Dissimilarity Vectors of Trees and Their Tropical Linear Spaces

Benjamin Iriarte

San Francisco State University
San Francisco CA USA

M.A. Mathematics, Thesis Defense
Basic Notions

• **A weighted n-tree** $T$: A tree with set of leaves $[n] = \{1, \ldots, n\}$ and s.t. all edges $e$ are given a **weight** $w(e)$:

  $$w(e) \in \mathbb{R}_{>0} \text{ if } e \text{ is internal.}$$

• $d \in \mathbb{R}^{[n]}$,

  $$d = (d_{\sigma})_{\sigma \in {n \choose m}}$$

and $d_{\sigma}$ is the **total weight** of the smallest subtree of $T$ which contains the leaves $\sigma$.

• $d \in \mathbb{R}^{[n \choose m]}$ is the **$m$-dissimilarity vector of $T$**.
Suppose we start with a tree, each of whose edges has a corresponding associated positive weight. This is shown for an 11-tree.
This introduces the notion of dissimilarity vector. The notion can be generalized.
Relation between 2-dissimilarity and $m$-dissimilarity vectors of $T$ for $m > 2$

(Cools)
Let $C_m \subseteq S_m$ be the set of cyclic permutations.

Take $\{i_1, i_2, \ldots, i_m\} \in \binom{[n]}{m}$.

Assume WLOG that $1 = i_1$, $2 = i_2$, $\ldots$, $m = i_m$.

Then

$$d_{12\ldots m} = \frac{1}{2} \cdot \min_{\sigma \in C_m} \left\{ d_{\sigma(1)} + d_{\sigma(1)\sigma^2(1)} + d_{\sigma^2(1)\sigma^3(1)} + d_{\sigma^3(1)\sigma^4(1)} + \cdots + d_{\sigma^{m-1}(1)\sigma^m(1)} \right\}$$
In particular, we have the nice equation

\[ d_{ijk} = \frac{1}{2}(d_{ij} + d_{ik} + d_{jk}) \]

for all \(1 \leq i < j < k \leq n\).
\[ d_{123456789(10)} = d_{12} + d_{23} + d_{34} + d_{45} + d_{56} + d_{67} + d_{78} + d_{89} + d_{9(10)} + d_{(10)1} \]
• (I.G.) Corollary

For all \( \{i, j, k, l, p\} \in \binom{[n]}{5} \) we have:

\[
d_{ij} = \frac{1}{3} (2d_{ijk} + 2d_{ijl} + 2d_{ijp} - d_{ipl} - d_{ikl} - d_{ikp} - d_{jpl} - d_{jkl} - d_{jkp} + 2d_{plk})
\]

(It is possible to characterize 3-dissimilarity vectors from this corollary, using Buneman)
Tropical Algebraic Geometry

- (Ardila, Pachter, Speyer, Sturmfels) **Tropical Geometry is a powerful tool to study weighted trees.**
- Let $\mathbb{K}$ be the a.c. field of elements

\[
\omega = \sum_{k=-\infty}^{k=p} c_k t^{k/q}
\]

$p \in \mathbb{Z}$, $c_p \neq 0$, $q \in \mathbb{Z}^+$ and $c_k \in \mathbb{C}$

Consider the valuation $\text{val}: \mathbb{K} \mapsto \mathbb{Q} \cup \{-\infty\}$ by which

\[
\text{val}(\omega) = p/q \text{ and } \text{val}(0) = -\infty
\]

**Examples:**

\[
\omega = t^2 + t^5 + t^6 \text{ and } \text{val}(\omega) = 6
\]

\[
\omega = (1 + i)t^{1/2} - (i)t^{2/3} + (2)t^{5/3} + (\sqrt{2} - \sqrt{3}i)t^{5/2} \text{ and } \text{val}(\omega) = 5/2
\]

\[
\omega = t^{1/2} + 1 + t^{-1/2} + t^{-1} + t^{-3/2} + t^{-2} + t^{-5/2} \ldots \text{ and } \text{val}(\omega) = 1/2
\]
• Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$.

• Tropical semiring $(\overline{\mathbb{R}}, \oplus, \odot)$.

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

\[ a \odot b = a + b \]
• Tropicalization of a Polynomial $f \in \mathbb{K}[X_1, \ldots, X_N]$.

• Let $A \subseteq \mathbb{Z}^N_{\geq 0}$ be finite.

$$f = \sum_{\alpha \in A} c_{\alpha} X_1^{\alpha_1} \cdots X_N^{\alpha_N}$$

$$\text{trop}(f) = \bigoplus_{\alpha \in A} \text{val}(c_{\alpha}) \odot X_1^{\alpha_1} \odot \cdots \odot X_N^{\alpha_N}$$

Example:

$$f \in \mathbb{K}[X, Y, Z]$$

$$f = iXY + (t^{\frac{1}{5}} + t^{-1} + 2i)XYZ^2 - Z^4$$

$$\text{trop}(f) = \max\{x + y, \frac{1}{5} + x + y + 2z, 4z\}$$
• Tropical Hypersurface defined by trop($f$).

\[ \mathcal{T}(\text{trop}(f)) := \{ x \in \mathbb{R}^N \mid \text{the maximum of trop}(f) \text{ is attained at least twice} \} \]

Example (Cont.):

\[ \mathcal{T}(\text{trop}(f)) = \{ x, y, z \mid x + y = \frac{1}{5} + x + y + 2z \geq 4z \} \]
\[ \cup \{ x, y, z \mid x + y = 4z \geq \frac{1}{5} + x + y + 2z \} \]
\[ \cup \{ x, y, z \mid \frac{1}{5} + x + y + 2z = 4z \geq x + y \} \]
• The *Tropical Variety* of an ideal $I$ of $\mathbb{K}[X_1, \ldots, X_N]$:

$$\mathcal{T}(I) = \bigcap_{f \in I} \mathcal{T} \left( \text{trop}(f) \right)$$

• *Fundamental Theorem of Tropical Algebraic Geometry:*

$$\mathcal{T}(I) = \text{closure of } \{(\text{val}(C_1), \ldots, \text{val}(C_N)) \mid (C_1, \ldots, C_N) \in V(I) \subseteq \mathbb{K}^N\}$$

• $\mathcal{T}(\text{trop}(f))$ and $\mathcal{T}(I)$ often considered in

$$\mathbb{T}P^{N-1} = \frac{\mathbb{R}^N}{(1, 1, \ldots, 1)\mathbb{R}}$$
Tropical Linear Spaces

- Plücker ideal $I_{m,n}$ in $\mathbb{K}[P_{i_1i_2\ldots i_m} \mid \{i_1, i_2, \ldots, i_m\} \in ([n]_m)]$, consisting of the algebraic relations-syzygies among the maximal minors of a generic matrix in $\mathbb{K}^{m \times n}$. Elements of $I_{m,n}$ are called Plücker relations.

- Quadrics form a Gröbner basis for $I_{m,n}$, among which we find the three-term Plücker relations:

$$P_{Sij}P_{Skl} - P_{Sik}P_{Sjl} + P_{Sil}P_{Sjk}$$

for all $S \in \left(\begin{bmatrix} [n] \\ m-2 \end{bmatrix}\right)$ and $\{i, j, k, l\} \in \left(\begin{bmatrix} [n] \setminus S \\ 4 \end{bmatrix}\right)$.
• The tropical Grassmannian-\((m, n)\):

\[ G_{m,n} = \mathcal{T}(I_{m,n}) \]

• A vector \( p \in \overline{\mathbb{R}}^{[n \choose m]} \) satisfies the tropical Plücker relations if and only if:

\[
p \in \bigcap_{S \in \binom{[n]}{m-2}} \mathcal{T} \left( \text{trop}(P_{Sij}P_{Skl} - P_{Sik}P_{Sjl} + P_{Sil}P_{Sjk}) \right)
\]

\[
\{i,j,k,l\} \in \binom{[n]}{4} \setminus S
\]

• Assume \( p \in \mathbb{R}^{[n \choose m]} \subsetneq \overline{\mathbb{R}}^{[n \choose m]} \) satisfies the tropical Plücker relations.
**notation:** Canonical basis of $\mathbb{R}^n$: $\{e_1, \ldots, e_n\}$. For $I \in \binom{[n]}{m}$, $e_I = \sum_{i \in I} e_i$.

- (Speyer) Every face of the regular subdivision of the $m$-hypersimplex $\mathcal{H}_m := \text{chull}\{e_I | I \in \binom{[n]}{m}\} \subseteq \mathbb{R}^n$ induced by $p$ is matroidal.
- Three objects:
  1. $L(p) := \bigcap_{\{i_1 < i_2 < \cdots < i_m < i_{m+1}\} \in \binom{[n]}{m+1}} \mathcal{T} \left( \bigoplus_{r=1}^{m+1} (p_{i_1 i_2 \cdots \hat{i}_r \cdots i_m i_{m+1}} \odot x_{i_r}) \right)$
  2. $P_p := \{x \in \mathbb{R}^n | x_{i_1} + x_{i_2} + \cdots + x_{i_m} \geq p_{i_1 i_2 \cdots i_m} \text{ for all } \{i_1, i_2, \ldots, i_m\} \in [n]_m\}$
  3. $P'_p := \{x \in P_p | \text{If } y \leq x \text{ and } y \in P_p, \text{ then } y = x\}$
• Facts (Speyer):
  i. \( L(p) = P'_p \text{ in } \mathbb{T}P^{n-1} \).
  ii. \( P'_p \) is an \((m - 1)\)-dimensional pure and contractible polyhedral complex.
  iii. Tropical linear space associated to \( p \).
• The tight span \( \mathcal{T}_p \) is the complex of bounded faces of \( P_p \) (or \( P'_p \), or \( L(p) \) projectively). To define \( \mathcal{T}_p \), we can relax the requirement that \( p \) satisfies the tropical Plücker relations. Face enumeration of \( \mathcal{T}_p \) and the \textit{f-vector conjecture} (Speyer).
Main Question

- Given a vector $d \in \mathbb{R}^{\binom{n}{m}}$, is $d$ the $m$-dissimilarity vector of an $n$-tree?
The case $m = 2$

**Observation:** For $p \in \mathbb{R}^{[n]}$, $p \in \mathcal{G}_{2,n}$ if and only if $p$ satisfies the tropical Plücker relations. Quadrics form a *Tropical Basis* for $I_{2,n}$.

- (Buneman) **Four point condition theorem:** A vector $d \in \mathbb{R}^{[n]}$ is a 2-dissimilarity vector if and only if $d \in \mathcal{G}_{2,n}$.

- (Dress) A vector $d \in \mathbb{R}^{[n]}$ is a 2-dissimilarity vector if and only if $T_d$ is a tree.
In general, quadratics DO NOT form a tropical basis of $I_{m,n}$ for $m > 2$, only a Gröbner Basis.

- (Pachter, Speyer) **Problem 1:** If $m > 2$, decide whether the set of $m$-dissimilarity vectors of $n$-trees is contained in $G_{m,n}$.

- (Ardila, Sturmfels) **Problem 2:** Given an $m$-dissimilarity vector $d \in \mathbb{R}^{\binom{n}{m}}$, study the tight span $T_d$. If **Problem 1** has a positive answer (or at least, if $d$ satisfies the tropical Plücker relations), this amounts to studying the bounded faces of $P'_d$. 

$m > 2$ and problems
Ultrametrics

- **Ultrametric n-tree:**
  1. rooted
  2. binary
  3. leaves \([n]\)
  4. nonnegative real weight \(w(e)\) for each edge \(e\)
  5. \(\lambda\)-equidistant, for some \(\lambda > 0\)
  6. induces a metric on \([n]\).

- **Ultrametric:** A metric space \(S\) with distance \(d: S \times S \rightarrow \mathbb{R}_{\geq 0}\) for which, \(d_{xz} \leq \max\{d_{xy}, d_{yz}\}\) for all \(x, y, z \in S\). Both definitions are “equivalent”.
The Tropical Grassmannian

**notation:** For a matrix $M \in \mathbb{K}^{m \times n}$ and $\rho \in \binom{[n]}{m}$, $M_\rho$ denotes the maximal minor of $M$ coming from columns $\rho$.

• (I.G.) **Theorem:** Let $m > 2$. If $d \in \mathbb{R}^{\binom{[n]}{m}}$ is an $m$-dissimilarity vector, then $d \in \mathcal{G}_{m,n}$.

**Sketch of proof:**

I. Fix a weighted $n$-tree $T$. Define $d' \in \mathbb{R}^{\binom{[n]}{2}}$ by:

$$d'_{ij} = 2t^* + d_{ij} - d_{jn} - d_{in} \text{ for all } \{i, j\} \subseteq \binom{[n]}{2}$$

For $t^* \in \mathbb{R}_{>0}$ sufficiently large, $d'$ induces an ultrametric on $[n]$, so $d'$ is the $m$-dissimilarity vector of an ultrametric tree.

II. There exists a full rank matrix $M' \in \mathbb{K}^{m \times n}$ s.t. $\text{val}(M'_\rho) = d'_\rho$ for all $\rho \in \binom{[n]}{m}$.

III. There exists a diagonal matrix $D \in \mathbb{K}^{n \times n}$ (given by $D_{ii} = d_{in} - t^*$ for all $i \in [n - 1]$ and $D_{nn} = -t^*$) such that $M = M'D$ satisfies $\text{val}(M_\rho) = d_\rho$ for all $\rho \in \binom{[n]}{m}$.
• (I.G.) **Corollary:** If $d \in \mathbb{R}^{[n \choose m]}$ is an $m$-dissimilarity vector, $d$ satisfies the tropical Plücker relations, $L(d)$ can be defined and studied through $\mathcal{T}_d$. The $f$-vector conjecture of Speyer applies.

• Our approach produces a complete flag of tropical linear spaces.
The tight span

Let $1 < m < n$. Fix a weighted $n$-tree $T$ with $m$-dissimilarity vector $d$. We will describe the tight span $\mathcal{T}_d$. Let $\text{in}(T)$ be the internal subtree of $T$.

Example:
notation: $\subseteq_{st}$ denotes the relation **being a subtree of**.

**definition:** For $S \subseteq_{st} in(T)$, the **extended tree** $\overline{S}$. The set of leaves of $\overline{S}$ is called the set of **extended leaves of** $S$.

**Example:**
• $S \subseteq_{st} \text{in}(T)$ determines two collections $\mathcal{H}$ and $\mathcal{R}$ of subsets of leaves of $T$.

i. members of $\mathcal{H}$ are indexed by the leaves of $S$. We have $i \in H_\ell \in \mathcal{H}$ if and only if the minimal path from $i$ to $S$ meets $S$ at $\ell$.

ii. members of $\mathcal{R}$ are indexed by the extended leaves of $S$. We have $i \in R_\ell \in \mathcal{R}$ if and only if the minimal path from $i$ to $\bar{S}$ meets $\bar{S}$ at $\ell$.

**observations:**

I. $\mathcal{R}$ is a partition of the set $[n]$ of leaves of $T$.

II. $\mathcal{H}$ is a collection of disjoint subsets of $[n]$, but generally not a partition of $[n]$.

III. Every member of $\mathcal{H}$ is the disjoint union of at least two elements of $\mathcal{R}$. 
Example:

\[ \mathcal{K} = \{ \{5,6\}, \{7,8,9,10\} \} \]

\[ \mathcal{R} = \{ \{1,2,3,4,5\}, \{6\}, \{7\}, \{8\}, \{9,10\}, \{11,12,13,14,15\} \} \]
**definition:** Let $0 \leq i \leq m - 1$. A **good pair** $(S, \mathcal{A})$ with $|\mathcal{A}| = i$ is given by:

i. $S \subseteq_{st} \text{in}(T)$ with at most $m - 1$ leaves.

ii. $\mathcal{A}$ a subset of the leaves of $S$.

iii. the number of extended leaves of $S$ not adjacent to an element of $\mathcal{A}$ is at least $m + 1 - i$. 

[Diagram showing an example of a good pair $(S, \mathcal{A})$ with $m=5$, $|\mathcal{A}|=2$, and $m+1-i \leq 4$ extended leaves not adjacent to $\mathcal{A}$]
• (I.G.) **Theorem:** A good pair \((S, A)\) with \(|A| = i\) uniquely determines an \(i\)-dimensional face of \(\mathcal{I}_d\). The face lattice of this face is isomorphic to the face lattice of the \(i\)-dimensional cube.

The set of bases of the matroid associated to our face is the collection of all \(m\)-sets \(I \in \binom{[n]}{m}\) satisfying:

\[
|I \cap H_\ell| = 1 \text{ for all leaves } \ell \in A.
\]
\[
|I \cap H_\ell| \geq 1 \text{ for all leaves } \ell \text{ of } S \text{ s.t. } \ell \notin A.
\]
\[
|I \cap R_\ell| \leq 1 \text{ for all leaves } \ell \text{ of } \overline{S}.
\]

Good pairs give rise to all \(i\)-dimensional faces of \(\mathcal{I}_d\) whenever \(0 \leq i < m - 1\). In general, good pairs give rise only to a subset of the \((m - 1)\)-dimensional faces of \(\mathcal{I}_d\).

If \(T\) is trivalent, \(S \subseteq_{st} \text{in}(T)\), and \(A\) is an \(i\)-subset of the leaves of \(S\), then \((S, A)\) is a good pair if and only if \(S\) has at most \(m - 1\) leaves and at least \(m - 1 + i\) vertices.
• **Corollary:** The vertices of $\mathcal{T}_d$ correspond to trees $S \subseteq_{st} \text{in}(T)$ with at most $m - 1$ leaves and at least $m + 1$ extended leaves.

For any such $S$, the coordinates of the associated vertex $x$ are given by:

$$x_i = \omega(i, S) + \frac{\omega(S)}{m} \text{ for all } i \in [n]$$

If $T$ is trivalent, the vertices of $\mathcal{T}_d$ correspond to trees $S \subseteq_{st} \text{in}(T)$ with at most $m - 1$ leaves and least $m - 1$ vertices.
**definition:** Suppose $S \subseteq_{st} \text{in}(T)$. If the collection $\mathcal{H}$ associated to $S$ is a partition of $[n]$, we say that $S$ is a **full subtree** of $\text{in}(T)$.

**Example:**

![Diagram](image)
• (I.G.) **Theorem:** The \((m - 1)\)-dimensional faces of \(\mathcal{T}_d\) which do not arise from good pairs, are in one to one correspondence with the full \(m\)-subtrees \(S \subseteq_{st} \text{in}(T)\).

This faces are line segments for \(m = 2\), hexagons for \(m = 3\), rhombic dodecahedrons for \(m = 4\), and they are simple if and only if \(m < 4\).

The matroid associated to such face, coming from \(S\), is the one whose set of bases are the full transversals of the collection \(\mathcal{H}\) associated to \(S\) (Notice \(|\mathcal{H}| = m\)). It is a transversal matroid.

The \(f\)-vector of the face is given by:

\[
\begin{align*}
    f_{-1} &= 1 \\
    f_i &= \binom{m-1}{i} \quad \text{for all } i \text{ with } 0 \leq i \leq m - 2 \\
    f_{m-1} &= 1
\end{align*}
\]

If \(T\) is trivalent, full \(m\)-subtrees of \(\text{in}(T)\) correspond to trees \(S \subseteq_{st} \text{in}(T)\) with \(m\) leaves and \(2(m - 1)\) vertices.
Let $S \subseteq_{st} \text{in}(T)$ and $A$ a subset of the leaves of $S$.

**definition:** Let $S_A$ be the subtree of $S$ spanned by the vertices in $\mathcal{V}(S) \setminus A$.

- (I.G.) **Theorem:** Let $F$ and $F'$ be faces of $\mathcal{T}_d$. Suppose that $F$ comes from a good pair $(S, A)$ and that $F'$ comes from a good pair $(S', A')$. Then, $F \cap F' \neq \emptyset$ if and only if:

$$S_A \subseteq_{st} S' \text{ and } S'_{A'} \subseteq_{st} S$$

Moreover, in the case when our faces meet, $F \cap F'$ comes from the good pair $(S \cap S', A \cap A')$. In particular, $F \cap F'$ is of dimension $|A \cap A'|$. 
Tight spans of trees
<table>
<thead>
<tr>
<th>Background</th>
<th>Problems</th>
<th>Results</th>
<th>Pictures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m = 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Background

Problems

Results

Pictures

\[ m = 4 \]
Consider a rank 8 matroid over $[14]$ whose bases satisfy the $\geq 1$-inequalities.

Problem: From the set of bases, identify the $\geq 1$-inequalities.
Thanks!