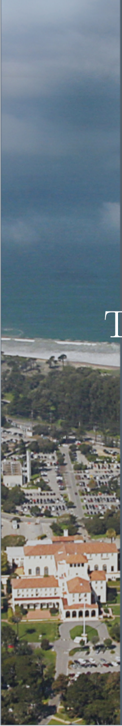


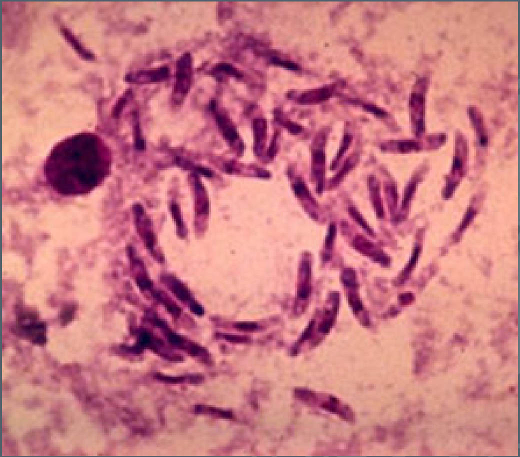
Plücker Coordinates of the best-fit Stiefel Tropical Linear Space to a Mixture of Gaussian Distributions

Ruriko Yoshida

Naval Postgraduate School

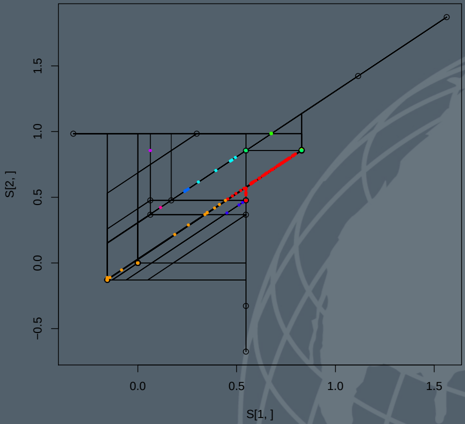
September, 2022

Apicomplexa



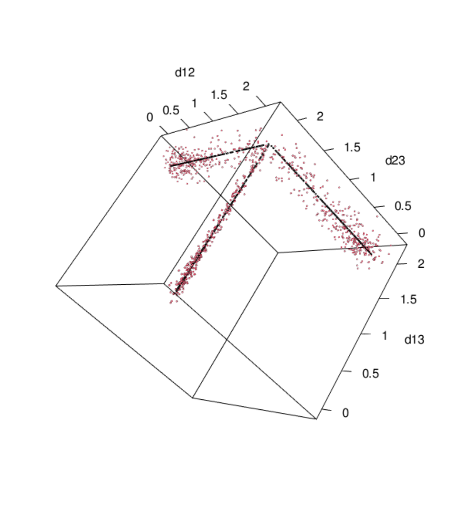
The Apicomplexa (also called Apicomplexia) are a large phylum of parasitic alveolates including malaria and Toxoplasma.

Apicomplexa



Visialization via tropical PCA (Yoshida, et al. 2019 and Page et al. 2020).

Phylogenetic Tree Data



Max-plus algebra

Here we use max-plus algebra, i.e.,

$$a \oplus b := \max\{a, b\} \text{ and } a \odot b := a + b.$$

Tropical Distance

Max-plus algebra

Here we use max-plus algebra, i.e.,

$$a \oplus b := \max\{a, b\} \text{ and } a \odot b := a + b.$$

Tropical distance

The tropical distance between two points is computed as follows:

$$d_{\mathrm{tr}}(v, w) = \max\{|v_i - w_i - v_j + w_j| : 1 \leq i < j \leq e\}.$$

This metric is also known as the generalized Hilbert projective metric.

Tropical Distance

Max-plus algebra

Here we use max-plus algebra, i.e.,

$$a \oplus b := \max\{a, b\} \text{ and } a \odot b := a + b.$$

Tropical distance

The tropical distance between two points is computed as follows:

$$d_{\mathrm{tr}}(v, w) = \max\{|v_i - w_i - v_j + w_j| : 1 \leq i < j \leq e\}.$$

This metric is also known as the generalized Hilbert projective metric.

Example

$$v = (0, 0, 0), \; w = (0, 3, 1), \; d_{\mathrm{tr}}(v, w) = 3.$$

Definition

If A is a $d \times d$ square matrix, we can also define its *tropical determinant* in analogy with the classical operation. We have

$$\operatorname{tdet} A = \bigoplus_{\sigma \in S_d} \left(\bigodot_{i=1}^{d} a_{i,\sigma(i)} \right).$$

If the tropical determinant of A is attained by at least two distinct permutations in S_d, we say that A is *tropically singular*.

Definition

If A is a $d \times d$ square matrix, we can also define its *tropical determinant* in analogy with the classical operation. We have

$$\operatorname{tdet} A = \bigoplus_{\sigma \in S_d} \left(\bigodot_{i=1}^{d} a_{i,\sigma(i)} \right).$$

If the tropical determinant of A is attained by at least two distinct permutations in S_d, we say that A is *tropically singular*.

Example

$$A = \begin{pmatrix} 0 & 2 & 2 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix}.$$

The tropical determinant of A is 4.

Definition

Let A be a tropical $d \times e$ matrix. Given a d-sized subset $\omega \subseteq [e]$, we write A_ω for the $d \times d$ matrix whose columns are the columns of A indexed by elements of ω. Then the map

$$p : [e]^d \to \mathbb{R} \cup \{-\infty\}$$

$$\omega \mapsto \operatorname{tdet}(A_\omega)$$

is a tropical Plücker vector. The corresponding tropical linear space is called the ***Stiefel tropical linear space*** given by A.

Tropical linear spaces

Definition

Let A be a tropical d e matrix. Given a d-sized subset $!$ $[e]$, we write $A_!$ for the d d matrix whose columns are the columns of A indexed by elements of $!$. Then the map

$$p : [e]^d \;!\; \mathrm{R} \;[\; f \quad 1\, g$$

$$! \;7!\; \mathrm{tdet}\,(A_!)$$

is a tropical Plücker vector. The corresponding tropical linear space is called the *Stiefel tropical linear space* given by A.

Definition

Let $p : [e]^d$ $!$ R $[$ f 1 g be a tropical Plücker vector. The *tropical linear space* L_p consists of all points x 2 TP^{e} 1 such that, for any $(d+1)$-subset of $[e]$, the maximum of the numbers $p($ f ${}_i g) + x_i$, for $i = 1; : : : ; d+1$, is attained at least twice.

Example

Let

$$A = \begin{matrix} 0 & 2 & 4 \\ 0 & 1 & 3 \end{matrix}$$

and let p be its associated tropical Plücker vector. Then

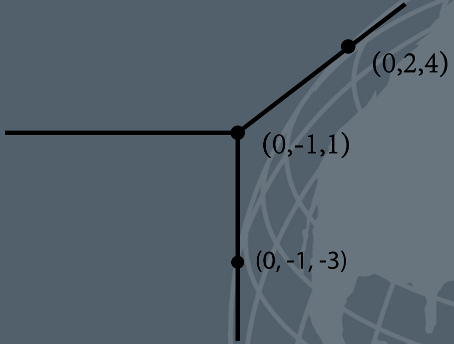
$$p(\{1,2\}) = \operatorname{tdet} \begin{matrix} 0 & 2 \\ 0 & 1 \end{matrix} = 2,$$

$$p(\{1,3\}) = \operatorname{tdet} \begin{matrix} 0 & 4 \\ 0 & 3 \end{matrix} = 4,$$

$$p(\{2,3\}) = \operatorname{tdet} \begin{matrix} 2 & 4 \\ 1 & 3 \end{matrix} = 3.$$

Tropical linear spaces

The Stiefel tropical linear space corresponding to A is a tropical line in R^3=R1.



Fitting to a Tropical Linear Space

- In order to fit to a tropical linear space, we project data points onto subspaces.

Fitting to a Tropical Linear Space

- In order to fit to a tropical linear space, we project data points onto subspaces.
- Let $L_{\mathbf{w}}$ be a *tropical linear space* of dimension $s \quad 1$ in R^d=R1 and $\mathbf{w} = (w \quad)$ be a vector in R^d=R1.

- In order to fit to a tropical linear space, we project data points onto subspaces.
- Let $L_{\mathbf{w}}$ be a *tropical linear space* of dimension $s \quad 1$ in R^d=R1 and $\mathbf{w} = (w \quad)$ be a vector in R^d=R1.
- If $L_{\mathbf{w}}$ is spanned by $x^1; : : : ; x^s$ in R^d=R1 then its Plücker coordinate $w \quad$ is the *tropical determinant* of the $s \quad s$-submatrix indexed by of the $s \quad d$-matrix $X \quad = (x^1; : : : ; x^s)$.

Fitting to a Tropical Linear Space

- In order to fit to a tropical linear space, we project data points onto subspaces.
- Let $L_{\mathbf{w}}$ be a *tropical linear space* of dimension $s \quad 1$ in $\mathrm{R}^d/\mathrm{R}1$ and $\mathbf{w} = (w \ )$ be a vector in $\mathrm{R}^d/\mathrm{R}1$.
- If $L_{\mathbf{w}}$ is spanned by $x^1;\ldots;x^s$ in $\mathrm{R}^d/\mathrm{R}1$ then its Plücker coordinate w is the *tropical determinant* of the $s \quad s$-submatrix indexed by of the $s \quad d$-matrix $X = (x^1;\ldots;x^s)$.

Theorem (The Blue Rule)

The i-th coordinate of the point in $L_{\mathbf{w}}$ nearest to u is equal to

$$_{L_{\mathbf{w}}}(u)_i \ = \ \max \ \min_{j2} \ u_j + w_{\ [i} \quad w_{\ [j} \ \text{for} \ i = 1;2;\ldots;d:$$

Here runs over all $(s \quad 1)$-subsets of $[d]$ that do not contain i.

Problem

Let $x_1, \dots, x_n$ be data points in $\mathbb{R}^d/\mathbb{R}\mathbf{1}$. The tropical regression plane of dimension $s-1$ is a solution to the optimization problem

$$\arg\min_{L_{\mathbf{w}}} \sum_{i=1}^{n} d_{\mathrm{tr}}(x_i, L_{\mathbf{w}}).$$

Here $\mathbf{w}$ runs over all points in $\mathbb{R}^d/\mathbb{R}\mathbf{1}$.

Theorem

Suppose $X = (\ _1 + \ _1; \ _2 + \ _2; \ _3; :::; \ _d) \ 2 \ \mathbb{R}^d/\mathbb{R}1$ *such that* $\ _1; \ _2 \ 2 \ \mathbb{R}$ *and* $\ _j \quad N(0; \ )$ *for* $j = 1; :::; d$. *Let* $X^0 \ 2 \ \mathbb{R}^d/\mathbb{R}1$ *be the projected point of* X *onto the one-dimensional Stiefel tropical linear space of the matrix*

$$A_1 = \begin{matrix} _1 & 1 & 0 & ::: & 0 \\ 1 & _2 & 0 & ::: & 0 \end{matrix} :$$

Then the tropical distance between X *and* X^0 *is*

$$d_{\mathrm{tr}}(X; X^0) = \max_{1 \ i \ d} (\ _i \qquad);$$

where $\quad$ *is the second smallest value in* $f \ _1; :::; \ _d g$, *and its expected value satisfies*

$$E[d_{\mathrm{tr}}(X; X^0)] \quad 2 \quad p \overline{2 \log(d)}:$$

One Gaussian Distribution

Theorem

Suppose $X = ({}_1 + {}_1, \ldots, {}_m + {}_m, {}_{m+1}, \ldots, {}_d) \in \mathbb{R}^d/\mathbb{R}1$ *such that* ${}_1, {}_2, \ldots, {}_m \in \mathbb{R}$ *and* ${}_j \;\; N(0, \;)$ *for* $j = 1, \ldots, d$. *Let* $X' \in \mathbb{R}^d/\mathbb{R}1$ *be the projected point of* X *onto the* $(m \;\; 1)$*-dimensional Stiefel tropical linear space of the* $m \;\; d$ *matrix* $A_{m \;\; 1}$,

$$A_{m \;\; 1} = \begin{pmatrix} {}_1 & \infty & \infty & \cdots & \infty & 0 & \cdots & 0 \\ \infty & {}_2 & \infty & \cdots & \infty & 0 & \cdots & 0 \\ \infty & \infty & {}_3 & \cdots & \infty & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \infty & \infty & \infty & \infty & {}_m & 0 & \cdots & 0 \end{pmatrix}. \quad (1)$$

Then

$$\lim_{\;\to 0} \mathrm{E}[d_{\mathrm{tr}}(X, X')] = 0.$$

Lemma

The Stiefel tropical linear space with the Plücker coordinates $P = (P_{12}; P_{13}; P_{23})$ *that passes through given two points* $= (\ _1; \ _2; \ _3)$ *and* $= (\ _1; \ _2; \ _3)$ 2 R^3=R1 *in a general position* ($_i$ $_i$ $\neq$ $_j$ $_j$ *for* 1 $i < j$ 3) *is unique.*

Mixture of Two Gaussians

Lemma

The Stiefel tropical linear space with the Plücker coordinates $P = (P_{12}; P_{13}; P_{23})$ that passes through given two points $= (\ _1;\ _2;\ _3)$ and $= (\ _1;\ _2;\ _3)$ 2 $\mathbb{R}^3 / \mathbb{R}1$ in a general position ($_i \quad _i \not 6 \quad _j \quad _j$ for 1 $\ i < j \ $ 3) is unique.

Theorem

The Stiefel tropical linear space with P_{ij} for 1 $\ i < j \ $ d that passes through given two points $= (\ _1;\ _2;\ _3; \ldots;\ _d)$ and $= (\ _1;\ _2;\ _3; \ldots;\ _d)$ 2 $\mathbb{R}^d / \mathbb{R}1$ in a general position ($_i \quad _i \not 6 \quad _j \quad _j$ for 1 $\ i < j \ $ d) is unique and obtained as the tropical determinant.

Mixture of Two Gaussians

In order to make it simple, suppose we have two random variables:

$$
\begin{aligned}
X_1 &= (5 + {}_{11}; \quad 5 + {}_{12}; \quad {}_{13}; \quad \ldots; \quad {}_{1d}); \\
X_2 &= (\quad 5 + {}_{21}; \quad 5 + {}_{22}; \quad {}_{23}; \quad \ldots; \quad {}_{2d});
\end{aligned}
$$

where ${}_{ij} \quad N(0; \quad)$ with small $\quad > 0$ for $i = 1; 2$ and $j = 1; \ldots; d$.

Let p_{A_0} be the Plücker coordinates of the Stiefel tropical linear space corresponding to the $2 \quad d$ matrix,

$$
A_0 = \begin{pmatrix} 5 & 5 & 0 & \ldots & 0 \\ 5 & 5 & 0 & \ldots & 0 \end{pmatrix}:
$$

Theorem

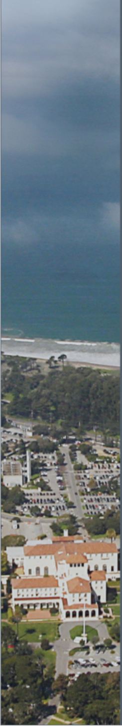
Suppose w is the projected point of either X_1 (or X_2) onto the Stiefel tropical linear space p_{A_0} and $P([_{i;j} fj \ _{ij}j \quad 5g)$ for $> 0; i = 1; 2$ and $j = 1; : : : ; d$. Then the expected value of the tropical distance between X_1 or X_2 and w is smaller than $2 \ \mathrm{p}\ \overline{2 \log(d \quad 1)}$ with the probability 1 .

Theorem

Suppose w is the projected point of either X_1 (or X_2) onto the Stiefel tropical linear space p_{A_0} and $P([\ _{i;j} fj\ _{ij} j \quad 5g)$ for $> 0; i = 1;2$ and $j = 1; : : : ; d$. Then the expected value of the tropical distance between X_1 or X_2 and w is smaller than $2\sqrt{2\log(d \quad 1)}$ with the probability $1 \quad$.

Remark

We also worked on a mixture of three or more Gaussians fitted by tropical polynomials over $R^3{=}R1$.



Thank you for your attention!

Questions?

Supported by NSF.

https://arxiv.org/abs/2112.11893