A tropical interpretation of $m$—dissimilarity maps

Cristiano Bocci ∗ and Filip Cools †

Abstract. Let $T$ be a weighted tree with $n$ numbered leaves and let $D = (D(i,j))_{i,j}$ be its distance matrix, so $D(i,j)$ is the distance between the leaves $i$ and $j$. If $m$ is an integer satisfying $2 \leq m \leq n$, we prove a tropical formula to compute the $m$-dissimilarity map of $T$ (i.e. the weights of the subtrees of $T$ with $m$ leaves), given $D$. For $m = 3$, we present a tropical description of the set of $m$-dissimilarity maps of trees. For $m = 4$, a partial result is given.

MSC. 05C05, 05C12, 14M15, 14Q99, 15A99, 92B05

1 Introduction

Let $D$ be a matrix whose rows and columns are indexed by a set $X$. We assume that $D$ is symmetric and has zero entries on the main diagonal. In phylogenetics, these kind of matrices are called dissimilarity matrices. Usually, we take $X = [n] := \{1, 2, \ldots, n\}$. Hence a dissimilarity matrix $D$ can also be seen as a map $D : [n]^2 \rightarrow \mathbb{R}$, with $D(i,j) = D(j,i)$ and $D(i,i) = 0$ for each $i, j \in [n]$.

A metric is a non-negative dissimilarity matrix which satisfies the triangle inequality $D(i, j) \leq D(i, k) + D(k, j)$ for all $i, j, k \in X$.

We say that $D$ has a graph realization if there is a weighted graph (so a non-negative weight is assigned to each edge) whose node set contains $X$ and such that the distance (i.e. the length of the shortest path) between nodes $i, j \in X$ is exactly $D(i, j)$. A distance matrix is a non-negative dissimilarity matrix that has a graph realization. In [3, 4], one can find some results on these kind of matrices.

In the case the graph is a tree and $X$ corresponds to the set of leaves, $D$ is called a tree metric. This case has been studied intensively and is well understood. The main result is the following (see [2] or [6, Theorem 2.36]).

Theorem 1.1 (Tree Metric Theorem). Let $D$ be a non-negative dissimilarity matrix on $[n]$. Then $D$ is a tree metric on $[n]$ if and only if, for every four (not necessarily distinct) elements $i, j, k, l \in [n]$, the maximum of the three numbers

\[ D(i, j), D(i, k), D(k, j) \]

is equal to

\[ D(i, k) + D(k, j) \]

References:

* I.T.I.S. “A. Avogadro”, Via case Nuove 27, 53021 Abbadia San Salvatore (SI), Italy, email: cristiano.bocci@gmail.com.
† K.U.Leuven, Department of Mathematics, Celestijnenlaan 200B, B-3001 Leuven, Belgium, email: Filip.Cools@wis.kuleuven.be.
\[ D(i, j) + D(k, l) \leq D(i, k) + D(j, l) \] and
\[ D(i, j) + D(k, l) \leq D(i, l) + D(j, k) \]
is attained at least twice. Moreover, the tree \( T \) with leaves \([n]\) that realizes \( D \) is unique.

The condition of the theorem is called the four-point condition. It is a necessary and sufficient condition on a matrix to be realized by a tree.

Tree metrics on \( n \) leaves are parameterized by the space of trees \( T_n \subset \mathbb{R}^{(\frac{n}{2})} \).

Theorem 1.2. The space of trees \( T_n \) is the union of \((2n-5)!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-5) \) orthants isomorphic to \( \mathbb{R}^{2n-3}_{\geq 0} \). More precisely, \( T_n \) is a simplicial fan of pure dimension \( 2n-3 \) in \( \mathbb{R}^{(\frac{n}{2})} \).

We can consider a generalisation of the concept of dissimilarity matrix. Let \( m \leq n \) be an integer. A map \( D : [n]^m \to \mathbb{R} \) is called an \( m \)-dissimilarity map if
\[ D(i_1, \ldots, i_m) = D(i_{\pi(1)}, \ldots, i_{\pi(m)}) \]
for all permutations \( \pi \in S_m \) and \( D(i_1, i_2, \ldots, i_m) = 0 \) if the numbers \( i_1, \ldots, i_m \) are not pairwise distinct.

We say that \( D \) is realized by a tree \( T \) if the leaf set of \( T \) is \([n]\) and if for each \( m \)-subset \( V = \{i_1, \ldots, i_m\} \subset [n] \), the weight of the smallest subtree of \( T \) containing \( V \) is equal to \( D(i_1, \ldots, i_m) \). An important result on \( m \)-dissimilarity maps of trees is given in [5].

Theorem 1.3. Let \( T \) be a tree with \( n \) leaves and no vertices of degree 2. Let \( m \geq 3 \) be an integer. If \( n \geq 2m - 1 \), then \( T \) is uniquely determined by its \( m \)-dissimilarity map \( D \). If \( n = 2m - 2 \), this is not true.

In this paper, we give a description of a map \( \phi^{(m)} : \mathbb{R}^{(\frac{n}{2})} \to \mathbb{R}^{(\frac{n}{3})} \), sending the distance matrix of a tree \( T \) to its corresponding \( m \)-dissimilarity map (see Theorem 3.2 in Section 3). In Section 4, we investigate the case \( m = 3 \). In particular, we show that \( \phi^{(3)}(T_n) \) is equal to the intersection of the tropical Grassmannian \( G_{3,n} \) with a linear space (see Theorem 4.6). In Section 5, we give a partial result on the case \( m = 4 \). An introduction to tropical geometry is given in Section 2.

To finish this section, we describe the relation with Phylogenetics. A classical problem in computational biology is to construct a phylogenetic tree from a sequence alignment of \( n \) species

| Species 1 | ACAATGTCATTAGCGATACGTACGATGC... |
| Species 2 | ACCGTGTCATTAGCGATACGTACGATGC... |
| Species 3 | ACCGTAGTCATTAGCGATACGTACGATGC... |
| Species 4 | GCACAGTCATAGCGATACGTACGATGC... |
| Species n | GAACTGTCATAGCGATACGTACGATGC... |

The main technique to select a tree model is computing the maximum likelihood estimate (MLE) for each of the \((2n-5)!! \) trees. Unluckily, all the MLE
computations are very difficult, even for a single tree, and this approach requires examining all exponentially many trees.

A popular way to avoid this problem is the so-called distance based approach, where one collapses the data to a dissimilarity matrix and obtains a tree via a projection onto tree space $T_n$ (by using the neighbor-joining algorithm). In fact, for such sequence data, computational biologists infer the distance between any two taxa. Thus, an interesting problem of phylogenetics concerns the construction of a weighted tree which represents this distance matrix, provided such a tree exists.

More general, we may think of an $m$-dissimilarity map as a measure of how dissimilar each subset of $m$ species is. As a generalization of the previous problem, we can search for a weighted tree such that the $m$-subtree weights represent the entries of the $m$-dissimilarity map. This problem has some natural relevance in Phylogenetics. Indeed, for example, it can be more reliable statistically to estimate the triple weights $D(i,j,k)$ rather than the pairwise distances $D(i,j)$ ([5], [6]).

2 Tropical geometry

The tropical semiring $(\mathbb{R} \cup \{-\infty\}, \oplus, \otimes)$ is the set of real numbers completed with $-\infty$, equipped with two binary operations: the tropical sum is the maximum of two numbers and the tropical multiplication is the ordinary sum.

Tropical monomials $x_1^{a_1} \cdots x_k^{a_k}$ represent ordinary linear forms $\sum_{i=1}^k a_i x_i$ and tropical polynomials

$$\bigoplus_{a \in A} \lambda_a \otimes x_1^{a_1} \otimes \cdots \otimes x_k^{a_k},$$

with $A \subset \mathbb{N}^k$ finite and $\lambda_a \in \mathbb{R}$, represent piecewise-linear convex functions

$$F : \mathbb{R}^k \to \mathbb{R} : (x_1, \ldots, x_k) \mapsto \max_{a \in A} \{ \lambda_a + \sum_{i=1}^k a_i x_i \}.$$
maximum of the collection of numbers

\[
\left\{ \sum_{i=1}^{k} a_i x_i + \lambda \right\}_{a \in A}
\]

is attained at least twice.

**Theorem 2.1.** If \( I \subset K[x_1, \ldots, x_n] \) is an ideal, the following two subsets of \( \mathbb{R}^k \) coincide:

1. the intersection of all tropical hypersurfaces \( T(\text{trop}(f)) \) with \( f \in I \);
2. the closure in \( \mathbb{R}^k \) of the set

\[
\{ (-\text{val}(y_1), \ldots, -\text{val}(y_k)) \mid (y_1, \ldots, y_k) \in V(I) \} \subset \mathbb{Q}^k.
\]

**Proof.** See [7, Theorem 2.1]. \( \square \)

For an ideal \( I \subset K[x_1, \ldots, x_k] \), we denote by \( T(I) \subset \mathbb{R}^k \) the set mentioned in Theorem 2.1. It is called the *tropical variety* of the ideal \( I \).

**Definition 2.2.** If \( T(I) \subset \mathbb{R}^k \) is a tropical variety, we say that \( \{f_1, \ldots, f_r\} \) is a tropical basis of \( T(I) \) if and only if \( I = \langle f_1, \ldots, f_r \rangle \) and

\[ T(I) = T(\text{trop}(f_1)) \cap \cdots \cap T(\text{trop}(f_r)). \]

**Remark 2.3.** In general, a set of generators of an ideal \( I \) is not a tropical basis for \( T(I) \). Of course, the singleton \( \{f\} \) is a tropical basis for the tropical hypersurface \( T(\text{trop}(f)) \).

We are mainly interested in the tropical variety \( T(I_{m,n}) \), where \( I_{m,n} \) is the ideal of the Grassmannian \( G(m, n) \subset \mathbb{R}^\binom{n}{m} \). To be more precise, we fix a polynomial ring

\[
\mathbb{Z}[x] = \mathbb{Z}[x_{i_1 i_2 \cdots i_m} \mid 1 \leq i_1 < i_2 < \cdots < i_m \leq n]
\]
in \( \binom{n}{m} \) variables with integer coefficients. The Plücker ideal \( I_{m,n} \) is the prime ideal in \( \mathbb{Z}[x] \), consisting of the algebraic relations among the determinants of the \((m \times m)\)-minors of any \((m \times n)\)-matrix with entries in a commutative ring. It is well-known that \( I_{m,n} \) is generated by quadrics (see for example [8]).

The affine variety defined by \( I_{m,n} \) is the Grassmannian \( G(m, n) \subset \mathbb{R}^\binom{n}{m} \), which parameterizes all \( m \)-dimensional linear subspaces of an \( n \)-dimensional vector space. It has dimension \((n - m)m + 1\).

**Definition 2.4.** The tropical variety \( T(I_{m,n}) \) is called a tropical Grassmannian and is denoted by \( G_{m,n} \).

**Theorem 2.5.** The tropical Grassmannian \( G_{m,n} \) is a polyhedral fan in \( \mathbb{R}^\binom{n}{m} \). Each of its maximal cones has the same dimension, namely \((n - m)m + 1\).
Proof. See [7, Corollary 3.1.]

Now we are going to fix our attention on the case $m = 2$.

**Example 2.6** ($m = 2$ and $n = 4$). The smallest non-zero Plücker ideal is the principal ideal $I_{2,4} = (x_{12}x_{34} - x_{13}x_{24} + x_{14}x_{23})$. Thus $G_{2,4}$ is a fan with three five-dimensional cones $\mathbb{R}^4 \times \mathbb{R}_{\leq 0}$ glued along $\mathbb{R}^4$.

**Theorem 2.7.** The ideal $I_{2,n}$ is generated by the quadratic polynomials

$$p_{ijkl} := x_{ik}x_{jl} - x_{ij}x_{kl} - x_{il}x_{jk} \quad (1 \leq i < j < k < l \leq n). \quad (3)$$

These polynomials form the reduced Gröbner basis if the underlined terms are leading.

Proof. See [8, Theorem 3.1.7 and Proposition 3.7.4].

For each quadruple $\{i, j, k, l\} \subset \{1, 2, \ldots, n\}$, we consider the tropical polynomial $\text{trop}(p_{ijkl}) = (x_{ij} \otimes x_{kl}) \oplus (x_{ik} \otimes x_{jl}) \oplus (x_{il} \otimes x_{jk})$.

This polynomial defines a tropical hypersurface $T(\text{trop}(p_{ijkl}))$. It turns out that the tropical Grassmannian $G_{2,n}$ is the intersection of these $\binom{n}{4}$ hypersurfaces, so the quadrics $p_{ijkl}$ form a tropical basis for $I_{2,n}$ (see [7]).

Let $D$ be an dissimilarity matrix on $[n]$ and $\{i, j, k, l\} \subset [n]$. The maximum of the three numbers $D(i, j) + D(k, l), D(i, k) + D(j, l)$ and $D(i, l) + D(j, k)$ is attained at least twice if and only if $D \in T(\text{trop}(p_{ijkl}))$. Thus Theorem 1.1 implies that a metric $D$ on $[n]$ is a tree metric if and only if $D$ belongs to $\mathcal{T}_n$.

In particular, one has the following result.

**Theorem 2.8.** The space of trees $\mathcal{T}_n$ is the tropical Grassmannian $G_{2,n}$.

Proof. See [7, Theorem 4.2] or the arguments above.

Now we come back to the general case (so the case where $m \leq n$ is arbitrary). The ideal $I_{m,n}$ is generated by quadratic polynomials, known as the Plücker relations. Among these are the three-term Plücker relations

$$p_{Rijkl} := x_{Rik}x_{Rjl} - x_{Rij}x_{Rkl} - x_{Ril}x_{Rjk},$$

which are closely related to (3). Hereby $R$ is any $(m - 2)$-subset of $[n]$ and $i, j, k, l \in [n] \setminus R$.

**Definition 2.9.** The three-term tropical Grassmannian $\mathcal{T}_{m,n}$ is the intersection

$$\mathcal{T}_{m,n} := \bigcap_{R,i,j,k,l} T(\text{trop}(p_{Rijkl})) \subset \mathbb{R}^\binom{n}{m}.$$

In general, the three-term Plücker relations do not generate $I_{m,n}$. If $m = 2$, then $S = \emptyset$ and $\mathcal{T}_{2,n} = G_{2,n}$. For $m \geq 3$, the tropical Grassmannian $G_{m,n}$ is contained in $\mathcal{T}_{m,n}$. This containment is proper for $n \geq m + 4$. 

5
3 A description on the m-subtree weight map

In this section, we are going to give an explicit description of a map

\[ \phi^{(m)} : \mathbb{R}^2 \rightarrow \mathbb{R}^m, \]

sending the dissimilarity matrix \( D \) of a tree \( T \) to its \( m \)-dissimilarity map.

Let \( \prec \) be the order relation on \( \mathbb{N}^\infty \) defined as follows. We have

\[(a_1, a_2, a_3, \ldots) \prec (b_1, b_2, b_3, \ldots)\]

if and only if there exists an \( n \in \mathbb{N} \) such that \( a_i = b_i \) for all \( i < n \) and \( a_n < b_n \).

Let \( T \) be a tree with \( n \) leaves. Let \( r \) be an inner node of \( T \) and consider \( T \) as a rooted tree (with root \( r \)). Let \( N \) be the set of nodes of \( T \). In particular, the set of leaves \([n] = \{1, \ldots, n\}\) is contained in \( N \).

**Lemma 3.1.** There exists a map \( \alpha : N \rightarrow \mathbb{N}^\infty \) such that the following properties hold:

1. \( \alpha \) is injective.
2. If \( n \in N \) is an ancestor of \( m \in N \), we have \( \alpha(m) \succ \alpha(n) \). So the root \( r \) of \( T \) gives rise to the minimum of \( \{\alpha(n)\mid n \in N\} \).
3. If \( n_1, n_2 \in N \) with \( n_2 \) not a descendant nor an ancestor of \( n_1 \), \( m_1 \in N \) a descendant of \( n_1 \) and \( m_2 \in N \) a descendant of \( n_2 \), we have \( \alpha(m_1) \prec \alpha(m_2) \) if and only if \( \alpha(n_1) \prec \alpha(n_2) \).

**Proof.** We will define \( \alpha \) inductively. Take \( \alpha(r) = (0, 0, 0, \ldots) \). For the induction step, if \( \alpha(n) = (a_1, \ldots, a_s, 0, 0, \ldots) \) is defined for some \( n \in N \) with \( a_s \neq 0 \) and if \( m_1, \ldots, m_t \) are the children of \( n \), take \( \alpha(n_i) = (a_1, \ldots, a_s, i, 0, \ldots) \). Note that all the properties hold and that the depth of \( n \in N \) in \( T \) is equal to the number of non-zero entries in \( \alpha(n) \). \( \square \)

We say that the leaves of \( T \) are well-numbered if and only if \( \alpha(i) \prec \alpha(j) \) for all \( i < j \).

A permutation \( \sigma \in S_m \) of \( \{1, \ldots, m\} \) is called cyclic if and only if the decomposition of \( \sigma \) into a product of disjoint cycles consists of only one cycle of order \( m \). Denote the set of cyclic permutations in \( S_m \) by \( C_m \). Note that \( \sigma^m = I_d \) if \( \sigma \in C_m \).

**Theorem 3.2.** Let \( n \) and \( m \) be integers such that \( n > m \geq 2 \). Let

\[ \phi^{(m)} : \mathbb{R}^2 \rightarrow \mathbb{R}^m : X = (X_{i,j}) \mapsto (X_{i_1, \ldots, i_m}) \]

be the map with

\[ X_{i_1, \ldots, i_m} = \frac{1}{2} \min_{\sigma \in C_m} \{X_{i_1, i_{\sigma(1)}} + X_{i_{\sigma(1)}, i_{\sigma(2)}} + \ldots + X_{i_{\sigma(m-1)}, i_{\sigma(m)}}\}. \]

If \( D \in \mathcal{G}_{2,n} \subset \mathbb{R}^2 \) is the dissimilarity matrix of an \( n \)-tree \( T \), then the \( m \)-dissimilarity map of \( T \) is equal to \( \phi^{(m)}(D) \). So the set of \( m \)-dissimilarity maps of \( n \)-trees is equal to \( \phi^{(m)}(\mathcal{G}_{2,n}) \).
Proof. Write
\[ f(X; \sigma; i_1, \ldots, i_m) = X_{i_1, i_{\sigma(1)}} + X_{i_{\sigma(2)}, i_{\sigma(2)}} + \ldots + X_{i_{m-1}, i_{m(1)}}. \]
Note that
\[ f(X; \sigma; i_{\pi(1)}, \ldots, i_{\pi(m)}) = f(X; \pi \sigma^{-1}; i_1, \ldots, i_m) \]
for all \( \pi \in S_m \), hence
\[ \min_{\sigma \in C_m} \{ f(X; \sigma; i_{\pi(1)}, \ldots, i_{\pi(m)}) \} = \min_{\sigma \in C_m} \{ f(X; \sigma; i_1, \ldots, i_m) \}. \]

We have to prove that the weight \( D(i_1, \ldots, i_m) \) of the smallest subtree \( T' \) of \( T \) containing the leaves \( i_1, \ldots, i_m \) is equal to \( \frac{1}{2} \cdot \min_{\sigma \in C_m} \{ f(D; \sigma; i_1, \ldots, i_m) \} \). It is enough to prove this for \( i_1 = 1, \ldots, i_m = m \) (the general case is proved completely analogously). By equation (4), we may also assume the leaves of \( T' \) are well-numbered.

Let \( e = (x, y) \) be an edge of \( T' \) with \( y \) a child of \( x \). We claim that for all \( \sigma \in C_m \), the weight \( w(e) \) of \( e \) is taken into account in at least two of the \( m \) terms of
\[ f(D; \sigma; 1, \ldots, m) = D(1, \sigma(1)) + D(\sigma(1), \sigma^2(1)) + \ldots + D(\sigma^{m-1}(1), 1) \]
and in exactly two of the summands of
\[ f(D; \tau; 1, \ldots, m) = D(1, 2) + D(2, 3) + \ldots + D(m, 1), \]
where
\[ \tau = \begin{pmatrix} 1 & 2 & \ldots & m-1 & m \\ 2 & 3 & \ldots & m & 1 \end{pmatrix} \in C_m. \]
Using this claim, we immediately see
\[ D(i_1, \ldots, i_m) = \frac{1}{2} \cdot f(D; \tau; 1, \ldots, m) = \frac{1}{2} \cdot \min_{\sigma \in C_m} \{ f(D; \sigma; 1, \ldots, m) \}. \]

To finish this theorem, we only need to prove the claim. Consider the split of \( T' \) induced by \( e \) and let \( T'' \) be the component of the split containing \( y \) (hence \( T'' \) is the maximal subtree of \( T' \) containing \( y \) but not \( x \)). Denote the set of leaves of \( T'' \) by \( L'' \). We may assume \( 1 \in L'' \) (the case \( 1 \not\in L'' \) is analogous). Note that in this case \( L'' \) is of the form \( \{1, \ldots, s\} \) for some \( s < m \).

The weight of \( e \) is taken into account in the term \( D(i, j) \) (i.e. the path between the leaves \( i \) and \( j \) of \( T'' \) passes \( e \)) if and only if \( i \in L'' \) and \( j \not\in L'' \) or vice versa. Thus \( w(e) \) is only counted in the two terms \( D(s, s+1) \) and \( D(m, 1) \) of \( f(D; \tau; 1, \ldots, m) \).

So it is enough to show that there exists a \( t \in \{0, \ldots, m-1\} \) such that \( \sigma^t(1) \in L'' \) and \( \sigma^{t+1}(1) \not\in L'' \) (the other case is proved analogously). If we assume this is not the case (so \( \sigma^t(1) \in L'' \) implies \( \sigma^{t+1}(1) \in L'' \)), we get \( L'' = \{1, \ldots, m\} \), a contradiction. \( \square \)
Corollary 3.3. If $D \in \mathcal{G}_{2,n} \subset \mathbb{R}^{(2)}$, we have that $D(i_1, \ldots, i_m)$ is equal to
\[
\left( \bigoplus_{\sigma \in C_m} (D(i_1, i_{\sigma(1)}) \otimes D(i_{\sigma(1)}, i_{\sigma^2(1)}) \otimes \cdots \otimes D(i_{\sigma^{m-1}(1)}, i_{\sigma^m(1)}))^{-\frac{1}{2}} \right).
\]

Remark 3.4. In each component $D(i_1, \ldots, i_m)$, the minimum is attained at least twice. Indeed, assume the minimum is attained for $\sigma = \tau$. Since
\[
f(D; \tau; i_1, \ldots, i_m) = f(D; \tau^{-1}; i_1, \ldots, i_m),
\]
the minimum is also attained for $\sigma = \tau^{-1}$. Note that this could be useful for computations, since it permits us to consider only $\frac{|S_m|}{2}$ permutations. Furthermore, if $\{i_1, i_k\}$ is a cherry of $T'$, the minimum is also attained for $\sigma = (jk) \circ \tau \circ (jk)$, whereby $(jk)$ is the transposition in $S_m$ switching $j$ and $k$.

Remark 3.5. The map $\phi^{(m)}$ is not injective on the whole domain $\mathbb{R}^{(2)}$. For example, consider $D, D' \in \mathbb{R}^{(2)}$, whereby $D(i, j) = 1$ for all $1 \leq i < j \leq n$ and $D'$ only differs from $D$ in the last coordinates, with $D'(n - 1,n) = 2$. Clearly, one has $D \in \mathcal{G}_{2,n}$, $D' \notin \mathcal{G}_{2,n}$ and $\phi^{(m)}(D) = \phi^{(m)}(D')$. However, Theorem 1.3 implies that the restriction of $\phi^{(m)}$ to $\mathcal{G}_{2,n}$ is injective if $n \geq 2m - 1$.

Proposition 3.6. $\phi^{(m)}(\mathcal{G}_{2,n}) \subseteq \mathcal{T}_{m,n} \cap \phi^{(m)}(\mathbb{R}^{(2)})$

Proof. The inclusion $\phi^{(m)}(\mathcal{G}_{2,n}) \subseteq \phi^{(m)}(\mathbb{R}^{(2)})$ is obvious, while $\phi^{(m)}(\mathcal{G}_{2,n}) \subset \mathcal{T}_{m,n}$ follows from [5]. For sake of completeness, we include the proof in this paper.

Consider a tree $T$ with leaf set $[n]$ and distance matrix $D$. Let $R$ be an $(m - 2)$-subset of $[n]$ and $i, j, k, l \in [n] \setminus R$. We have to prove that
\[
\phi^{(m)}(D) \in \mathcal{T}(\text{trop}(p_{R,i,j,kl})).
\]

Let $[R]$ be the smallest subtree of $T$ containing the leaves in $R$ and let $T'$ be the tree obtained from $T$ by contracting $[R]$ to a point. Denote by $i', j'$, etc. the images of respectively $i, j$, etc. in $T'$. Note that $R'$ is a leaf of $T'$. We have
\[
D(R,i,j) = D'(R',i',j') + D(R),
\]
hence $\phi^{(m)}(D) \in \mathcal{T}(\text{trop}(p_{R,i,j,kl}))$ if and only if $\phi^{(3)}(D') \in \mathcal{T}(\text{trop}(p_{R',i',j',kl})))$, where $D'$ is the distance matrix of $T'$.

Now Remark 4.1 below implies
\[
D'(R',i',j') = \frac{1}{2} D'(i',j') + D'(i',R') + D'(j',R'),
\]
so $\phi^{(3)}(D') \in \mathcal{T}(\text{trop}(p_{R',i',j',kl})))$ if and only if $D' \in \mathcal{T}(\text{trop}(p_{R',i',j',kl})))$. Hence the statement follows from Theorem 1.1.
4 The 3-dissimilarity maps of trees

Denote the coordinates of $\mathbb{R}^3$ by $X(i,j)$ (here we index over all integers $i, j$ with $1 \leq i < j \leq n$) and the coordinates of $\mathbb{R}^5$ by $X(i,j,k)$ (here we index over all integers $i, j, k$ with $1 \leq i < j < k \leq n$). Recall that if $D \in \mathcal{G}_{2,n}$ is a tree, $D(i,j)$ is the distance between leaf $i$ and leaf $j$.

Remark 4.1. Since $\mathcal{C}_3 = \{\sigma_1, \sigma_2\}$ with

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{and} \quad \sigma_2 = (\sigma_1)^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix},$$

the map $\phi^{(3)}$ sends $X = (X(i,j))_{i,j}$ to $(X(i,j,k))_{i,j,k}$ with

$$X(i,j,k) = \frac{1}{2} \cdot (X(i,j) + X(i,k) + X(j,k)).$$

So if $D \in \mathcal{G}_{2,n}$, the 3-subtree weights of the tree $D$ are given by $D(i,j,k) = \frac{1}{2} \cdot (D(i,j) + D(i,k) + D(j,k))$.

The following results states that for the case $m = 3$ the equality holds in Proposition 3.6 if $n \geq 5$.

Proposition 4.2. If $n \geq 5$, we have $\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{T}_{3,n} \cap \phi^{(3)}(\mathbb{R}^3)$

Proof. By Proposition 3.6, it is enough to show that for a general point $P \in \phi^{(3)}(\mathbb{R}^3) \cap \mathcal{T}_{3,n}$, there exists a point $D \in \mathcal{G}_{2,n}$ such that $\phi^{(3)}(D) = P$. Since $P \in \phi^{(3)}(\mathbb{R}^3)$, there exists a point $D \in \mathbb{R}^3$ such that $\phi^{(3)}(D) = P$. It suffices to prove that $D \in \mathcal{G}_{2,n}$. In order to do this, we show that in each triplet

$$\{D(i,j) + D(k,l), D(i,k) + D(j,l), D(i,k) + D(j,k)\},$$

the maximum is attained at least twice. Fix $S \in [n] \setminus \{i, j, k, l\}$ $(n \geq 5)$. Since $P \in \mathcal{T}_{3,n}$, in the triplet

$$\{P(S,i,j) + P(S,k,l), P(S,i,k) + P(S,j,l), P(S,i,l) + P(S,j,k)\},$$

the maximum is attained at least twice. Note that

$$P(S,i,j) + P(S,k,l) = \frac{1}{2} \cdot (C + D(i,j) + D(k,l)),
$$

$$P(S,i,k) + P(S,j,l) = \frac{1}{2} \cdot (C + D(i,k) + D(j,l)),
$$

$$P(S,i,l) + P(S,j,k) = \frac{1}{2} \cdot (C + D(i,k) + D(j,k)),$$

where $C = D(S,i) + D(S,j) + D(S,k) + D(S,l)$. Hence the maximum in $\{D(i,j) + D(k,l), D(i,k) + D(j,l), D(i,k) + D(j,k)\}$ is also attained at least twice, thus $D \in \mathcal{G}_{2,n}$ and $P \in \phi^{(3)}(\mathcal{G}_{2,n})$. \[ Q.E.D. \]
Definition 4.3. An ultrametric $D$ on $[n]$ is a metric which satisfies the following strengthened version of the triangle inequality:

$$\forall i, j, k \in [n] : D(i, j) \leq \max\{D(i, k), D(j, k)\}.$$ 

Equivalently, at least two of the three terms $D(i, j), D(i, k), D(j, k)$ are the same.

Remark 4.4. In general, the dissimilarity matrix $D$ of a tree $T$ is not an ultrametric. In case $D \in G_{2,n}$, it is equal. In particular, $2F = \max\{D(i, j) \mid i, j \in X \text{ and } i \neq j\}$.

Proposition 4.5. $\phi^{(3)}(G_{2,n}) \subset G_{3,n}$

Proof. Let $T$ be a tree with 3-dissimilarity map $P = (D(i, j))_{i,j} = \phi^{(3)}((D(i, j))_{i,j}) \in \phi^{(3)}(G_{2,n}) \subset \mathbb{R}^{(3)}$.

If $M \in K^{3 \times n}$, we denote the $(3 \times 3)$-minor with columns $i, j, k$ by $M(i, j, k)$. By Theorem 2.1, $G_{3,n}$ is the closure in $\mathbb{R}^{(3)}$ of the set

$$S := \{(-\text{val}(\det(M(i, j, k))))_{i,j,k} \mid M \in K^{3 \times n} \} \subset \mathbb{Q}^{(3)}.$$ 

Assume first that all the edges of $T$ have rational weights, a fortiori $P \in \mathbb{Q}^{(3)}$.

We are going to show there exists a matrix $M \in K^{3 \times n}$ such that

$$D(i, j, k) = -\text{val}(\det(M(i, j, k))).$$

Fix a rational number $E$ with $E \geq D(i, n)$ for all $i \in \{1, \ldots, n - 1\}$ and define a new metric $D'$ by

$$D'(i, j) = 2E + D(i, j) - D(i, n) - D(j, n)$$

for all different $i, j \in [n]$ (in particular, $D'(i, n) = 2E$ for $i \neq n$). Note that $D' \in G_{2,n}$ and that $D'$ is an ultrametric on $\{1, \ldots, n - 1\}$, so it can be realized by an equidistant ($n - 1$)-tree $T''$ with root $r$. Each edge $e$ of $T''$ has a well-defined height $h(e)$, which is the distance from the top node of $e$ to each leaf below $e$. Pick a random rational number $a(e)$ and associate the label $a(e)t^{2h(e)}$ to $e$. If $i \in \{1, \ldots, n - 1\}$ is a leaf of $T''$, define the polynomial $x_i(t)$ by adding the labels of all edges between $r$ and $i$. It is easy to see that $D'(i, j) = \deg(x_j(t) - x_i(t))$ for all $i, j \in \{1, \ldots, n - 1\}$.

Denote the distance from $r$ to each edge by $F$. Since

$$2F = \max\{D'(i, j) \mid 1 \leq i < j \leq n - 1\} < 2E,$$

we have $F < E$. The metric $D'$ on $[n]$ can be realized by a tree $T'$, where $T'$ is the tree obtained from $T''$ by adding the leave $n$ together with an edge $(r, n)$ of
length $2E - F$. If we define $x_n(t) = t^{2E}$, we get that $D'(i, j) = \deg(x_j(t) - x_i(t))$ for all $i, j \in [n]$.

Now consider the matrix

$$M' = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ x_1(t) & x_2(t) & x_3(t) & \ldots & x_n(t) \\ x_1(t)^2 & x_2(t)^2 & x_3(t)^2 & \ldots & x_n(t)^2 \end{bmatrix}.$$  

We have $\det(M'(i, j, k)) = (x_j(t) - x_i(t))(x_k(t) - x_i(t))(x_k(t) - x_j(t))$, hence

$$D'(i, j) + D'(i, k) + D'(j, k) = \deg(\det(M'(i, j, k))).$$

Let $M$ be the matrix obtained from $M'$ by multiplying, for each $i$, the $i$-th column of $M'$ by $(t^{D(i,n)} - E)^2$. Since

$$D(i, j) = D'(i, j) + (D(i, n) - E) + (D(j, n) - E) = \deg(t^{D(i,n)} - E \cdot t^{D(j,n)} - E \cdot (x_i(t) - x_j(t))),$$

we get that $D(i, j) + D(i, k) + D(j, k) = \deg(\det(M(i, j, k)))$. If we replace each $t$ in $M$ by $t^{-1/2}$, we get

$$D(i, j, k) = -\text{val}(\det(M(i, j, k))).$$

Now assume $T$ has irrational edge weights. We can approximate $T$ arbitrarily close by a tree $\tilde{T}$ with rational edge weights. From the arguments above, it follows that the 3-dissimilarity map $D$ of $\tilde{T}$ belongs to $S$, hence $D \in G_{3,n}$. □

**Theorem 4.6.** If $n \geq 5$, we have $\phi^{(3)}(G_{2,n}) = \phi^{(3)}(R(5)) \cap G_{3,n}$.

**Proof.** The statement follows from Proposition 4.2, Proposition 4.5 and the fact that $G_{3,n} \subset T_{3,n}$. □

## 5 The 4-dissimilarity maps of trees

In this section, we give a geometric description of $\phi^{(4)}(G_{2,n})$.

**Remark 5.1.** The set $C_4 = \{\sigma_1, \sigma_1^{-1}, \sigma_2, \sigma_2^{-1}, \sigma_3, \sigma_3^{-1}\}$ with

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}.$$

Hence the map $\phi^{(4)}$ sends $(X(i, j))_{i,j}$ to $(X(i, j, k, l))_{i,j,k,l}$ where $X(i, j, k, l)$ is equal to the minimum of the three terms

$$X(1, 2) + X(2, 3) + X(3, 4) + X(4, 1),$$

$$X(1, 2) + X(2, 4) + X(4, 3) + X(3, 1),$$

$$X(1, 3) + X(3, 2) + X(2, 4) + X(4, 1),$$

divided by two.
Consider $M = \mathbb{R}^{\binom{n}{2}(n-2)}$ and take $X(i, j; k, l)$, with $\{i, j, k, l\} \subset [n]$ a quadruple, as coordinates on $M$. For example, $X(j, i; l, k) = X(i, j; k, l)$ and $X(k, l; i, j) \neq X(i, j; k, l)$.

Let $\pi : \mathbb{R}^{\binom{n}{2}} \to M : (X(i, j))_{i,j} \mapsto (X(i, j; k, l))_{i,j,k,l}$ with

$$X(i, j; k, l) = \frac{1}{2} \cdot (X(i, j) + X(k, l) + \min\{X(i, k) + X(j, l), X(i, l) + X(j, k)\}).$$

Let $L$ be the linear subspace of $M$ consisting of points $X(i, j; k, l)$ with

$$X(i, j; k, l) = X(i, k; j, l) = X(i, l; j, k) = X(j, i; k, l) = X(k, l; i, j)$$

for all different $i, j, k, l \in [n]$. Points in $L$ can be projected naturally to $\mathbb{R}^{\binom{n}{4}}$ by sending $X(i, j; k, l)$ to $X(i, j, k, l)$. Denote this projection by $p$.

**Proposition 5.2.** $\phi^{(4)}(G_{2,n}) = p(\pi(\mathbb{R}^{\binom{n}{2}}) \cap L)$.

**Proof.** Note that for any real numbers $a, b, c$, we have

$$a + \min\{b, c\} = b + \min\{a, c\} = c + \min\{a, b\} \quad (5)$$

if and only if $\max\{a, b, c\}$ is attained at least twice. If the latter holds, the terms in (5) are equal to $\min\{a + b, a + c, b + c\}$.

If we take $a = X(i, j) + X(k, l)$, $b = X(i, k) + X(j, l)$ and $c = X(i, l) + X(j, k)$, the statement follows from the Tree Metric Theorem. 

**Acknowledgments**

We thank Ruriko Yoshida, Anders Jensen and especially Bernd Sturmfels for their many comments and suggestions which improved this manuscript. The second author is a postdoctoral fellow of the Research Foundation - Flanders (FWO).

**References**

[1] L. Billera, S. Holmes, K. Vogtman: *Geometry of the space of phylogenetic trees*, Advances in Applied Mathematics 27 (2001), 733-767.

[2] P. Buneman, *A Note on the Metric Properties of Trees*, J. Combinatorial Theory 17 (1974), 48-50.

[3] F. Chung, M. Garrett, R.L. Graham, D. Shalcross, *Distance realization problems with applications to Internet tomography*, J. Computer Systems and Sciences 63 (2001), 432-448.

[4] S.L. Hakimi, S.S. Yau, *Distance matrix of a graph and its realizability*, Quart. Appl. Math. 22 (1965), 305-317.
[5] L. Pachter, D. Speyer, *Reconstructing trees from subtree weights*, Applied Mathematics Letters 17 (2004), 615-621.

[6] L. Pachter, B. Sturmfels, *Algebraic statistics for computational biology*, Cambridge University Press, New York 2005

[7] D. Speyer, B. Sturmfels, *The Tropical Grassmannian*, Adv. Geom. 4 (2004), 389-411.

[8] B. Sturmfels, *Algorithms in Invariant Theory*, Texts and Monographs in Symbolic Computation, Springer-Verlag, Vienna, 1993.