L-Infinity Optimization in Tropical Geometry and Phylogenetics

Daniel Irving Bernstein* and Colby Long

North Carolina State University
dibernst@ncsu.edu
https://arxiv.org/abs/1606.03702

October 19, 2016
Phylogenetics Motivation

- Find a tree representing the evolutionary history among a set of species from data
- Sometimes the data is a list of pairwise “distances” between species

Example
- Species: human, gorilla, banana
- Data:
  - $d(\text{human}, \text{gorilla}) = 1$
  - $d(\text{human}, \text{banana}) = 5$
  - $d(\text{gorilla}, \text{banana}) = 5$

What if our data were perturbed and we instead observed $d(\text{human}, \text{gorilla}) = 1$, $d(\text{human}, \text{banana}) = 5$, and $d(\text{gorilla}, \text{banana}) = 6$?
**Ultrametrics**

### Definition

A *dissimilarity map* on a set $X$ is a function $d : X \times X \rightarrow \mathbb{R}$ such that

1. $d(x, y) = d(y, x)$ and
2. $d(x, x) = 0$

We say that $d(\cdot, \cdot)$ is an *ultrametric* if for all $x, y, z \in X$ the maximum of $d(x, y), d(y, z)$ and $d(x, z)$ is attained twice.

### Example

Let $X = \{A, B, C, D\}$. The following is an ultrametric on $X$.

$$
\begin{pmatrix}
A & B & C & D \\
A & 0 & 5 & 7 & 9 \\
B & 5 & 0 & 7 & 9 \\
C & 7 & 7 & 0 & 9 \\
D & 9 & 9 & 9 & 0
\end{pmatrix}
$$
Proposition

Every ultrametric on a finite set can be expressed as a rooted tree and vice versa. The tree structure associated to an ultrametric $d$ is called the topology of $d$.
Which ultrametrics are $l^\infty$-nearest to the dissimilarity map below?

\[
\begin{array}{cccc}
A & B & C & D \\
\hline
A & 0 & 2 & 4 & 6 \\
B & 2 & 0 & 7 & 10 \\
C & 4 & 7 & 0 & 12 \\
D & 6 & 10 & 12 & 0 \\
\end{array}
\]
Finding $l^\infty$-Nearest Ultrametrics

**Theorem (Chepoi and Fichet 2000)**

Let $d$ be a dissimilarity map on a finite set $X$. Then the following algorithm produces an ultrametric on $X$ that is nearest to $d$ in the $l^\infty$ norm.

1. **Draw the complete graph on vertex set $X$**
2. **Label the edge between $x$ and $y$ by $d(x, y)$**
3. **For each $x, y \in X$ define**
   
   $$d_u(x, y) = \min_{\text{paths } P \text{ from } x \text{ to } y} \left( \max_{\text{edges } (i, j) \text{ of } P} d(i, j) \right)$$

4. **Let $\delta = \|d_u - d\|_\infty$ and let $1$ be the ultrametric such that $1(x, y) = 1$ for all $x, y \in X$**
5. **Then $d_u(x, y) + \frac{\delta}{2} 1$ is an ultrametric that is $l^\infty$-nearest to $d$.**
Example

\[ d = \begin{pmatrix} A & B & C & D \\ A & 0 & 2 & 4 & 6 \\ B & 2 & 0 & 7 & 10 \\ C & 4 & 7 & 0 & 12 \\ D & 6 & 10 & 12 & 0 \end{pmatrix} \]

\[ d_u = \begin{pmatrix} A & B & C & D \\ A & 0 & 2 & 4 & 6 \\ B & 2 & 0 & 4 & 6 \\ C & 4 & 4 & 0 & 6 \\ D & 6 & 6 & 6 & 0 \end{pmatrix} \]

\[ d_u + \frac{\|d_u - d\|_\infty}{2} \mathbf{1} = \begin{pmatrix} A & B & C & D \\ A & 0 & 5 & 7 & 9 \\ B & 5 & 0 & 7 & 9 \\ C & 7 & 7 & 0 & 9 \\ D & 9 & 9 & 9 & 0 \end{pmatrix} \]
Remark

There exist other ultrametrics that are also $l^\infty$ distance 3 from $d$. Some have different tree topologies.
Problems

Let $U(n)$ denote the set of ultrametrics on a set of size $n$. If $d$ is a dissimilarity map, let $C(d, U(n))$ denote the collection of ultrametrics that are $l_\infty$-nearest to $d$.

**Proposition (B.-Long 2016)**

$C(d, U(n))$ is a tropical polytope.

**Problem**

Describe the tropical vertices of $C(d, U(n))$.

We give an algorithmic solution to the above problem in a more general setting (Bergman fans of matroids). We also investigate the problem below.

**Problem**

Given a linear space $L \subseteq \mathbb{R}^N$ and some $x \in \mathbb{R}^N$, describe the set of points in $L$ that are $l_\infty$-nearest to $x$. 
Definition

The *tropical semiring* is the extended real numbers \( \mathbb{R} \cup \{-\infty\} \) where tropical addition is defined as

\[
a \oplus b := \max\{a, b\}
\]

and tropical multiplication is defined as

\[
a \odot b := a + b.
\]

Definition

The *tropical semi-module* is \((\mathbb{R} \cup \{-\infty\})^n\). If \(x, y \in (\mathbb{R} \cup \{-\infty\})^n\), then \(x \oplus y\) is the vector whose \(i\)th entry is \(x_i \oplus y_i\). If \(\alpha \in \mathbb{R} \cup \{-\infty\}\) then the \(i\)th entry of \(\alpha \odot x\) is \(\alpha \odot x_i\).

\[
\begin{pmatrix}
1 \\
2
\end{pmatrix} \oplus 1 \odot \begin{pmatrix}
-\infty \\
2
\end{pmatrix} = \begin{pmatrix}
1 \\
2
\end{pmatrix} \oplus \begin{pmatrix}
-\infty \\
3
\end{pmatrix} = \begin{pmatrix}
1 \\
3
\end{pmatrix}
\]
In what follows, \( v_1, \ldots, v_k \in (\mathbb{R} \cup \{-\infty\})^n \).

**Definition**

A tropical polytope is a set of the form

\[
t\text{conv}\{v_1, \ldots, v_k\} := \{\lambda_1 \otimes v_1 \oplus \cdots \oplus \lambda_k \otimes v_k : \lambda_1, \ldots, \lambda_k \in \mathbb{R} \cup \{-\infty\} \text{ and } \lambda_1 \oplus \cdots \oplus \lambda_k = 0\}.
\]

**Definition**

A tropical polyhedral cone is a set of the form

\[
t\text{cone}\{v_1, \ldots, v_k\} := \{\lambda_1 \otimes v_1 \oplus \cdots \oplus \lambda_k \otimes v_k : \lambda_1, \ldots, \lambda_k \in \mathbb{R} \cup \{-\infty\}\}.
\]
Example

tconv\{(1, 0), (0, 1)\} is displayed below. tcone\{(1, 0), (0, 1)\} is the entire plane.
Bergman Fans

Definition

Let $\mathcal{M}$ be a matroid on ground set $E$. The Bergman fan of $\mathcal{M}$ is the subset of $\mathbb{R}^E$:

$$\tilde{\mathcal{B}}(\mathcal{M}) := \{ x \in \mathbb{R}^E : \text{if } C \text{ is a circuit of } \mathcal{M}, \text{ then the maximum of } \{x_i : i \in C\} \text{ is attained twice} \}.$$ 

Example

Let $\mathcal{M}$ be the matroid underlying the complete graph on 4 vertices. The edge-labeled graph on the left represents an element of $\tilde{\mathcal{B}}(\mathcal{M}(K_4))$ whereas the one on the right does not.
Theorem (Ardila and Klivans, 2006)

The collection of ultrametrics on a finite set of size $n$ is exactly the Bergman fan of the complete graph on $n$ vertices.

Theorem (Ardila, 2004)

Bergman fans of matroids are tropical polyhedral cones. The Chipoi-Fichet algorithm for ultrametrics extends to Bergman Fans of matroids.

Theorem (B.-Long 2016)

Let $\mathcal{M}$ be a matroid on ground set $E$ and let $x \in \mathbb{R}^E$. Then the subset of the Bergman fan $\tilde{\mathcal{B}}(\mathcal{M})$ of points that are $l^\infty$-nearest to $x$ is a tropical polytope. We have an algorithm for computing its tropical vertices.

Proposition (B.-Long 2016)

All $l^\infty$-nearest ultrametrics to $d$ have the same topology if and only if all the vertices have the same topology.
Algorithm idea:

1. Compute the $l^\infty$-nearest ultrametric given by Chepoi-Fichet
2. Slide internal nodes of the tree down until either creating a new polytomy, or sliding any further would increase $l^\infty$ distance
3. Repeat
4. If at most one internal node can still be moved down, then that ultrametric is a tropical vertex
Tropical Polytope of Nearest Ultrametrics
Other Things

- Algorithm extends to Bergman fans of arbitrary matroids by generalizing the notion of tree topology
- Question for future work: how probable is it that a given dissimilarity map is $\ell^\infty$-nearest to ultrametrics with different topologies?
Let $L \subseteq \mathbb{R}^n$ be a linear space and let $x \in \mathbb{R}^n$.

Denote $C(x, L) := \{y \in L : \|x - y\|_\infty \text{ is minimized}\}$

If $\delta$ is the $l^\infty$ distance from $x$ to $L$ then

$$C(x, L) = (x + [-\delta, \delta]^n) \cap L$$
Oriented Matroids

**Definition**
The oriented matroid associated to a linear space $L$ is the collection of sign vectors that appear as coefficients of linear forms vanishing on $L$.

**Example**
Let $L = \{(t, t, 0) : t \in \mathbb{R}\} \subset \mathbb{R}^3$. Then the oriented matroid associated to $L$ is the sign vectors:

$$(+,-,0) \quad (-,+,-,0) \quad (0,0,0)$$

$$(+,-,+,-) \quad (-,+,-,0) \quad (0,0,+)$$

$$(+,-,+,0) \quad (-,+,-,0) \quad (0,0,-).$$

Some linear forms that vanish on $L$ are

$$x - y \quad - 2x + 2y + 5z \quad - 4z$$
Rank Functions of Oriented Matroids

**Definition**
If $O_L$ is the oriented matroid associated to $L$ and $\sigma \in O_L$, then we define $\text{rank}(\sigma)$ to be the size of the largest subset of the support of $\sigma$ that can be arbitrarily specified in $L$.

**Example**
Let $L = \{(t, t, 0) : t \in \mathbb{R}\} \subset \mathbb{R}^3$.
- $\text{rank}(0, 0, +) = 0$
- $\text{rank}(+, -, 0) = 1$
- $\text{rank}(-, +, +) = 1$. 
We can associate a sign vector to each face of a cube $x + [-\delta, \delta]^n$.

If $x \in \mathbb{R}^n$ is distance $\delta$ from $L$, we define $\text{type}_L(x)$ to be the sign vector of the minimal face of $x + [-\delta, \delta]^n$ that contains $C(x, L)$. 

\begin{align*}
&(-,+) \quad (0,+) \quad (+,+) \\
&(-,0) \quad (0,0) \quad (+,0) \\
&(-,-) \quad (0,-) \quad (+,-)
\end{align*}
Theorem (B.-Long)

Let \( L \subseteq \mathbb{R}^n \) be a linear space of dimension \( d \).

1. The set of possible values of \( \text{type}_L(x) \) is the oriented matroid underlying \( L \).
2. The dimension of the set of points in \( L \) that are closest to \( x \) in \( L \) is \( d - \text{rank(\text{type}(x))} \).

Corollary

The \( l^\infty \)-nearest point in \( L \) to a given \( x \in \mathbb{R}^n \) is unique for all \( x \) if and only if the matroid underlying \( L \) is uniform.
References

Federico Ardila.
Subdominant matroid ultrametrics.

Federico Ardila and Caroline J. Klivans.
The Bergman complex of a matroid and phylogenetic trees.

Daniel Irving Bernstein and Colby Long.
L-infinity optimization in tropical geometry and phylogenetics.

Victor Chepoi and Bernard Fichet.
L-infinity optimization via subdominants.
*Journal of Mathematical Psychology, 44, 600-616, 2000.*