as the union of pairwise intersections of data classes, although the max operator does not strictly speaking correspond to a union. We also remark that when  $n = 2$ ,  $T$  is the same as previously defined.

Let  $(a, \rho(T))$  be the eigenpair of  $T$  approximated by the Krasnoselskii-Mann algorithm, verifying  $T(a) = \rho(T) + a$ . We define the sectors:

$$(4.2) \quad I^k := \{1 \leq i \leq d, \quad T^k(a)_i > \rho(T) + a_i\}.$$

Given the definition of  $T$ , it is clear that these sets are disjoint. As in the binary case,  $T$  is linked with separating tropical hyperplanes by the following result:

**Lemma 5.** *If a tropical hyperplane of apex  $\alpha$  separates our finite point clouds with a margin of at least  $-\mu > 0$ , then  $T(\alpha) \leq \alpha + \mu$ .*

*Proof.* We consider two different classes  $k$  and  $\ell$  in  $\{1, \dots, n\}$ , and the sectors  $\sigma' := \{I'^j\}_{1 \leq j \leq n}$  defined by  $H_\alpha$ . Then, by definition, for  $x \in X^k$ ,  $(x - \alpha)$  reaches its maximum coordinate in  $I'^k$  and for any  $i$  in  $I'^\ell$ ,  $d_H(S^i, x) \geq -\mu$ , which can be rewritten as  $x_i - \max(x - \alpha)_i \leq \alpha_i + \mu$ . Maximizing over all  $x$  in  $X^k$  then yields  $T^k(\alpha)_i \leq \alpha_i + \mu$ . As this holds for all  $k \neq \ell$  at least, we have  $T(\alpha)_i \leq \alpha_i + \mu$ , (which also holds for  $\ell = 0$ , i.e when we consider unassigned sectors) we conclude that  $T(\alpha) \leq \alpha + \mu$ .  $\square$

**Proposition 6.** *The tropical parameterized hyperplane of apex  $a$  and configuration  $\sigma$ , given the configuration  $\sigma = \{I^k\}_{1 \leq k \leq n}$ , separates all point clouds with a margin of at least  $-\rho(T)$ , if and only if  $\rho(T) < 0$ . Moreover, this margin is optimal among all tropical hyperplanes.*

*Proof.* We consider class  $k \in \{1, \dots, n\}$  and sector  $\mathcal{S}^\ell$  with  $\ell \neq k$ . For  $j \in I^\ell$ , we have  $T^k(a)_i \leq a_i + \rho(T)$  and therefore can use the same reasoning as in the proof of Proposition 2 to obtain that for all  $v^k$  in  $\mathcal{V}^k$ ,

$$d_H(\mathcal{S}^\ell, v^k) = \max(v^k - a) - \max_{j \in I^\ell} (v^k - a)_j \geq -\rho(T),$$

and  $v^k$  cannot belong to sector  $\mathcal{S}^\ell$ , from which, as again, the sectors are well-defined.  $\square$

Since this method optimizes the worst margin, it works best on previously balanced data. If necessary, it is still possible to use the binary setting of the previous section with one-vs-all approaches, in order to maximize all margins 2 by 2.

*Remark 7.* According to Remark 3, we extend the sector assignment policy as follows in the non-separable case. Relative to some apex  $a$ , the sector corresponding to the maximum coordinate  $i$  is assigned to the class with the greatest number of points at distance  $\max(0, -\rho(T))$  from it. Unassigned sectors are removed for more sparsity, which preserves the margin in the separable case.

**Conclusion.** We therefore introduce Algorithm 2 to solve the tropical linear classification problem.

---

#### Algorithm 2 Determining tropical linear classifiers by solving mean-payoff games

---

- 1: **input:** Data classes  $X^1, \dots, X^n$  corresponding to labels  $1, \dots, n$ , and a relative convergence criterion  $\eta$  for the iteration scheme.
  - 2:  $T := \max_2 (T^k)_{1 \leq k \leq n}$
  - 3:  $(a, \lambda) := \text{KRASNOSELSKII-MANN}(T, \eta)$
  - 4: **assign** sectors  $\sigma$  to classes based on Remark 7
  - 5: **return** classifier  $H_a^\sigma$  and margin  $\max(0, -\lambda)$
-

In the separable case, we have margin optimality. In the non-separable case, by Proposition 4, the overlap zone is contained around the (non-parameterized) hyperplane  $H_a$  in a tubular envelope of radius  $\lambda$ , the (positive) eigenvalue approximated by the algorithm.

Let us analyze the complexity of Algorithm 2. Step 3 typically involves a few dozen iterations, in which  $n$  tropical projections have to be calculated each time, for a total cost in  $\mathcal{O}(dp)$ , where  $p$  is the total number of points in the dataset. We therefore have a pseudo-polynomial algorithm for calculating an  $\varepsilon$ -approximation of the classifier described above.

This gain in efficiency is achieved at the cost that tropical SVMs are ultimately very sensitive to outliers, which count as much as any point in the margin calculation.

**Why not tropicalizing classical SVMs?** Tropical hyperplanes can be obtained as the limit of classical hyperplanes seen on log-log paper, thanks to what we call *Maslov dequantization* [7] or Viro's method [8]. A natural way of doing Tropical SVM is therefore to take our data to the power  $\beta > 0$ , apply a classical SVM (which would maximize the margin in the new space), then go back to the logarithm. As  $\beta$  tends towards infinity, we are approaching some limiting tropical hyperplane. However, this method is a source of significant numerical error, especially when  $\beta$  is large. Moreover, it does not guarantee a good margin in the starting space, as illustrated in Figure 4.1, where the hypersurface in yellow tends towards the suboptimal tropical one in red.

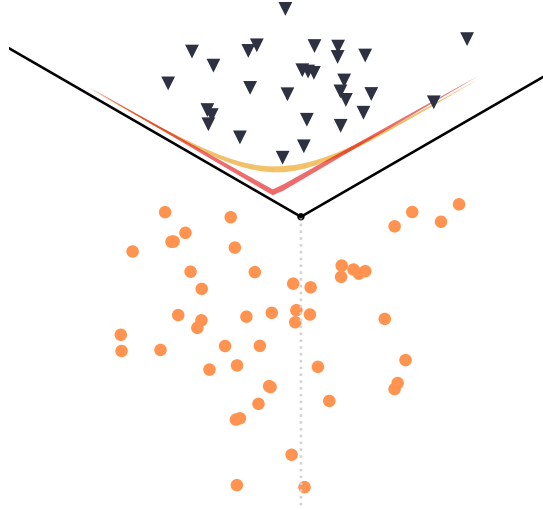


FIGURE 4.1. Tropicalized hyperplane tends towards a suboptimal tropical hyperplane

## 5. TROPICAL POLYNOMIALS AS PIECEWISE LINEAR CLASSIFIERS

In this section, we extend our previous approach to piecewise linear fitting, using tropical polynomials.

A *tropical polynomial function* over  $\mathbb{R}_{\max}^d$  is defined by

$$f(x) = \bigoplus_{\alpha \in \mathbb{Z}^d} f_{\alpha} x^{\alpha} = \max_{\alpha \in \mathbb{Z}^d} (f_{\alpha} + \langle x, \alpha \rangle),$$

where  $f_{\alpha} = -\infty$  except for finitely many values of  $\alpha$ , and with  $\langle x, \alpha \rangle = \sum_i x_i \cdot \alpha_i$  under the convention  $(-\infty) \cdot 0 = 0$ . The *tropical hypersurface* associated with  $f$  is the set of points from  $\mathbb{R}_{\max}^d$  where the maximum in  $f$  is attained twice. Tropical hypersurfaces are piecewise linear and divide  $\mathbb{R}_{\max}^d$  between sectors corresponding to the maximal monomial: they can perform more complex classification.

If  $\mathcal{A} \subset \mathbb{Z}^d$  is a set of vectors, we define the *Veronese embedding* of  $x$  as:

$$\text{ver}_{\mathcal{A}}(x) := (\langle x, \alpha \rangle)_{\alpha \in \mathcal{A}} \in \mathbb{R}_{\max}^{\mathcal{A}},$$

allowing us to map our point space into a larger space, made up of various integer combinations of features. To take a fairly exhaustive subset of monomials, we consider the integer combinations defined by the integer points of the dilated simplex. Noting  $s \in \mathbb{N}^*$  a scale parameter, we define

$$\mathcal{A}_s := \{x \in \mathbb{N}^d, \quad x_1 + \cdots + x_d = s\},$$



which grows polynomially with  $s$ . Note that we choose homogeneous monomials to preserve the translation invariance property in the base space. The Veronese embedding defined by  $\mathcal{A}_s$  is invertible and if  $\mathcal{H}_a$  is a tropical hyperplane of apex  $a$  in  $\mathbb{R}_{\max}^A$ ,  $\text{ver}_{\mathcal{A}_s}^{-1}(\mathcal{H}_a)$  is a tropical polynomial hypersurface in the initial space, corresponding to the polynomial

$$f_{-a, \mathcal{A}_s}(x) = \bigoplus_{\alpha \in \mathcal{A}_s} (-a_\alpha) x^\alpha$$

Hence,  $\text{ver}_{\mathcal{A}_s}$  is a feature map enabling us to perform piecewise linear classification in  $\mathbb{R}_{\max}^d$ . This brings us close to the framework of neural networks with piecewise linear activations, e.g. relaxed linear units [13].  $\mathcal{A}_1$  corresponds to tropical hyperplanes: we are generalizing the approach previously developed.

The initial metric can be easily linked to the metric in the augmented space: for  $x \in \mathbb{R}_{\max}^d$ , noting  $V$  the (injective) Veronese embedding, we have

$$\|(V^T V)^{-1} V^T\|_{\text{op}, \infty}^{-1} \leq \frac{\|\text{ver}(x)\|_{H, \mathbb{R}_{\max}^{\mathcal{A}_s}}}{\|x\|_{H, \mathbb{R}_{\max}^d}} \leq \|V\|_{\text{op}, \infty},$$

which gives, using  $\mathcal{A}_s$ , according to Appendix A:

$$\left[ \frac{d}{s+d} + \frac{sd^2}{(s+d)^2} \right]^{-1} \leq \frac{\|\text{ver}(x)\|_{H, \mathbb{R}_{\max}^{\mathcal{A}_s}}}{\|x\|_{H, \mathbb{R}_{\max}^d}} \leq s.$$

We deduce the following result:

**Proposition 8.** *Noting  $(a, \rho(T))$  the eigenpair of  $T$  applied to new point clouds in  $\mathbb{R}_{\max}^A$ , data is separable in the initial space by the tropical hypersurface of  $f_{-a, \mathcal{A}}$  with a margin of at least  $-\rho(T)/\|V\|_{\text{op}, \infty}$ , if and only if  $\rho(T) < 0$ .*

In Figures 1.1b, 1.1c and 1.1d, we visualize the decision boundaries of the piecewise linear classifier using monomials from  $\mathcal{A}_3$ . Dotted lines in light gray indicate merged sectors after assignment, and margin lower estimation is represented by the Hilbert balls around each point. These methods obviously work in higher dimensions, but we can only visualize them in lower ones.

**Complexity analysis.** The learning speed depends essentially on the dimension of the feature space, i.e. the starting dimension and the number of monomials considered. Although our iteration scheme is efficient, given that

$$|\mathcal{A}^s| = \binom{s+d-1}{s},$$

the complexity grows very quickly, opening the way to feature selection questions for directly learning *sparse* polynomials.

On the other hand, when the tropical polynomial is trained, since only active monomials are kept, inference is in  $\mathcal{O}(rd)$  only, where  $r$  is the number of monomials kept.  $r$  is always lower than the number of data points, and it is equal to it when every sector contains exactly one point.

**Feature selection.** A heuristic for adaptive monomial choice might be to sample pairs of points of different classes, and make perpendicular slopes admissible. In many cases, this gives better results (and even better margins), as we can see in Figure 5.1, in which we allow for homogeneous positive real monomials. Here, points are sampled randomly, but we may perform better by choosing a more representative sample, with points relatively distant from each other, for example.

## 6. CONCLUDING REMARKS

We have formulated a pseudo-polynomial algorithm to determine a classifying tropical hyperplane, with geometric guarantees. It would be interesting to conduct further work on monomial selection heuristics, to better separate complex datasets in the real world.

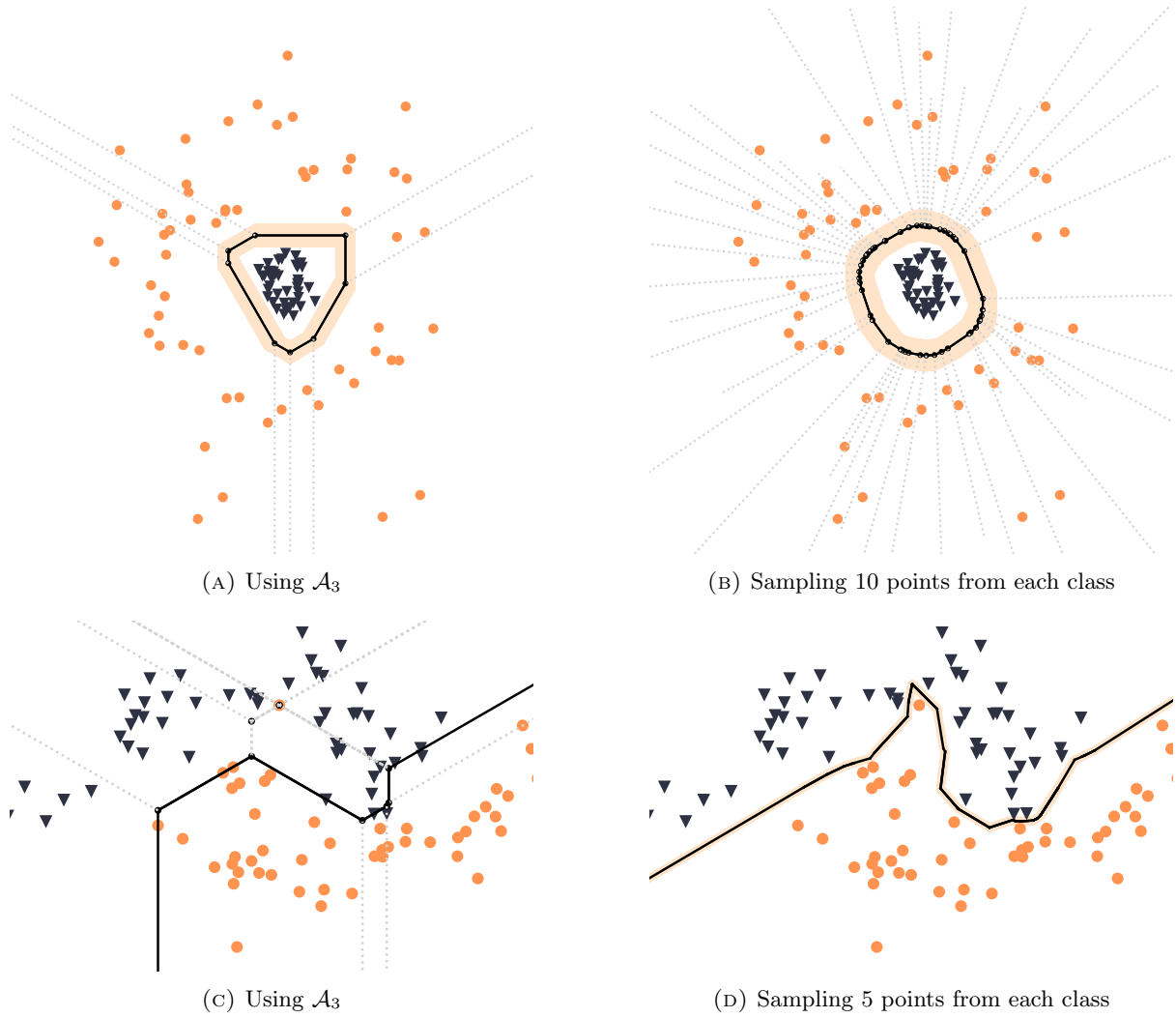


FIGURE 5.1. Feature engineering might yield better results

## REFERENCES

- [1] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92*, page 144–152, New York, NY, USA, 1992. Association for Computing Machinery.
- [2] Marianne Akian, Stéphane Gaubert, Yang Qi, and Omar Saadi. Tropical linear regression and mean payoff games: or, how to measure the distance to equilibria. *SIAM J. Discret. Math.*, 37:632–674, 2021.
- [3] Petros Maragos, Vasileios Charisopoulos, and Manos Theodosis. Tropical geometry and machine learning. *Proceedings of the IEEE*, PP:1–28, 04 2021.
- [4] Robert Page, Leon Zhang, and Ruriko Yoshida. Tropical principal component analysis on the space of ultrametrics, 2019.
- [5] Bernd Gärtner and Martin Jaggi. Tropical support vector machines. Technical report, ACS-TR-362502-01.
- [6] Xiaoxian Tang, Houjie Wang, and Ruriko Yoshida. Tropical support vector machine and its applications to phylogenomics, 2020.
- [7] G. L. Litvinov. The maslov dequantization, idempotent and tropical mathematics: A brief introduction, 2005.
- [8] Oleg Viro. Dequantization of real algebraic geometry on logarithmic paper, 2000.
- [9] Li Zhang and Weida Zhou. On the sparseness of 1-norm support vector machines. *Neural Networks*, 23(3):373–385, 2010.
- [10] Diane Maclagan and Bernd Sturmfels. *Introduction to Tropical Geometry*. American Mathematical Society, Apr 2015.
- [11] Marianne Akian, Stéphane Gaubert, and Alexander Guterman. Tropical polyhedra are equivalent to mean payoff games. *International Journal of Algebra and Computation*, 22(01):1250001, Feb 2012.

- [12] Xavier Allamigeon, Stephane Gaubert, Ricardo D. Katz, and Mateusz Skomra. Condition numbers of stochastic mean payoff games and what they say about nonarchimedean semidefinite programming. 2018.
- [13] Liwen Zhang, Gregory Naitzat, and Lek-Heng Lim. Tropical geometry of deep neural networks, 2018.

## APPENDIX A. NORM EQUIVALENCE BETWEEN VERONESE AND STARTING SPACES

Firstly,

$$\begin{aligned}\|\text{ver}(x)\|_{H, \mathbb{R}_{\max}^{\mathcal{A}_s}} &= \max_{\alpha \in \mathcal{A}_s} \sum_i \alpha_i x_i - \min_{\alpha \in \mathcal{A}_s} \sum_i \alpha_i x_i \\ &\leq s(\max x - \min x) = s\|x\|_{H, \mathbb{R}_{\max}^d}.\end{aligned}$$

In general, using

$$\|Y\|_H = 2 \inf_{\alpha \in \mathbb{R}} \|Y + \alpha e\|_\infty,$$

and denoting  $V$  the Veronese (linear) embedding, knowing that

$$\|VY\|_\infty \leq \|V\|_{\text{op}, \infty} \|Y\|_\infty,$$

we have the first side.

The other inequality is obtained using the pseudo-inverse of (non-square)  $V$ .

$V$  is a  $\ell_{s,d} \times n$  matrix, where  $\ell_{s,d}$  is the cardinal of  $\mathcal{A}^s$ , i.e. the number of homogenous polynomials of degree  $d$  with  $n$  variables:

$$\ell_{s,d} = \binom{s+d-1}{d-1}.$$

Rows of  $V$  are points from the external face of the dilated simplex. As the latter is invariant by permutating axes, we have

$$V^T V = (V_i^T V_j)_{ij} = \begin{pmatrix} \alpha + \beta & & \beta \\ & \ddots & \\ \beta & & \alpha + \beta \end{pmatrix} = \alpha I_{d \times d} + \beta \mathbf{1}_{d \times d},$$

where  $\alpha + \beta = \|V_{\cdot 1}\|^2$  and  $\beta = \langle V_{\cdot 1}, V_{\cdot 2} \rangle$ . We find that:

$$\alpha = \binom{s+d}{s-1} \quad \text{and} \quad \beta = \binom{s+d-1}{s-2},$$

and as  $\mathbf{1}_{d \times d}$  is of rank 1,  $V^T V$  is easy to invert, and

$$(V^T V)^{-1} = \alpha^{-1} (I - \beta \alpha^{-1} (1 + d\beta \alpha^{-1})^{-1} \mathbf{1}_{d \times d}),$$

As  $\mathbf{1}_{d \times d} V^T = s \mathbf{1}_{d \times \ell_{s,d}}$ ,

$$M := (V^T V)^{-1} V^T = \alpha^{-1} (I - B \mathbf{1}_{d \times \ell_{s,d}}),$$

where

$$B = s \cdot \beta \alpha^{-1} (1 + d\beta \alpha^{-1})^{-1} = \frac{s-1}{d+1}.$$

is a  $d \times \ell_{s,d}$  matrix whose lines are permutations of each other. Hence, their  $L^1$  norms are equal and:

$$\|M\|_{\text{op}, \infty} = \alpha^{-1} \sum_{i=1}^d |V_{i1} - B| = \alpha^{-1} \sum_{k=0}^s \ell_{s-k, d-1} |k - B|.$$

Splitting this sum with respect to the sign of  $k - B$  yields, noting  $b = \lfloor B \rfloor$ :

$$\begin{aligned}\|M\|_{\text{op}, \infty} &= \alpha^{-1} \left( \sum_{k=0}^b \ell_{s-k, d-1} (B - k) + \sum_{k=b+1}^s \ell_{s-k, d-1} (k - B) \right) \\ &\leq \alpha^{-1} \left( \sum_{k=0}^b \binom{s-k+d-2}{d-2} B + \sum_{k=b+1}^s \binom{s-k+d-2}{d-2} (s - B) \right)\end{aligned}$$

Choosing  $p+1$  items among  $q+1$  ones amounts to first choosing the maximum, and then the  $n$  remaining elements below it. Hence

$$\sum_{k=p}^q \binom{k}{p} = \binom{q+1}{p+1},$$

and

$$\sum_{b+1 \leq k \leq s} \binom{s-k+d-2}{d-2} = \binom{s-b+d-2}{d-1},$$

and similarly

$$\sum_{0 \leq k \leq b} \binom{s-k+d-2}{d-2} = \ell_{s,d} - \binom{s+d-b-2}{d-1}.$$

Hence

$$\|M\|_{\text{op},\infty} \leq \alpha^{-1} \left[ B\ell_{s,d} + (s-2B) \binom{s-b+d-2}{d-1} \right],$$

but

$$B\alpha^{-1}\ell_{s,d} = \frac{(s-1)d}{s(s+d)}, \quad s-2B = \frac{s(d-1)+2}{d+1},$$

and

$$\begin{aligned} \frac{\binom{s-b+d-2}{d-1}}{\binom{s+d}{d+1}} &= d(d+1) \frac{(s-b+d-2)(s-b+d-3) \cdots (s-b)}{(s+d)(s+d-1) \cdots (s+1)s} \\ &= \frac{d(d+1)}{(s+d-1)(s+d)} \frac{s-b}{s} \frac{s+1-b}{s+1} \cdots \frac{s+d-2-b}{s+d-2} \\ &\leq \frac{d(d+1)}{(s+d-1)(s+d)}, \end{aligned}$$

as  $b \leq s$ . Finally, for  $s \geq 2$ :

$$\begin{aligned} \|(V^T V)^{-1} V^T\|_{\text{op},\infty} &\leq \frac{(s-1)d}{s(s+d)} + \frac{d(s(d-1)+2)}{(s+d-1)(s+d)} \\ &\leq \frac{d}{s+d} + \frac{sd^2}{(s+d)^2} \end{aligned}$$